NBER WORKING PAPER SERIES

GLP-1–INDUCED WEIGHT LOSS AND THE FEMALE OBESITY PENALTY

Rebecca Diamond

Working Paper 35387
http://www.nber.org/papers/w35387

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2026

The project described in this paper relies on data from survey(s) administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The content of this paper is solely the responsibility of the author and does not necessarily represent the official views of USC, UAS, or the National Bureau of Economic Research. All errors are my own.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Rebecca Diamond. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

GLP-1–Induced Weight Loss and the Female Obesity Penalty
Rebecca Diamond
NBER Working Paper No. 35387
June 2026
JEL No. I10, J01, J12

ABSTRACT

GLP-1 medications generate large weight loss and may also alter social and economic outcomes. Using the Understanding America Study, I compare women starting GLP-1s for weight loss with matched women who would like to start a GLP-1 but have not. Single women's marriage/ cohabitation rates rise by 29 percentage points and employment among baseline non-employed women rises 27 percentage points after six or more quarters. Existing partnerships do not dissolve, and already-employed women show no upward job mobility. The pattern suggests that part of the female obesity penalty operates at new-match formation rather than only through health or incumbent productivity.

Rebecca Diamond
Harvard University
Department of Economics
and NBER
rdiamond@fas.harvard.edu

GLP-1 receptor agonists let many people lose large amounts of weight that they could not lose on their own, and Americans have taken them up with remarkable speed. In the Understanding America Study, by the end of 2024, ten percent of adults had started a GLP-1, and six percent of women had started one specifically to lose weight. Take-up rises sharply with body mass. Nine percent of women with a history of obesity have started a GLP-1 to lose weight, against fewer than three percent of women with no history (Figure 1, Panel B). Clinical trials show why the drugs are so popular: sustained use produces weight loss of 12–18 percent (Wilding et al., 2021; Jastreboff et al., 2022).

Body weight is a visible health trait that may matter in markets through two different channels. It may affect own health, mobility, and productivity; or it may affect how others treat a person. A large economics literature documents that heavier women earn less, work less, and are less likely to marry or cohabit, with penalties that fall disproportionately on women (Averett and Korenman, 1996; Cawley, 2004; Mukhopadhyay, 2008; Chiappori, Oreffice and Quintana-Domeque, 2012; Cawley, 2015). But the causal interpretation of these gaps is difficult because body weight is correlated with health, family background, human capital, and local opportunity. Quasi-experimental and genetic-instrument evidence often suggests that cross-sectional wage penalties overstate the causal effect of weight (Brunello and d'Hombres, 2007; Lindeboom, Lundborg and van der Klaauw, 2010; Böckerman et al., 2019). At the same time, field-experimental evidence makes discrimination at first contact plausible: otherwise similar job applications receive fewer callbacks when the applicant is presented as obese (Rooth, 2009). The unresolved question is therefore not only whether weight matters, but *where* it matters: inside existing jobs and relationships, or at the moment new jobs and relationships are formed.

GLP-1 drugs create a new empirical setting. They generate large weight loss in a broader, non-surgical population and diffused during a period when access was constrained by cost, insurance coverage, physician prescribing, and drug shortages. I use the Understanding America Study to compare women who start a GLP-1 to lose weight with women who would like to start a GLP-1 but have not yet done so. This comparison restricts attention to women who have expressed demand for medical weight loss. Within that demand pool, I match on pre-treatment body mass, health, income, employment, partnership, and well-being, so the

comparison also holds fixed many of the factors that make some women more likely to seek or obtain treatment. I focus on women because GLP-1 use for weight loss is much more common among women in the data, leaving too few male weight-loss starters to estimate the same specifications I do for women with precision. I also exclude people who start a GLP-1 to manage diabetes, since diabetes starts appear to be triggered by illness onset rather than by a desire to lose weight.

The estimates show that GLP-1 weight loss changes outcomes on precisely the margins where visible body weight should affect first impressions. Single women form new partnerships. Among women who are single when they start, the probability of being married or living with a partner rises 18 percentage points overall and 29 percentage points after six or more quarters. The gain builds gradually as weight comes off. Women not employed at baseline also move into work. Their employment rate rises 13 percentage points overall and 27 percentage points after six or more quarters, and their weekly hours rise by nearly ten hours at the longer horizon. Much of the employment gain is movement out of unemployment rather than out of retirement or disability.

What does not change for women is equally informative. Among women already partnered at initiation, partnership status and separation or divorce are essentially unchanged.[1] Nor do I find evidence of upward job mobility among incumbent workers. Already-employed women do not increase hours, income per household member, or job-to-job moves. If anything, employer changes decline, possibly due to a desire to retain access to health insurance that covers these drugs. These null effects help distinguish a new-match channel from a general productivity or well-being channel.

Improved work capacity may explain part of the employment response among women not employed at baseline. At longer horizons, reported work-limiting health problems fall in this group. But the broader pattern is difficult to reconcile with a general health, mood, or motivation channel. Self-rated health does not improve, and life satisfaction, loneliness, and depression are largely flat or move in the wrong direction over the same horizon. These

[1]Appendix Table A16 reports analogous relationship-status estimates for men. The male weight-loss sample is much smaller, and the baseline-partnered male cell is smaller still, so I treat these estimates as suggestive. In contrast to women, baseline-partnered men show a 7 percentage point increase in leaving their baseline partner after treatment. This grows to 12 points by six or more quarters after initiation. The no-dissolution result should therefore be read as specific to women.

subjective-health patterns also distinguish GLP-1 weight loss from much of the bariatric-surgery weight loss evidence, where surgery often improves perceived health, health-related quality of life, self-esteem, mood, and depression (Konttinen et al., 2024; Hult et al., 2019; Dawes et al., 2016).

The markets that respond are the ones where someone forms a fresh impression of a woman's body weight: a prospective partner, or an employer considering an applicant who is not employed. For women, the arrangements that do not respond are the ones already in place, where any first impression occurred long ago and where weight is one characteristic embedded in a much richer stock of information. This pattern is consistent with an important part of the female obesity penalty operating as a first-impression discount. Reducing weight appears to remove part of that discount, but over the horizon I observe it does not raise income of the already employed or make women report feeling better off. The headline event studies are flat before initiation and the effects build only as weight falls. The results survive moderate Rambachan–Roth departures from parallel trends, randomization inference, alternative control pools, and external validation with the experimental medical trials in which my research design recovers large, trial-relevant weight loss and the trial-comparable physical-function pattern.

This paper contributes to three literatures. First, it adds to the economics of obesity by showing that part of the female obesity penalty operates at new-match formation rather than through incumbent productivity or wages. Prior work documents that heavier women have lower wages and family income, weaker employment outcomes, and lower rates of marriage or cohabitation (e.g., Averett and Korenman, 1996; Cawley, 2004; Conley and Glauber, 2006; Mukhopadhyay, 2008; Chiappori, Oreffice and Quintana-Domeque, 2012; Cawley, 2015), but estimates based on observed weight differences are difficult to interpret causally (Brunello and d'Hombres, 2007; Lindeboom, Lundborg and van der Klaauw, 2010; Böckerman et al., 2019). The results here use large pharmacological weight loss to separate two margins that are usually bundled together: the formation of new matches and outcomes inside existing matches.

Second, the paper extends the medical weight-loss literature from bariatric surgery to pharmacologic, non-surgical weight loss. Bariatric-surgery patients are a more selected and more

severely obese population than GLP-1 weight-loss starters. In Bruze et al. (2018), mean pre-treatment BMI among surgery patients was about 42, compared with 35 among GLP-1 weight-loss starters in my sample. Existing surgery evidence shows that large weight loss can change relationship status, but in both directions. In Swedish registry and cohort data, surgery patients who are single at baseline are more likely to marry or enter a new relationship, while those already partnered are also more likely to separate or divorce (Bruze et al., 2018).[2] In women, I find the first margin but not the second: GLP-1 initiation is followed by a large increase in partnership formation among initially single women, with no detectable increase in dissolution among initially partnered women. This contrast is consistent with GLP-1s being a less invasive treatment used by a less severely obese population, and with the shorter follow-up horizon in my data. It also suggests that pharmacologic weight loss for women operates primarily through new match formation rather than by destabilizing existing relationships.[3]

The labor-market comparison points in the same direction. The bariatric-surgery literature finds its clearest gains on extensive-margin employment outcomes, while evidence for sustained employment and earnings gains among the employed is more mixed. A recent systematic review concludes that employment often rises in the short run after surgery but returns toward baseline over longer horizons, as weight is partially regained (Xia et al., 2026). Large matched-register studies from Sweden find little evidence that surgery improves earnings relative to matched controls, including in female samples (Norrbäck et al., 2021, 2023). My GLP-1 estimates mirror the extensive-margin pattern in a non-surgical population. Women not employed at initiation enter work, while women already employed at initiation do not increase hours or change jobs. This pattern is consistent with an obesity penalty that is strongest at hiring rather than inside existing jobs, although improved work capacity likely explains part of the employment-entry response.

Third, the paper provides early panel evidence on the social and economic consequences of GLP-1 diffusion. A rapidly growing health-economics literature asks whether GLP-1s im-

[2]The main U.S. evidence on relationship status after bariatric surgery is King, Hinerman and White (2022), which tracks surgical patients over time and benchmarks marital transitions to general-population rates rather than to an individual-level matched control group.

[3]Appendix E suggests that, in the smaller male sample, partnered men are about 6.8 percentage points more likely to leave their baseline partner after GLP-1 initiation.

prove clinical outcomes enough to reduce medical spending or offset their high costs (Bock, Moshfegh and Zhang, 2026; Wing et al., 2026), and related work studies how GLP-1s change consumer demand (Hristakeva, Liaukonytė and Feler, 2026). I instead study outcomes outside the health-care system and consumer-goods markets: partnership formation, employment, hours, income, job mobility, and well-being. These non-medical outcomes are potentially important for the private and social value of medical weight loss, but they are largely absent from the existing evidence.

The closest related paper is Kaestner and Schiman (2026), who use MEPS prescription data to study GLP-1 diffusion among adults with diabetes and find little average short-run change in mental health, self-rated health, employment, or marriage. My results are complementary. Their setting is diabetes treatment, whereas I isolate initiation for weight loss; their MEPS design is also best suited to short-horizon average effects. When I aggregate my estimates across baseline employment and partnership status and use their short-run horizon, my confidence intervals closely overlap theirs for employment, partnership, and health. The new evidence here comes from following weight-loss starters over longer horizons and splitting women by baseline employment and partnership. Among women single at initiation, partnership rises as weight loss accumulates; among women not employed at initiation, employment and hours rise at longer horizons.

I. Data

The UAS is a nationally representative internet panel of roughly 15,000 U.S. adults (Alattar, Messel and Rogofsky, 2018). The GLP-1 module, fielded between December 2024 and February 2025 to 10,085 respondents, records each person's GLP-1 use status (current user, former user, never used but interested, never used and not interested), the month they first started a GLP-1, the reason for starting (diabetes management or weight loss), and out-of-pocket cost. I distinguish four sample counts. The raw module fielding count is 10,085 respondents. The positive-weight module sample, used for descriptive tables, is 9,758 respondents with positive post-stratification weights and valid GLP-1 use status. In that positive-weight module sample, 369 women and 90 men report starting a GLP-1 for weight loss, while 319 women and 211 men report starting for diabetes management. Event-time

analyses additionally require usable first-start timing; that leaves 367 female and 88 male weight-loss starters, and 302 female and 202 male diabetes starters, before the later analysis restrictions described below.

I focus exclusively on the treatment effects of GLP-1 used for weight loss, as opposed to for diabetes management. Diabetes usage is confounded by diabetes onset or worsening. The advent of GLP-1s for weight loss marks a major change in the medical treatment of obesity, making most take-up driven by the new innovation in medical weight loss rather than by worsening health or obesity.

The UAS is uniquely well-suited for studying the effects of GLP-1s on social and economic outcomes. It combines high-frequency panel data with a wide set of health, economic, and social outcomes. Most importantly, the data are available for very recent periods (through May 2026), making it possible to study the recent surge in GLP-1 utilization for weight loss. The quarterly household panel supplies labor-force status, weekly hours, household income, and marital and cohabitation status from 2021 to 2026.[4] A biennial panel supplies self-reported BMI, diabetes, and questions related to health, depression, loneliness and mood.[5] A monthly survey that started in 2023 covers life satisfaction, self-rated health, depression screener questions, and loneliness. These variables were measured biennially before 2023, so I combine the sources. A May 2026 survey re-measures GLP-1 status, which lets me extend the analysis window to May 2026.

I code partnership as being married or cohabiting; employment as working or on paid leave; unemployment as on layoff or actively looking; weekly hours are unconditional, with nonworkers assigned zero; and health, depression, loneliness, and life satisfaction are standardized after harmonizing biennial and monthly measures.

II. Who Starts a GLP-1 to Lose Weight?

GLP-1 use has risen sharply over the past few years. Figure 1, Panel A, shows that 17 percent of women with a history of obesity have tried a GLP-1 as of the end of 2024.[6] Fourteen

[4] I start my analysis in 2021 to drop COVID-era lockdowns.

[5] There are also Health and Retirement Survey-style questions on wealth, individual labor market earnings and food consumption, but due to the amount of noise in these data and the sample size, they do not offer any insight.

[6] The May 2026 follow-up re-measures GLP-1 status but does not collect first-start dates, so I do not extend the quarterly adoption curves beyond 2024Q4.

percent of men with a history of obesity have tried a GLP-1. Even among respondents never observed with obesity, adoption rates are rising: nearly 5 percent of women and 2.5 percent of men in this group have tried these drugs. Panel B shows adoption for use exclusively for weight loss rather than diabetes management. Nine percent of women and 4 percent of men with a history of obesity have tried GLP-1s for weight loss. Usage among respondents never observed with obesity is 2.5 percent among women and less than 1 percent among men. The main analysis focuses on women because the female weight-loss starter sample is much larger; Appendix E discusses men briefly.

Table 1 compares the women who started a GLP-1 to lose weight with two groups of never-users: the women who would like to start a GLP-1 but have not yet, and the women who report no interest. Three facts stand out.

First, interest in and use of GLP-1s for weight loss are concentrated among women with high BMI. The baseline BMI of GLP-1 users is 35, very similar to that of women who report interest in taking a GLP-1 and well above the uninterested group, whose average BMI is near 28.

Second, GLP-1 adopters' health and well-being look more like those of women not interested in taking GLP-1s than those of interested non-adopters. They rate their own health (z-scored) better than the interested do (-0.07 against -0.29), and only slightly below the uninterested (0.07). They are less likely to have diabetes than the interested and have a diabetes rate similar to the uninterested. GLP-1 users are more satisfied with their lives, less lonely, and less depressed than interested non-adopters, and very similar to the uninterested.

Third, adopters are the most financially advantaged. Their household income and their own earnings are the highest of the three groups. Adopters' households take in $83,000 a year, against $60,000 for the women who want the drug. Cost looks like a real barrier. Forty percent of GLP-1 users pay for the drug fully out of pocket, at $299 a month, so the women who start are disproportionately the ones who can afford to. Similar summary statistics for men are reported in Appendix Table A15. Due to small samples, many comparisons are statistically imprecise. However, male users also have higher income than the other two groups and have the highest BMI. Relative to female GLP-1 users, male GLP-1 users look sicker: they report higher BMI, a higher diabetes rate, and lower self-rated health.

III. Empirical Design

I focus on women aged 25 to 61 at treatment to avoid changes associated with Medicare eligibility.[7] Treatment is defined as starting a GLP-1 for weight loss and does not condition on continuing use through the end of the sample. Table 1 shows that 53% of women who start treatment report current use at the GLP-1 module. I define the treatment quarter, g, as one quarter before the reported start month, to allow for the possibility of slight measurement error in retrospective data reporting. Using the exact reported start month shows essentially identical results. The control group comes from women who stated that they would like to start a GLP-1 but have not as of a given event time. This pool combines interested never-users with later adopters, the latter observed only in the quarters before their own start. For each cohort I take baseline characteristics from the last quarterly interview before g, and form exact-match cells on employment and partnership status. Within cells, controls are propensity score re-weighted, with propensity scores estimated using race dummies, lagged BMI, diabetes status, self-rated health, depression score, whether health limits work, life satisfaction, household income, and cohort fixed effects. The benchmark sample is 242 treated women and 850 matched controls. Balance is tight after reweighting. The variables in the propensity score: race, lagged BMI, diabetes, self-rated health, depression, whether health limits work, life satisfaction, and household income—have standardized treated-control differences at or below 0.03 after matching and reweighting within the baseline employment and partnership splits used below. Variables excluded from the propensity score also balance, including age, college share, cumulative employer and job changes since panel entry, and unemployment.

Stacking the cohorts, I estimate

$$y_{ipt} = \alpha_{ip} + \lambda_{pt} + \sum_{e \neq -1} \beta_e \, D_{ip} \mathbf{1}\{t - g_p = e\} + \varepsilon_{ipt}, \tag{1}$$

where α_{ip} are stack-by-person fixed effects, λ_{pt} are cohort-by-calendar-quarter fixed effects, and e is event time. To smooth out some quarterly variation in survey non-response, I measure event time at the half-yearly level, so each e is a two-quarter period. For post-

[7] As of May 2026, Medicare does not cover GLP-1s for weight loss.

period averages I replace the event-time terms with a single post indicator, split at six quarters. Standard errors cluster by person. Section V reports studentized randomization-inference p-values for the headline estimates. Outcomes run through 2026Q2. Because the GLP-1 module is observed once in early 2025 while outcomes extend through 2026Q2, the benchmark uses an interim verified-control filter: May-2026-confirmed never-users contribute through 2026Q2, and controls whose late-2025 tracking-wave record still shows never-use contribute through 2025Q3.

IV. Results

Weight loss. BMI is flat before women start and then falls (Figure 2, Panel A). Splitting women by baseline employment status (Table 3) and separately by baseline partnership status (Table 2) shows no meaningful difference in weight loss across groups. All groups lose between 1.4 and 2.5 BMI points in the post period. These effects grow to between 2.7 and 4.2 BMI points at 1.5+ years after initiation.

The marriage market. The largest magnitude change I find is in the marriage market. Women who are single when they start (Figure 2, Panel B) increase their probability of being married or living with a partner by 18.3 percentage points overall and 28.6 percentage points (s.e. 7.1) at six or more quarters, building gradually as they lose weight. The gradual build suggests reverse causality is unlikely. If single women began weight loss drugs because they had just started a non-cohabiting romantic relationship, we should find most of the increase in cohabitation or marriage near initiation, not a gradual increase over three years as weight drops.

Partnered women's relationships, by contrast, are stable. Their partnership probability barely moves (0.1 points, s.e. 1.8), and there is no excess relationship dissolution (0.0, s.e. 1.7; Table 2). The marriage-market response is the making of new matches, not the breaking or remaking of old ones. Existing relationship matches are presumably based on characteristics beyond a woman's weight, so weight loss does not reveal the same new information to her partner. In contrast, single women's weight loss seems to dramatically affect first impressions of potential partners, leading to a large increase in marriage and cohabitation. This is consistent with prior work on the obesity penalty negatively affecting women in

the marriage market. Similar results are found in studies of marriage-market responses to bariatric surgery-driven weight loss (Bruze et al., 2018).

The dramatic increase in partnership rates after single women start a GLP-1 could be driven by a change in mood and motivation to date, rather than by changes in marriage-market body-weight discrimination. To test this alternative hypothesis, I look at GLP-1 impacts on a z-score depression screener, a loneliness screener, and life satisfaction. Table 2 shows that single women experience no statistically significant change in depression, loneliness, or life satisfaction, although these effects are noisy. The point estimate for depression is positive post treatment, and the lower bound of the 95 percent CI rejects an improvement in depression of more than 0.09 SDs. The point estimate for loneliness suggests an improvement of 0.1 standard deviations after 1.5+ years. Standard errors prevent ruling out improvements as large as 0.42 SDs post treatment, possibly indicating the new partnerships lower loneliness. The point estimate on life satisfaction is negative, and the 95 percent CI rejects an effect of 0.06 SD improvement or larger. The 1.5+ year point estimates are mixed rather than coherently positive, and standard errors grow as well. Despite the noise, it would be hard to rationalize the 29 percentage point increase in partnership rates among single women by changes in mood, but the estimates are imprecise. A first-impression mechanism in which potential partners respond to body weight is consistent with these effects.

The labor market. Labor-market responses differ dramatically by baseline employment status. Table 3 shows that women not employed at baseline increase their employment rate by 13.2 percentage points (s.e. 5.6) post treatment, growing to 26.9 percentage points (s.e. 8.4) after 1.5+ years. Hours worked increase by 4.7 per week post treatment (s.e. 2.1) and by 9.9 hours after 1.5+ years (s.e. 3.1). Household log income grows 12–18 percent, although standard errors prevent me from rejecting a zero effect. Much of this increase in labor supply comes from previously unemployed women looking for work. Unemployment falls by 8.2 percentage points post treatment and 16.1 percentage points after 1.5+ years. The rest of the labor-supply increase appears to come from women previously out of the labor force, as declines in disability status and retirement are essentially zero, although standard errors cannot rule out modest effects.

Baseline-employed women look very different. Their employment rates fall by 2.6 percent-

age points (s.e. 1.5) post treatment and 4.4 percentage points (s.e. 2.2) after 1.5+ years. Weekly hours fall 1–1.5 hours per week (a 2.5–4 percent decline from a pre-period mean of 37.8 hours per week). Increases in unemployment, disability, retirement, and not being in the labor force all appear to absorb this employment decline, although statistical power is low for detecting where these no-longer-employed women go. Despite this decrease in labor supply, household income *rises* by 7 percent post treatment and 10 percent after 1.5+ years. This tracks the group's increased partnership rate, with the new partner bringing significant income to the household. Even standardizing household income by number of household members shows an increase, suggesting that the new partners are higher earners than these women were.

Despite the small labor-supply reduction among previously employed women, I find that these women change employers 0.13 fewer times overall and 0.18 fewer times after 1.5+ years.[8] I do not see any change in employer-change frequency among women not employed at baseline. The reduction is concentrated among women whose insurance covers their GLP-1, suggesting they may be staying at their baseline employer to maintain access to health insurance that will cover these medications (Appendix Table A20).

There are two plausible reasons women not employed at baseline work more after weight loss: a decline in weight-related physical or health-related work impediments, or a decrease in weight-related labor-market discrimination. I can directly test the first hypothesis by looking at treatment effects on the share of women reporting an impairment that limits the kind or amount of work they can do. I find that women employed at baseline and women not employed at baseline report a 3 to 4.5 percentage point *increase* in the share saying their health limits their work. This may reflect the very common GI side effects of GLP-1s, such as nausea, vomiting, diarrhea, and fatigue, especially when users' bodies are first adapting to these drugs. After 1.5+ years, the effect for women not employed at baseline reverses to a 14.2 percentage point *decline* in reporting that health limits their ability to work, while women employed at baseline continue to report a 3 percentage point increase, although I cannot reject a zero effect at this horizon. The decline in reported work limitations suggests that improved work capacity may be an important component of the employment response.

[8] I count a switch into not employed as an "employer" change.

At the same time, the employment increase is larger than the decline in reported work limitations, so the pattern is also consistent with changes in hiring associated with visible weight loss.

It is possible that weight loss changes women's mental health and life outlook, making them more desirable hires regardless of weight. Treatment effects on life satisfaction, depression, and loneliness make it unlikely that the estimates are driven solely by this alternative hypothesis. Women not employed at baseline are 0.44 standard deviations more depressed at baseline than the national average, and their depression z-score increases slightly after 1.5 years by 0.02 (s.e. 0.17). Their loneliness z-score increases by 0.28 standard deviations (s.e. 0.19), and life satisfaction falls by 0.17 SDs (s.e. 0.13). While these effects are noisy, they move in the opposite direction of a mood and motivation improvement, leaving little evidence for substantial positive effects. Women employed at baseline do not show broad well-being gains either: depression rises, life satisfaction falls, and loneliness is a noisy zero. Both groups also report *declines* in self-reported health and an increase in the share reporting that their health is fair or poor. Despite the weight loss, employment gains, and partner gains, these women do not appear to feel healthier or happier.

Both the employment gains for women not employed at baseline and the partnership gains for women single at baseline occur through newly formed relationships that are likely influenced by first impressions. This is exactly where discrimination based on body weight would be most likely to show up. Existing employment relationships seem to become more stable (fewer employer changes), and existing partnerships do not dissolve. These incumbent relationships are based on more than weight, and they remain durable.

V. Identification and Robustness

The main threat to my identification assumption is that women who start a GLP-1 may have been on different untreated trajectories from matched women who wanted a GLP-1 but had not yet started one. I assess this concern in several ways. First, I estimate differential linear pre-trends using only the pre-period from the same matched stacks, event clocks, and baseline subgroups as in Tables 2 and 3. Across the full Tables 2–3 outcome grid, only 1 of 63 pre-trend slopes reject at the five-percent level, compared with about 3 expected by chance.

See Appendix Table A4. Covariate balance is also tight after matching and reweighting: the propensity-score covariates have standardized treated-control differences at or below 0.03, and direct balance tests do not approach conventional significance levels. See Appendix Figures A1 and A2.

I next ask how much differential post-treatment drift would be needed to overturn the headline effects. I use the relative-magnitude sensitivity analysis of Rambachan and Roth (2023), where $\bar{M}$ bounds the possible post-treatment violation of parallel trends as a fraction of the largest pre-treatment deviation. For the average effect over quarters 6–18, the partnership effect among baseline-single women remains positive at $\bar{M} = 0.25$, with a 95 percent confidence set of $[0.16, 0.76]$, and the lower bound rounds to zero at $\bar{M} = 0.5$, $[0.00, 0.93]$. The employment effect among women not employed at baseline is even more robust: its confidence set is $[0.21, 0.65]$ at $\bar{M} = 0.25$ and $[0.11, 0.76]$ at $\bar{M} = 0.5$. Weekly hours remain significant at $\bar{M} = 0.25$, but confidence intervals start to touch zero at $\bar{M} = 0.5$.

Finite-sample inference gives similar conclusions. I implement randomization inference by reshuffling treatment within cohort $\times$ exact-match cells and using the studentized placebo statistic. The two-sided randomization-inference p-value is 0.006 for the quarters 6–18 partnership effect among baseline-single women. The corresponding p-values for employment, hours, and unemployment among baseline-not-employed women are 0.053, 0.044, and 0.043, respectively. See Appendix Table A12. Thus the partnership result is strongly supported by the randomization-inference exercise, while the labor-market extensive-margin results are supported but closer to conventional significance thresholds.

The results are not driven by the choice of using women who desire a GLP-1 as the control group. Re-estimating the main specification with interested never-users only, all non-users, or only uninterested never-users preserves the same patterns: partnership remains positive, employment remains small on average, and weight loss remains clear whenever a never-user pool anchors the comparison (Appendix Tables A6–A11). Estimating treatment-effect heterogeneity by baseline employment and partnership status in a single pooled interaction model recovers the same headline results (Appendix Tables A18 and A19). When pooling all women into a single sample and not cutting by baseline employment or partnership status, the effect on employment is smaller and no longer statistically significant, as 72 percent of

the sample is employed at baseline (Appendix Table A13). The marriage and cohabitating effects fall to 6.2 percentage points post treatment and 10.2 percentage points after 1.5+ years and remain statistically significant. Again, these are smaller because 66 percent of the pooled sample is partnered at baseline (Appendix Table A13).

As a final check of my research design and methods, I benchmark my findings against the randomized medical trials on these drugs (Wilding et al., 2021; Jastreboff et al., 2022) and the medical literature studying the real-world use of GLP-1s (Gasoyan et al., 2025; Peters et al., 2026). When I estimate effects on weight loss at the same time horizon (about 1.25 years) as the semaglutide and tirzepatide trials and pool my sample of GLP- 1users across genders, I find GLP-1 usage leads to an 8.6 percent body-weight decline, rising to 9.6 percent in a no-diabetes BMI/comorbidity sample that approximates the medical trials eligibility. These estimates are below the trial point estimates of 12.4 percent for semiglutide and 11.9–17.8 percent for tirzepatide, but they line up closely with real-world clinical-practice evidence. Gasoyan et al. (2025) estimate 8.7 percent weight loss at one year among real-world GLP-1 patients. This comparison is useful because real-world use differs in important ways from the clinical trials. Real world uses more often discontinue treatment and use lower doses than those in the clinical trials (Gasoyan et al., 2025). Another key difference was those in the trials also received intensive lifestyle support around nutrition and exercise. Consistent with that implementation gap, UAS effects on physical, general, and mental health are small and near zero, consistent with real-world clinical evidence (Peters et al., 2026), but lower than what was found in the clinical trials (Rubino et al., 2024). Appendix F gives the full construction and discusses of these findings and how relate to the clinical trial and real-world results.

A remaining concern is endogenous treatment timing. A woman may start a GLP-1 at the same time that she re-enters dating, prepares for a partnership transition, searches for work, or makes a broader investment in self-improvement. Such a contemporaneous shock would not necessarily generate differential pre-trends and could bias the difference-in-differences estimates even if the preceding diagnostics pass. Several patterns make this explanation less consistent with the data. First, the headline partnership and employment effects are not immediate: they emerge at longer horizons, when weight loss has accumulated,

rather than jumping in the first quarters after initiation. Second, GLP-1 initiation is not accompanied by broad improvements in subjective health or well-being. I find no significant improvements in self-rated health, depressive symptoms, loneliness, or life satisfaction, and the pre-period slopes for these outcomes are small and statistically insignificant. Third, the responses are concentrated in the baseline groups at risk of forming new matches. If treatment coincided with a general desire to transform one's life, similar changes should also appear among women already employed or already partnered, who make up most of the treated sample. Instead, already-employed women do not increase hours, household income per member, or job and employer changes. Baseline-partnered women are no more likely to remain partnered or to leave their baseline partner. The margins that respond are the ones where a new evaluator forms an impression: prospective partners for initially single women and employers for initially non-employed women. Finally, the weight-loss change is not unusually large. The UAS estimate lines up closely with real-world clinical-practice evidence and remains below the randomized-trial efficacy benchmarks, consistent with ordinary real-world GLP-1 use rather than an unusually intensive lifestyle-change episode among treated women. These diagnostics do not rule out all unobserved selection into treatment timing, but they make the leading "life-change" explanation less consistent with the timing, magnitude, and cross-outcome pattern of the estimates.

Taken together, these checks make it less likely that the headline partnership and employment-entry and hours effects are driven by differential pre-trends, finite-sample inference, the particular control pool, or endogenous timing of treatment.

VI. Men

Because the analysis sample contains a smaller sample of treated men, estimates are too noisy for strong conclusions and cannot be split by baseline employment and partner status. The most striking result I find among men is that, unlike women, they leave their baseline partners after treatment. Among men partnered at baseline, treatment increases the probability that they are no longer with that partner by 6.8 percentage points (s.e. 2.9). This grows to 12.1 percentage points (s.e. 6.1) after 1.5+ years. I also find that men 1.5+ years after treatment are 16 percentage points (s.e. 7.6) less likely to report that their health limits

the work they can do. This magnitude is quite similar to what I find for women not employed at baseline. The remaining male estimates are too noisy to support strong conclusions. See Appendix E for further discussion of the male effects.

VII. Conclusion

This paper uses the rapid diffusion of GLP-1s for weight loss to study whether large, non-surgical reductions in body weight change women's social and labor-market outcomes. The estimates point to a simple pattern. Outcomes that require a new match change. Among women who are single when they start treatment, marriage or cohabitation rises by 29 percentage points after six or more quarters. Among women not employed at initiation, employment rises by 27 percentage points and weekly hours rise by nearly ten hours at the same horizon. By contrast, outcomes embedded in existing matches change little. Existing partnerships do not dissolve, already-employed women do not move to better jobs, and subjective health and well-being do not improve in parallel.

This pattern helps interpret the female obesity penalty. Body weight can matter because it affects health, mobility, productivity, confidence, or because other people attach penalties to a visible trait. The results here are most consistent with a substantial first-impression obesity penalty. The response is concentrated where another person is forming a new assessment of the woman—a prospective partner or an employer considering an applicant who is not employed—and is muted where richer information has already accumulated.

The findings also broaden the welfare questions raised by GLP-1 diffusion. Evaluating these drugs only through clinical endpoints or medical spending omits margins that are central to economic well-being: employment entry, hours, household resources, and family formation. At the same time, these nonmedical returns raise distributional questions. Early adopters are disproportionately financially advantaged, and access is shaped by insurance coverage, out-of-pocket costs, and physician prescribing. If medical weight loss changes how women are treated in markets, then unequal access to the drugs may also shape who can avoid the social and economic penalties attached to body size. It is also unsettling to think about using medical treatment as a possible remedy for discrimination.

Over the horizon observed here, GLP-1 initiation changes new social and labor-market

matches more than it changes incumbent productivity, perceived health, or well-being. Future work will show whether these extensive-margin gains persist, whether earnings gains emerge, and whether similar effects arise for men as sample sizes grow. The early evidence suggests that mass pharmacological weight loss is not only a health shock. It is also a shock to the social and labor-market valuation of body weight.

REFERENCES

Alattar, Laith, Matt Messel, and David Rogofsky. 2018. "An Introduction to the Understanding America Study Internet Panel." *Social Security Bulletin*, 78(2): 13–28.

Averett, Susan, and Sanders Korenman. 1996. "The Economic Reality of the Beauty Myth." *Journal of Human Resources*, 31(2): 304–330.

Böckerman, Petri, John Cawley, Jutta Viinikainen, Terho Lehtimäki, Suvi Rovio, Ilkka Seppälä, Jaakko Pehkonen, and Olli Raitakari. 2019. "The Effect of Weight on Labor Market Outcomes: An Application of Genetic Instrumental Variables." *Health Economics*, 28(1): 65–77.

Bock, Sam, Jasmin Moshfegh, and Jonathan Zhang. 2026. "Weighing the Impacts of GLP-1s: Quasi-Experimental Evidence from Provider Adoption." National Bureau of Economic Research Working Paper 34667.

Brunello, Giorgio, and Béatrice d'Hombres. 2007. "Does Body Weight Affect Wages? Evidence from Europe." *Economics and Human Biology*, 5(1): 1–19.

Bruze, Gustaf, Tobias E. Holmin, Markku Peltonen, Johan Ottosson, Kajsa Sjöholm, Ingmar Näslund, Martin Neovius, Lena M. S. Carlsson, and Per-Arne Svensson. 2018. "Associations of Bariatric Surgery With Changes in Interpersonal Relationship Status: Results From 2 Swedish Cohort Studies." *JAMA Surgery*, 153(7): 654–661.

Cawley, John. 2004. "The Impact of Obesity on Wages." *Journal of Human Resources*, 39(2): 451–474.

Cawley, John. 2015. "An Economy of Scales: A Selective Review of Obesity's Economic Causes, Consequences, and Solutions." *Journal of Health Economics*, 43: 244–268.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque. 2012. "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market." *Journal of Political Economy*, 120(4): 659–695.

Conley, Dalton, and Rebecca Glauber. 2006. "Gender, Body Mass, and Socioeconomic Status: New Evidence from the PSID." *Advances in Health Economics and Health Services Research*, 253–275.

Dawes, Aaron J., Melinda Maggard-Gibbons, Alicia R. Maher, Marika J. Booth, Isomi Miake-Lye, Jessica M. Beroes, and Paul G. Shekelle. 2016. "Mental Health Conditions Among Patients Seeking and Undergoing Bariatric Surgery: A Meta-analysis." *JAMA*, 315(2): 150–163.

Gasoyan, Hamlet, W Scott Butsch, Rebecca Schulte, Nicholas J Casacchia, Phuc Le, Christopher B Boyer, Marcio L Griebeler, Bartolome Burguera, and Michael B Rothberg. 2025. "Changes in weight and glycemic control following obesity treatment with semaglutide or tirzepatide by discontinuation status." *Obesity*, 33(9): 1657–1667.

Hristakeva, Sylvia, Jūra Liaukonytė, and Leo Feler. 2026. "The No-Hunger Games: How GLP-1 Medication Adoption Is Changing Consumer Food Demand." *Journal of Marketing Research*, 63(3): 427–444.

Hult, Mari, Stephanie E. Bonn, Lena Brandt, Mikael Wirén, and Ylva Trolle Lagerros. 2019. "Women's Satisfaction with and Reasons to Seek Bariatric Surgery—a Prospective Study in Sweden with 1-Year Follow-up." *Obesity Surgery*, 29: 2059–2070.

Jastreboff, Ania M., Louis J. Aronne, Nadia N. Ahmad, Sean Wharton, Lisa Connery, Breno Alves, Arihiro Kiyosue, Shuyu Zhang, Bing Liu, Mathijs C. Bunck, and Adam Stefanski. 2022. "Tirzepatide Once Weekly for the Treatment of Obesity." *New England Journal of Medicine*, 387(3): 205–216.

Kaestner, Robert, and Cuiping Schiman. 2026. "The Effects of GLP-1 Use on Mental Health, Self-rated Health, Employment and Marriage." National Bureau of Economic Research Working Paper 35198.

King, Wendy C, Amanda S Hinerman, and Gretchen E White. 2022. "Changes in marital status following Roux-en-Y gastric bypass and sleeve gastrectomy: a US multicenter prospective cohort study." *Annals of Surgery Open*, 3(3): e182.

Konttinen, Hanna, Kajsa Sjöholm, Lena M. S. Carlsson, Markku Peltonen, and Per-Arne Svensson. 2024. "Fifteen-year changes in health-related quality of life after bariatric surgery and non-surgical obesity treatment." *International Journal of Obesity*, 48: 1447–1456.

Lindeboom, Maarten, Petter Lundborg, and Bas van der Klaauw. 2010. "Assessing the Impact of Obesity on Labor Market Outcomes." *Economics and Human Biology*, 8(3): 309–319.

Mukhopadhyay, Sankar. 2008. "Do Women Value Marriage More? The Effect of Obesity on Cohabitation and Marriage in the USA." *Review of Economics of the Household*, 6(2): 111–126.

Norrbäck, Mattias, Martin Neovius, Johan Ottosson, Ingmar Näslund, and Gustaf Bruze. 2021. "Earnings and Employment for Women after Bariatric Surgery: A Matched Cohort Study." *International Journal of Obesity*, 45(4): 766–775.

Norrbäck, Mattias, Martin Neovius, Johan Ottosson, Ingmar Näslund, and Gustaf Bruze. 2023. "Earnings and Work Loss from 5 Years before to 5 Years after Bariatric Surgery: A Cohort Study." *PLoS ONE*, 18(5): e0285379.

Peters, Warren, Tea Monk Fries, Shilpa Middolu, Lauren Wilson, and Scott H. Lee. 2026. "Real-World Comparative Effectiveness of Semaglutide 2.4 mg Versus Other Commercially Available Medications for Chronic Weight Management in Adults with Obesity in the United States: A Pragmatic Clinical Study." Philadelphia, PA, USA. Presentation CTI, Code P6. Published in Value in Health, Volume 29, Issue S6.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies*, 90(5): 2555–2591.

Rooth, Dan-Olof. 2009. "Obesity, Attractiveness, and Differential Treatment in Hiring: A Field Experiment." *Journal of Human Resources*, 44(3): 710–735.

Rubino, Domenica M., Frank L. Greenway, Usman Khalid, Patrick M. O'Neil, Julio Rosenstock, Rasmus Sørrig, Thomas A. Wadden, Sean Wharton, and John P. H. Wilding. 2024. "Effect of Once-Weekly Semaglutide on Patient-Reported Outcomes in Adults with Overweight or Obesity: The STEP 1 to 4 Trials." *Diabetes, Obesity and Metabolism*, 26(6): 2129–2138.

Wilding, John P. H., Rachel L. Batterham, Salvatore Calanna, Melanie Davies, Luc F. Van Gaal, Ildiko Lingvay, Barbara M. McGowan, Julio Rosenstock, Marie T. D. Tran, Thomas A. Wadden, Sean Wharton, Koutaro Yokote, Niels Zeuthen, and Robert F. Kushner. 2021. "Once-Weekly Semaglutide in Adults with Overweight or Obesity." *New England Journal of Medicine*, 384(11): 989–1002.

Wing, Coady, Sih-Ting Cai, Daniel W. Sacks, and Kosali I. Simon. 2026. "Do GLP-1 Medications Pay for Themselves?" National Bureau of Economic Research Working Paper 34678.

Xia, Qing, Thomasina Donovan, Jiahao Diao, Xiao Yang, Xingzhong Jin, Stella T. Lartey, and Hannah E. Carter. 2026. "Bariatric Surgery, Employment, and Productivity Outcomes: A Systematic Review." *JAMA Surgery*, 161(1): 84–94.

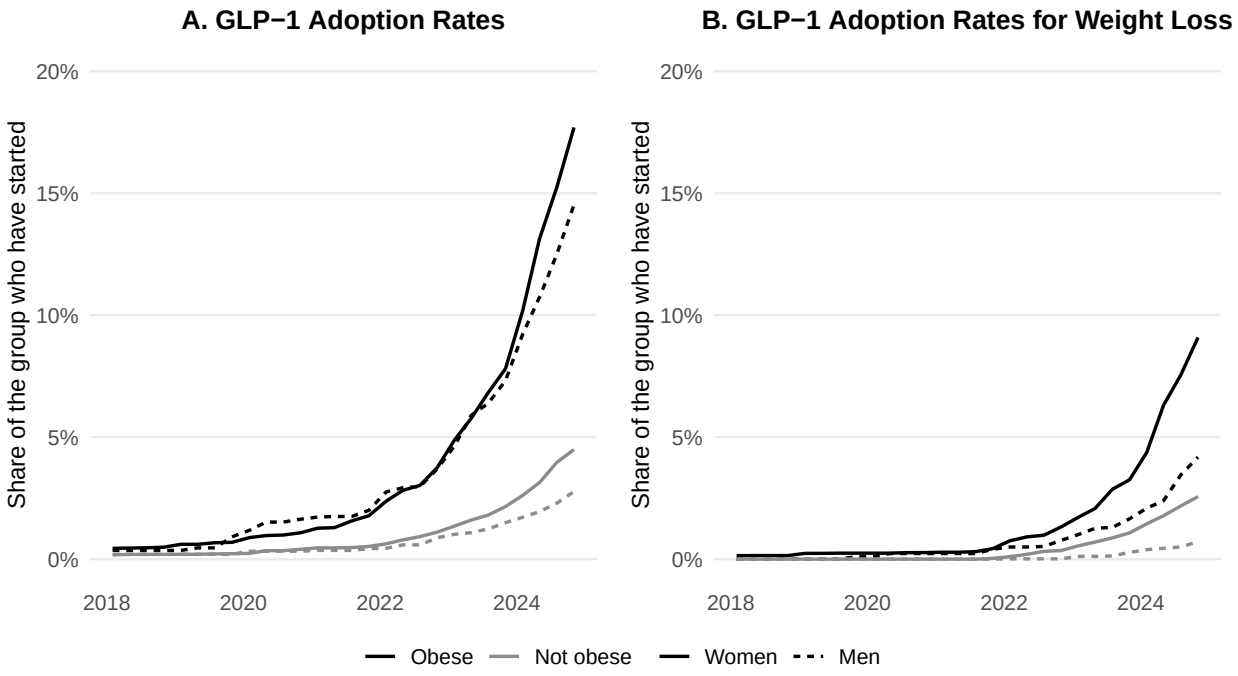


FIGURE 1. GLP-1 ADOPTION, BY SEX AND OBESITY HISTORY

Note: Within-group share of each sex and obesity group who have started a GLP-1, by quarter through the GLP-1 survey (2024Q4), from retrospective first-start dates with survey weights. *Panel A* counts every GLP-1 starter. *Panel B* counts only those who started a GLP-1 to lose weight. "Obese" means ever observed with body-mass index at least 30 in the biennial panel, prior to the GLP-1 start.

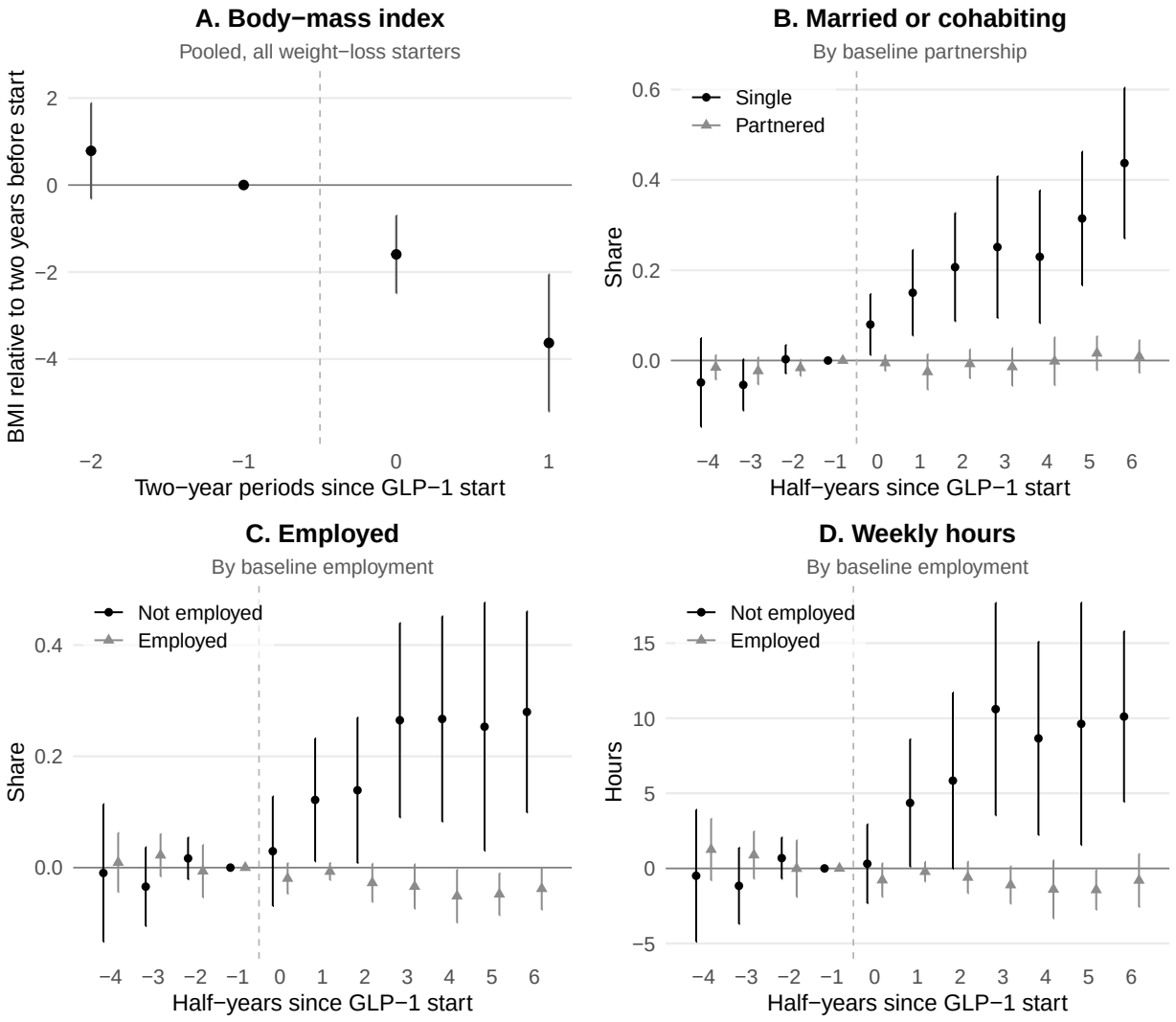


FIGURE 2. WEIGHT LOSS, WORK, AND PARTNERING AROUND GLP-1 INITIATION: WOMEN STARTING FOR WEIGHT LOSS

Note: Stacked matched event studies for women first starting a GLP-1 for weight loss. Controls are later adopters, observed before their own start, plus interested never-users, propensity-score reweighted within cohort × employment × partnership cells. Bars are 95 percent confidence intervals, clustered by person. *Panel A* shows GLP-1 impacts on BMI. *Panels B–D* use the quarterly household panel in half-year event periods, relative to the half-year before start, with stack-person and cohort-by-quarter fixed effects. Panel B (married or cohabiting) splits by baseline partnership. Panels C (employment) and D (weekly hours worked) split by baseline employment status.

TABLE 1—DESCRIPTIVE STATISTICS: WOMEN BY GLP-1 STATUS

	Weight-loss starters	Interested, no take-up	Not interested	p-value, starters vs. interested	p-value, starters vs. not interested
Panel A. Demographics (GLP-1 module interview)					
Age (years)	47.3	45.9	47.1	0.227	0.858
	[12.8]	[15.5]	[17.1]		
Non-Hispanic White	0.65	0.58	0.62	0.106	0.415
	[0.48]	[0.49]	[0.49]		
Non-Hispanic Black	0.09	0.15	0.11	0.016	0.260
	[0.28]	[0.36]	[0.32]		
Hispanic/Latino	0.22	0.18	0.18	0.273	0.263
	[0.41]	[0.38]	[0.38]		
BA or higher	0.35	0.24	0.39	0.005	0.218
	[0.48]	[0.43]	[0.49]		
Employed or on leave	0.66	0.56	0.55	0.013	0.001
	[0.47]	[0.50]	[0.50]		
Panel B. Health and well-being (UAS core, pre-baseline)					
Body-mass index	34.8	34.0	27.9	0.345	<0.001
	[8.1]	[7.9]	[7.0]		
Obese (BMI $\geq$ 30)	0.69	0.66	0.31	0.567	<0.001
	[0.46]	[0.47]	[0.46]		
Diabetes	0.10	0.22	0.10	<0.001	0.943
	[0.29]	[0.41]	[0.30]		
Self-rated health (SD)	-0.07	-0.29	0.07	0.012	0.066
	[0.91]	[0.90]	[0.93]		
Poor or fair health	0.18	0.25	0.15	0.063	0.341
	[0.38]	[0.43]	[0.36]		
Health limits work	0.20	0.21	0.17	0.754	0.406
	[0.40]	[0.41]	[0.38]		
CES-D depression (SD)	0.17	0.51	0.04	0.002	0.144
	[1.07]	[1.16]	[1.02]		
Loneliness (SD)	0.22	0.47	0.07	0.004	0.046
	[0.85]	[1.14]	[1.07]		
Satisfaction with life (SD)	-0.01	-0.31	-0.04	<0.001	0.657
	[0.87]	[1.00]	[0.99]		
Panel C. Economic resources (panel surveys, pre-baseline)					
Household income ($)	$82,750	$59,116	$74,795	<0.001	0.065
	[$53,873]	[$47,494]	[$54,625]		
Own annual earnings ($)	$45,541	$29,939	$37,378	<0.001	0.048
	[$56,579]	[$51,192]	[$64,432]		
Weekly work hours	26.2	21.5	19.8	0.009	<0.001
	[19.3]	[20.4]	[19.8]		
Unemployed	0.06	0.08	0.08	0.402	0.427
	[0.24]	[0.27]	[0.27]		
Disabled	0.10	0.12	0.07	0.644	0.202
	[0.31]	[0.32]	[0.26]		
Retired	0.08	0.10	0.16	0.275	<0.001
	[0.27]	[0.30]	[0.36]		
Married or cohabiting	0.64	0.58	0.58	0.218	0.197
	[0.48]	[0.49]	[0.49]		
Panel D. GLP-1 module					
Currently using (vs. former)	0.53				
	[0.50]				
Months since first start	14.2				
	[21.4]				
Body weight lost (%)	9.8				
	[10.7]				
Pays fully out of pocket	0.40				
	[0.49]				
Monthly out-of-pocket ($)	$299				
	[$245]				
Respondents (positive-weight module)	369	798	3,610		

Note: Survey-weighted means by GLP-1 status in the GLP-1 module (December 2024–February 2025); standard deviations in brackets. Columns are women who started a GLP-1 to lose weight, interested never-users ("would like to take" a GLP-1), and uninterested never-users. The respondent row is the positive-weight module count, before requiring valid first-start timing. Diabetes-management starters and ever-users starting for another reason are excluded throughout. Panels B–C take each respondent's latest pre-baseline panel observation (the pre-start quarter for starters with valid timing, the module interview for never-users and for the few starters without valid timing). Panel D's out-of-pocket cost is among users paying fully out of pocket. The last two columns report p-values for the difference in weighted means between weight-loss starters and each never-user group.

Table 2—Effects of GLP-1 initiation for weight loss: women, by baseline partnership

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all quarters)	-1.423	0.051	0.048	0.029	1.078	-0.027	0.001	0.000
	(0.408)	(0.045)	(0.052)	(0.029)	(1.029)	(0.021)	(0.018)	(0.017)
Post, quarters 6–18	-2.658	0.058	0.037	0.076	2.932	-0.050	0.009	0.005
	(0.578)	(0.055)	(0.068)	(0.043)	(1.536)	(0.030)	(0.024)	(0.022)
Pre-period treated mean	34.6	11.13	10.03	0.66	25.2	0.05	0.98	0.02
Person-waves	7,047	43,608	41,860	43,608	43,584	43,608	43,608	43,604
Panel B. Women single at baseline								
Post GLP-1 start (all quarters)	-2.516	0.110	0.009	-0.009	-0.823	0.008	0.183	
	(0.735)	(0.049)	(0.043)	(0.021)	(0.978)	(0.018)	(0.047)	
Post, quarters 6–18	-4.199	0.201	0.114	-0.012	-0.758	0.020	0.286	
	(0.817)	(0.087)	(0.064)	(0.034)	(1.574)	(0.028)	(0.071)	
Pre-period treated mean	34.9	10.83	10.14	0.82	33.2	0.05	0.01	
Person-waves	3,367	21,237	20,450	21,255	21,236	21,255	21,255	
Test of equal effects, partnered vs. single (p-value)								
Post GLP-1 start (all quarters)	0.368	0.501	0.686	0.446	0.221	0.390	0.000	
Post, quarters 6–18	0.990	0.247	0.394	0.236	0.183	0.329	0.000	

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all waves)	0.045	-0.010	0.101	-0.009	-0.118	0.048	-0.024	-0.032
	(0.042)	(0.021)	(0.052)	(0.082)	(0.078)	(0.056)	(0.050)	(0.054)
Post, quarters 6–18	-0.035	0.046	0.060	-0.016	-0.049	-0.044	-0.014	-0.002
	(0.075)	(0.034)	(0.047)	(0.121)	(0.116)	(0.075)	(0.076)	(0.083)
Pre-period treated mean	-0.07	0.16	0.17	0.21	0.20	-0.11	0.49	0.64
Person-waves	34,305	34,316	6,936	27,323	24,341	34,125	7,171	7,171
Panel D. Women single at baseline								
Post GLP-1 start (all waves)	-0.019	0.054	-0.024	0.070	-0.038	-0.088	-0.280	-0.225
	(0.086)	(0.041)	(0.071)	(0.083)	(0.132)	(0.076)	(0.108)	(0.111)
Post, quarters 6–18	-0.061	0.076	-0.067	0.094	-0.097	-0.055	-0.354	-0.289
	(0.113)	(0.058)	(0.039)	(0.153)	(0.164)	(0.152)	(0.132)	(0.141)
Pre-period treated mean	-0.27	0.21	0.15	0.07	0.20	-0.12	0.36	0.41
Person-waves	17,282	17,307	3,333	13,944	12,412	17,291	3,439	3,439
Test of equal effects, partnered vs. single (p-value)								
Post GLP-1 start (all waves)	0.834	0.555	0.331	0.531	0.754	0.154	0.060	0.243
Post, quarters 6–18	0.823	0.861	0.030	0.656	0.577	0.720	0.012	0.134

Note: Stacked matched difference-in-differences estimates for women starting a GLP-1 for weight loss, by baseline partnership. Controls come from later adopters, observed before their own start, plus interested never-users, exactly matched on employment × partnership and propensity-reweighted on race, lagged BMI, diabetes, self-rated health, depression, whether health limits work, life satisfaction, and household income. Outcomes run through 2026Q2 using the interim verified-control filter described in the text. Standard errors in parentheses, clustered by person.

Table 3—Effects of GLP-1 initiation for weight loss: women, by baseline employment

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women employed at baseline								
Post GLP-1 start (all quarters)	-1.685	0.068	0.017	-0.026	-1.048	0.006	0.078	0.001
	(0.532)	(0.029)	(0.033)	(0.015)	(0.605)	(0.007)	(0.033)	(0.025)
Post, quarters 6–18	-3.702	0.105	0.035	-0.044	-1.512	0.013	0.137	0.003
	(0.620)	(0.052)	(0.050)	(0.022)	(0.815)	(0.011)	(0.052)	(0.028)
Pre-period treated mean	34.3	11.33	10.38	0.97	37.8	0.02	0.61	0.02
Person-waves	7,946	49,653	47,929	49,653	49,650	49,653	49,653	34,335
Panel B. Women not employed at baseline								
Post GLP-1 start (all quarters)	-1.453	0.121	0.138	0.132	4.697	-0.082	0.017	0.001
	(0.766)	(0.094)	(0.091)	(0.056)	(2.096)	(0.050)	(0.012)	(0.011)
Post, quarters 6–18	-3.458	0.180	0.216	0.269	9.866	-0.161	-0.002	0.023
	(1.397)	(0.112)	(0.109)	(0.084)	(3.127)	(0.070)	(0.030)	(0.031)
Pre-period treated mean	36.2	10.22	9.22	0.05	1.5	0.14	0.80	0.00
Person-waves	2,050	12,642	11,917	12,660	12,624	12,660	12,660	6,944
Test of equal effects, employed vs. not employed (p-value)								
Post GLP-1 start (all quarters)	0.886	0.658	0.438	0.085	0.104	0.255	0.020	0.655
Post, quarters 6–18	0.947	0.737	0.551	0.022	0.035	0.169	0.009	0.415

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Biennial health, well-being, and job-mobility outcomes, women employed at baseline								
Post GLP-1 start (all waves)	0.016	0.018	0.030	0.097	-0.085	-0.035	-0.125	-0.098
	(0.054)	(0.023)	(0.062)	(0.065)	(0.078)	(0.054)	(0.065)	(0.068)
Post, quarters 6–18	-0.017	0.042	0.032	0.112	-0.132	-0.064	-0.180	-0.101
	(0.083)	(0.036)	(0.037)	(0.114)	(0.115)	(0.091)	(0.083)	(0.092)
Pre-period treated mean	0.02	0.10	0.10	0.08	0.20	-0.07	0.42	0.57
Person-waves	38,200	38,215	7,836	31,026	27,535	37,962	8,030	8,030
Panel D. Biennial health, well-being, and job-mobility outcomes, women not employed at baseline								
Post GLP-1 start (all waves)	0.030	0.001	0.045	-0.128	0.067	0.090	0.087	0.059
	(0.056)	(0.037)	(0.061)	(0.132)	(0.167)	(0.096)	(0.087)	(0.094)
Post, quarters 6–18	-0.184	0.109	-0.142	0.018	0.278	-0.165	0.049	0.040
	(0.099)	(0.054)	(0.106)	(0.171)	(0.194)	(0.133)	(0.223)	(0.224)
Pre-period treated mean	-0.50	0.34	0.36	0.44	0.20	-0.20	0.52	0.55
Person-waves	11,210	11,231	2,016	8,778	7,900	11,248	2,156	2,156
Test of equal effects, employed vs. not employed (p-value)								
Post GLP-1 start (all waves)	0.576	0.537	0.583	0.230	0.427	0.354	0.340	0.746
Post, quarters 6–18	0.535	0.645	0.346	0.561	0.216	0.898	0.319	0.875

Note: Stacked matched difference-in-differences estimates for women starting a GLP-1 for weight loss, by baseline employment. Controls come from later adopters, observed before their own start, plus interested never-users, exactly matched on employment × partnership and propensity-reweighted on race, lagged BMI, diabetes, self-rated health, depression, whether health limits work, life satisfaction, and household income. Outcomes run through 2026Q2 using the interim verified-control filter described in the text. Standard errors in parentheses, clustered by person.

Supplemental Appendix

GLP-1–Induced Weight Loss and the Female Obesity Penalty

Rebecca Diamond

June 19, 2026

Contents

A Sample construction and the GLP-1 module **1**

A.1 Constructing the outcome variables 2

A.2 Re-measuring control status after the module 5

B Design diagnostics **5**

B.1 Pre-trends 5

B.2 Rambachan–Roth sensitivity bounds 6

B.3 Balance 8

C Robustness **11**

C.1 Alternative control groups 11

C.2 Randomization inference 18

D Additional outcomes and figures **18**

E GLP-1 Effects on Men **22**

F Validating the design against the GLP-1 weight-loss trials **25**

A Sample construction and the GLP-1 module

The Understanding America Study GLP-1 module was fielded December 18, 2024 through February 9, 2025 to 10,085 panel members; 9,758 respondents with positive post-stratification weights answered the GLP-1 awareness question and constitute the positive-weight module sample used in descriptive tables. Respondents who have heard of GLP-1 medications report use status (currently taking; used to take; never but would like to; never, no interest), the month and year they *first* started any GLP-1, and, when the exact date is not recalled, an estimate of how many months earlier they started. The main reason for starting distinguishes weight loss, diabetes management, and other. Payment route, monthly copays, and monthly out-of-pocket costs are also asked. Table A1 reports starter counts first by start reason in the positive-weight module sample, and then shows the subset with valid first-start timing.

Treatment timing uses the exact start month when reported (80 percent of ever-users) and otherwise the months-ago estimate. Because all dates are recalled retrospectively, the analysis

event clock assigns treatment to the quarter before the reported start, allowing one quarter of recall error; undoing the allowance, so event 0 falls in the quarter of the reported start, leaves the estimates essentially unchanged. Treatment is intent-to-treat at *first initiation*: former users remain treated from their start date. I restrict all analysis to users who start a GLP-1 in 2021Q4 or later (dropping 13 women) and all pre-period analysis to 2021Q1 and later to remove the most extreme periods of COVID. The main difference-in-differences analysis restricts the sample to adults ages 25–61 at baseline to avoid changes associated with Medicare eligibility, since Medicare does not cover GLP-1s for weight loss. Finally, I restrict the sample to GLP-1 users who are in the quarterly My Household panel at least once in the 4 quarters prior to starting a GLP-1. The USC UAS has turnover in panel participation, and not all GLP-1 users were part of the USC UAS survey the year before starting a GLP-1. Table A2 documents the path from the 369 positive-weight female weight-loss starters, to 367 with valid first-start timing, to the 242-women benchmark.

Table A1: GLP-1 use in the GLP-1 survey module

	Women	Men
UAS672 positive-weight module sample	5,922	3,836
Ever used a GLP-1	715	315
start reason: weight loss	369	90
start reason: diabetes management	319	211
start reason: other/unknown	27	14
with valid first-start timing (ITT treated)	692	302
weight loss, valid timing	367	88
diabetes management, valid timing	302	202
other/unknown, valid timing	23	12
Never used, would like to (interested never-users)	798	346

Notes: Counts are unweighted. Start-reason rows immediately below "Ever used a GLP-1" are positive-weight module counts before imposing valid first-start timing. Valid-timing rows are the subset eligible for event-time treatment definitions.

Table A2: Sample funnel: from module respondents to the benchmark

Step	Women
Positive-weight female weight-loss starters (module)	369
Valid first-start timing	367
Start quarter 2021Q4 or later (panel window)	354
My Household baseline interview within 4 quarters of start	298
Baseline age 25-61	242
Benchmark analysis sample (common support, outcome observed)	242

A.1 Constructing the outcome variables

The outcomes come from a quarterly household panel, a biennial panel, and two monthly surveys. The quarterly My Household Over Time panel supplies labor-force status, weekly hours,

Table A3: How weight-loss starters obtain and pay for GLP-1s

	Women	Men
Weight-loss starters (positive-weight module)	369	90
Insurance, no copay	0.21	0.20
Insurance with copay	0.32	0.34
median copay ($/month)	25	25
Not covered, out of pocket	0.41	0.43
median out-of-pocket ($/month)	275	300

Notes: Shares are among positive-weight GLP-1 module respondents who report starting for weight loss, so the denominators match the descriptive module counts rather than the valid-timing subset used for event-time analyses. The remainder report "other" arrangements (samples, trials, foreign purchase). Medians are conditional on the payment route.

a sixteen-bracket household income measure, and marital and cohabitation status. The biennial Comprehensive Panel Dataset, which fields a Health and Retirement Study instrument, supplies self-reported height and weight, own annual earnings, self-rated health, a depression screener, a loneliness item, and whether health limits work. Two monthly surveys repeat several of the health and well-being items at higher frequency: a general panel fielded since 2023, and an older-member event panel that runs from 2019 through 2023.

Weight. Body mass index is self-reported. Respondents give their height in feet and inches and their weight in pounds, which I convert to meters and kilograms and combine as weight divided by height squared.

Labor-force status. The My Household panel asks each respondent for a single labor-force status: currently working, on sick or other leave, unemployed and on layoff, unemployed and looking, retired, disabled, or another status. From this one item I build an employment indicator (currently working or on paid leave), an unemployment indicator (on layoff or actively looking), and a retirement indicator. Disability is the panel's separate yes-or-no "disabled" flag rather than the residual status category, so a woman can report a disability while still attached to the labor force.

Weekly hours. Hours come from the panel's "hours of work per week." I measure them unconditionally: a woman who is not employed is assigned zero hours. I cap reported hours at 80 to limit the pull of implausible values.

Household income. The panel records household income in sixteen brackets, from under $5,000 to $150,000 or more. I assign each bracket its midpoint, set the open top bracket to $175,000, and take logs; the estimates are not sensitive to alternative assignments for that top bracket. Household income per member divides the midpoint by household size.

Own earnings. Annual labor earnings come from the HRS income module. Earnings sum five reported components, wages and salary, income from a professional practice or trade, tips, bonuses and commissions, earnings from a second job or the military, and self-employment income. Earnings refer to the calendar year before the interview and are dated to that reference year. I keep zeros and winsorize at the 99th percentile of the distribution.

Partnership. A woman is coded as partnered if she reports being married, whether her spouse lives with her or elsewhere, or reports living with a partner. The women who are separated, divorced, widowed, or never married and not cohabiting are coded as not partnered.

Combining the biennial and monthly well-being scales. Four outcomes, self-rated health, depression, loneliness, and life satisfaction, are measured by a biennial item for the early part of the sample and by a monthly survey starting in 2023. I put the different survey instruments on one scale using respondents who answer both instruments within about three months of each other: I line up each linking instrument to the reference instrument by matching their means and standard deviations over that overlap, then map the linking readings onto the reference scale. The one binary item, loneliness, is calibrated instead by its two conditional means. Each harmonized variable is then standardized once, so its coefficients are read in standard-deviation units.

Self-rated health. The survey asks whether the respondent would rate their health as excellent, very good, good, fair, or poor. I reverse the scale so higher values are healthier and combine the biennial and monthly readings as above. I also report a plain indicator for fair or poor health, built as a level from the same question and left unstandardized.

Health limits work. The jobs module asks whether the respondent has an impairment or health problem that limits the kind or amount of paid work they can do. I use the yes-or-no answer directly.

Depression. The biennial source is the eight-item CES-D screener, which counts how many of eight feelings the respondent reported much of the time in the past week, scoring the two positively worded items (felt happy, enjoyed life) in reverse, for a zero-to-eight scale. The monthly source is the two depression stems of the PHQ-4, "feeling down, depressed, or hopeless" and "little interest or pleasure in doing things," each rated one to four and summed. Higher values are more depressed. I equate and standardize the two as above.

Loneliness. The biennial source is the CES-D "felt lonely" item, a yes-or-no report of feeling lonely much of the time in the past week. The monthly source is the single BRFSS question on how often the respondent feels lonely, from always to never, which I reverse so higher values are lonelier and treat as the reference scale. The binary biennial item is mapped onto it by its conditional means before the combined series is standardized.

Life satisfaction and happiness. Life satisfaction comes from a zero-to-ten "how satisfied are you with your life" item in the biennial well-being module, combined with the monthly "overall life satisfaction" question. I align both measures in the same direction and standardize the combined series. Happiness is the parallel zero-to-ten "how happy are you" item. It has no monthly counterpart, so it stays on its single biennial source; I report its effect in units of its baseline standard deviation (about two points on the ten-point scale) so it sits on the same footing as the other scales.

Job mobility. Two outcomes count cross-wave mobility since a woman entered the panel. The employer-change count increments when a woman is employed in consecutive waves and reports that she no longer works for the same employer as at the prior interview. It also increments when she crosses the employment boundary, either leaving a job for non-employment or starting a job

from non-employment. It does not increment for a same-employer change in the kind of work she does, and it does not increment when she stays with the same employer or remains out of work. The employer-or-position-change count is the broader measure: it includes every employer-change event just described and also increments for a same-employer change in kind of work or position. I treat non-employment as a single state, so a move across the employment boundary counts in both measures, while remaining non-employed does not. Each outcome is the running total since panel entry. Because the stacked regressions carry a person fixed effect, the post-period coefficients measure the changes accumulated since the matching baseline.

A.2 Re-measuring control status after the module

The GLP-1 module is a single snapshot: it records each control's use status once, at the December 2024 to February 2025 interview. The outcomes run through 2026Q2. A woman who reports no GLP-1 use in early 2025 may start one by 2026, so her later quarters are not clean control observations until I can verify that she did not. I obtain that verification from the May 2026 UAS792 follow-up and from interim GLP-1 tracking waves collected between the original module and late 2025. If UAS792 re-confirms never-use in May 2026, I keep that control's quarters through 2026Q2. If the interim tracking-wave record still shows never-use, I keep that control's quarters through 2025Q3; the earliest start revealed by the interim records is 2025Q4, so those quarters are verified clean. Treated quarters are always kept.

B Design diagnostics

B.1 Pre-trends

Table A4 reports the pre-trend test inside the exact designs of main-text Tables 2 and 3. For every outcome and every baseline subgroup those tables report, it gives the differential linear pre-slope: the coefficient on treated $\times$ linear event time across the pre-period, in half-yearly bins with person and cohort-quarter fixed effects, estimated on the same matched subgroup stacks and event clock as the outcome it certifies. This smooth slope is the pre-trend that matters for a difference-in-differences design, because it is the trend that would extrapolate across the treatment date into a post-period bias.

The slopes have small magnitudes, and confidence intervals contain zero for all key findings. Of the sixty-three slopes estimated across the grid, one rejects at five percent, against about three expected by chance: self-rated health among baseline-not-employed women (+0.06 standard deviations per half-year, $p = 0.04$). The pre-trend tests here strongly support a lack of pre-treatment trends.

Table A4: Differential linear pre-slope tests, by baseline subgroup

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Baseline partnership (main-text Table 2)								
Partnered	-0.2933	-0.0150	-0.0160	0.0086	0.0352	0.0028	0.0026	-0.0028
	(0.1680)	(0.0176)	(0.0166)	(0.0086)	(0.2902)	(0.0054)	(0.0034)	(0.0035)
Single	0.0842	-0.0012	0.0055	0.0008	-0.0578	-0.0044	0.0181	
	(0.2528)	(0.0175)	(0.0184)	(0.0130)	(0.5790)	(0.0114)	(0.0167)	
Baseline employment (main-text Table 3)								
Employed	-0.2018	-0.0043	0.0033	-0.0002	-0.2780	-0.0034	0.0118	-0.0066
	(0.1379)	(0.0116)	(0.0141)	(0.0083)	(0.3327)	(0.0061)	(0.0098)	(0.0053)
Not employed	0.1554	-0.0268	-0.0339	0.0174	0.6547	0.0145	-0.0042	0.0069
	(0.3041)	(0.0332)	(0.0271)	(0.0181)	(0.6264)	(0.0111)	(0.0036)	(0.0042)

	Self-rated health	Poor/fair health	Health limits work	Depression	Lonely	Life satisf.	Employer changes	Job changes
Baseline partnership (main-text Table 2)								
Partnered	-0.0013	-0.0044	0.0109	0.0688	0.0295	-0.0136	-0.0143	-0.0061
	(0.0165)	(0.0097)	(0.0137)	(0.0426)	(0.0390)	(0.0295)	(0.0259)	(0.0267)
Single	0.0225	-0.0015	0.0139	0.0032	0.0787	0.0752	0.0059	0.0305
	(0.0312)	(0.0180)	(0.0114)	(0.0535)	(0.0473)	(0.0514)	(0.0403)	(0.0404)
Baseline employment (main-text Table 3)								
Employed	-0.0149	-0.0017	0.0057	0.0569	0.0640	-0.0251	-0.0371	-0.0308
	(0.0183)	(0.0084)	(0.0072)	(0.0318)	(0.0339)	(0.0240)	(0.0205)	(0.0239)
Not employed	0.0637	-0.0230	0.0481	0.0726	-0.0164	0.0722	0.0133	0.0133
	(0.0302)	(0.0177)	(0.0403)	(0.0697)	(0.0562)	(0.0498)	(0.0464)	(0.0464)

Notes: Each cell reports the coefficient on treated × linear event time, estimated on all pre-treatment half-years with person and cohort-quarter fixed effects; standard errors are clustered by person and reported in parentheses. The slope is in outcome units per half-year. The relationship-dissolution outcome is defined for women partnered at baseline; the baseline-single cell is therefore blank.

B.2 Rambachan–Roth sensitivity bounds

The pre-trend tests in Table A4 are useful diagnostics, but they do not by themselves prove parallel trends. A pre-trend test can fail to reject because the pre-period estimates are noisy, and the relevant identifying threat is a post-treatment difference in untreated outcome trends. I therefore complement the pre-trend tests with the sensitivity analysis of Rambachan and Roth (2023), which asks how much differential post-treatment drift the main estimates can tolerate.

I implement the exercise using the same matched event-study specification as in main-text Tables 2 and 3. The treatment effect of interest is the average longer-run effect over quarters 6–18 after GLP-1 initiation. The pre-treatment event-study coefficients provide the scale for plausible violations of parallel trends, while the post-treatment coefficients are allowed to combine the treatment effect with a possible counterfactual drift between treated women and their matched controls.

My primary sensitivity exercise uses the relative-magnitude class, $\Delta^{RM}(\bar{M})$. In this class, $\bar{M}$ indexes how large the post-treatment violation of parallel trends may be relative to the largest pre-treatment deviation observed in the event study. Thus $\bar{M} = 0$ corresponds to the usual no-post-treatment-violation case, while $\bar{M} = 0.25$, 0.5, and 0.75 allow post-treatment trend violations that are one-quarter, one-half, and three-quarters as large as the largest supported pre-treatment deviation.

Table A5 reports the resulting 95 percent confidence sets for the longer-run headline effects. I focus on the cells where the main tables show economically important changes: partnership among women single at baseline, and employment, weekly hours, and unemployment among women not employed at baseline. For outcomes whose main estimates are already close to zero, the same exercise would mainly widen confidence sets around null effects.

The relative-magnitude bounds support the direction of the main extensive-margin results under moderate departures from parallel trends. For baseline-single women, the quarters 6–18 partnership effect remains positive at $\bar{M} = 0.25$, with confidence set $[0.16, 0.76]$. At $\bar{M} = 0.5$, the lower bound rounds to zero, $[0.00, 0.93]$, and by $\bar{M} = 0.75$ the interval clearly includes zero. For baseline-not-employed women, the employment effect is more robust: the confidence set is $[0.21, 0.65]$ at $\bar{M} = 0.25$ and remains positive at $\bar{M} = 0.5$, $[0.11, 0.76]$, before becoming borderline at $\bar{M} = 0.75$. Weekly hours remain positive at $\bar{M} = 0.25$, $[1.5, 18.2]$, but include zero at larger values of $\bar{M}$. The unemployment decline has the expected negative sign, but its relative-magnitude interval already just includes zero at $\bar{M} = 0.25$, $[-0.46, 0.02]$.

Panel B reports a complementary smoothness restriction. The $M = 0$ smoothness case does not impose exact parallel trends; it allows the treated-control difference in untreated outcomes to continue on a linear path through the treatment date, while ruling out curvature in that differential trend. Under this restriction, the partnership effect for baseline-single women remains positive $[0.08, 0.52]$, the employment effect for baseline-not-employed women remains positive $[0.10, 0.57]$, and the unemployment effect remains negative $[-0.50, -0.05]$. Weekly hours, however, include zero under the smoothness check.

I therefore interpret the sensitivity analysis as follows. The partnership and employment-entry results do not rely on exact parallel trends and survive moderate violations calibrated to the observed pre-period deviations. The evidence is not robust to very large violations of parallel trends, and the hours result is less robust than the employment indicator itself. Accordingly, I place the most weight on the extensive-margin findings: GLP-1 initiation is followed by increased partnership formation among women single at baseline and increased employment among women not employed at baseline.

Table A5: Rambachan–Roth sensitivity for headline longer-run effects

Panel A. Relative magnitudes

Sample	Outcome	Pre bins	$\bar{M} = 0$	$\bar{M} = 0.25$	$\bar{M} = 0.5$	$\bar{M} = 0.75$
Single baseline	Partnered	-4,-3,-2	[0.26, 0.63]	[0.16, 0.76]	[0.00, 0.93]	[-0.17, 1.11]
Not-employed baseline	Employed	-4,-3,-2	[0.27, 0.56]	[0.21, 0.65]	[0.11, 0.76]	[0.00, 0.89]
Not-employed baseline	Weekly hours	-4,-3,-2	[3.8, 15.6]	[1.5, 18.2]	[-1.6, 21.4]	[-5.5, 25.3]
Not-employed baseline	Unemployed	-4,-3,-2	[-0.33, -0.09]	[-0.46, 0.02]	[-0.63, 0.18]	[-0.81, 0.36]

Panel B. Smoothness

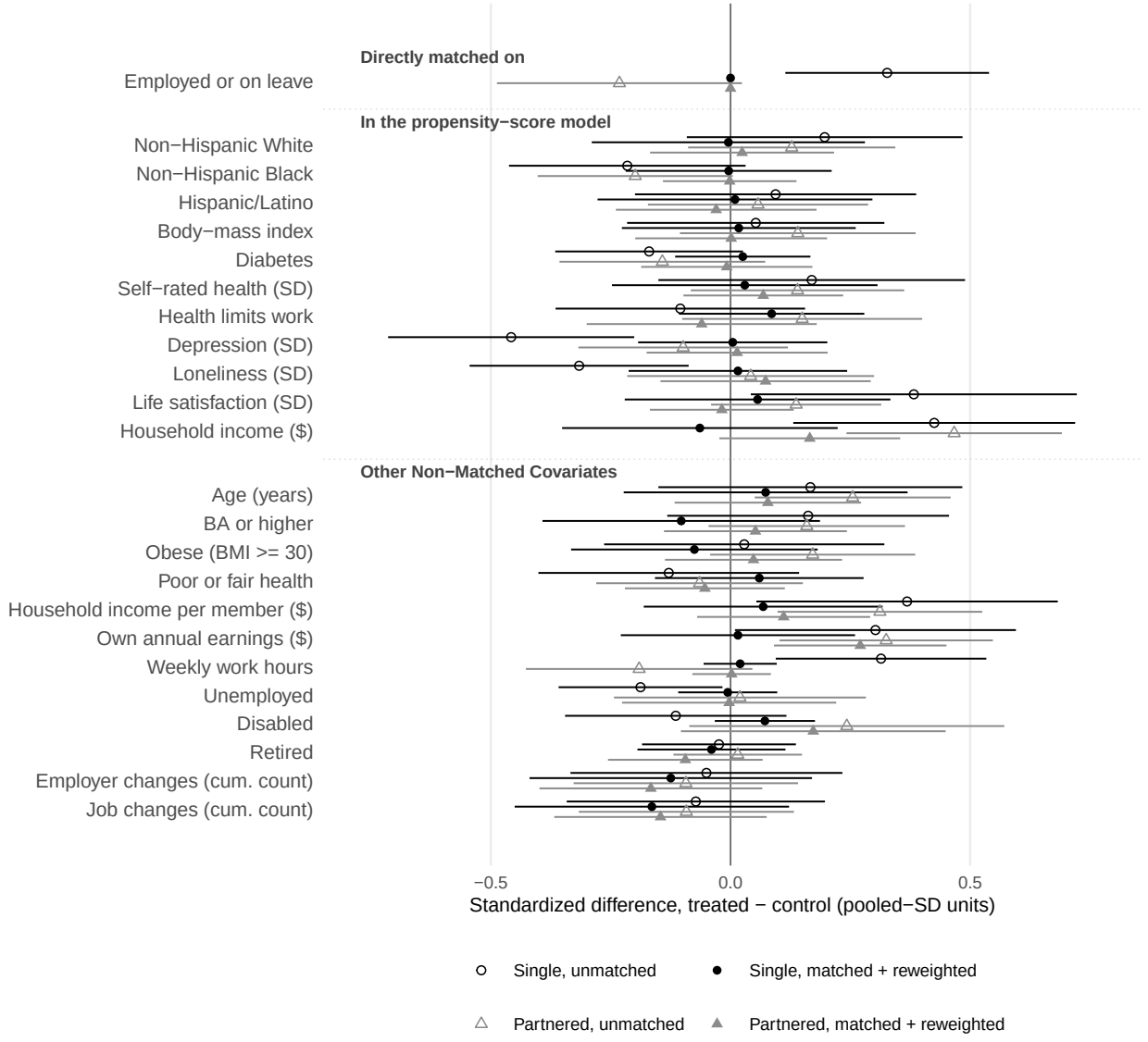
Sample	Outcome	Pre bins	Smoothness $M = 0$
Single baseline	Partnered	-4,-3,-2	[0.08, 0.52]
Not-employed baseline	Employed	-4,-3,-2	[0.10, 0.57]
Not-employed baseline	Weekly hours	-4,-3,-2	[-1.6, 16.5]
Not-employed baseline	Unemployed	-4,-3,-2	[-0.50, -0.05]

Notes: Each cell reports a 95 percent confidence set for the quarters 6–18 effect from Rambachan and Roth (2023), using the same matched subgroup stack and event clock as main-text Tables 2 and 3. Panel A reports relative-magnitudes bounds; $\bar{M}$ scales the largest allowed post-period violation of parallel trends by the largest supported pre-period violation. Panel B reports the smoothness class with $M = 0$, which permits a linear differential trend but no curvature in the post-period violation. Pre bins are the event-study bins used for calibration after imposing the support rule that at least 25 treated women contribute to the bin.

B.3 Balance

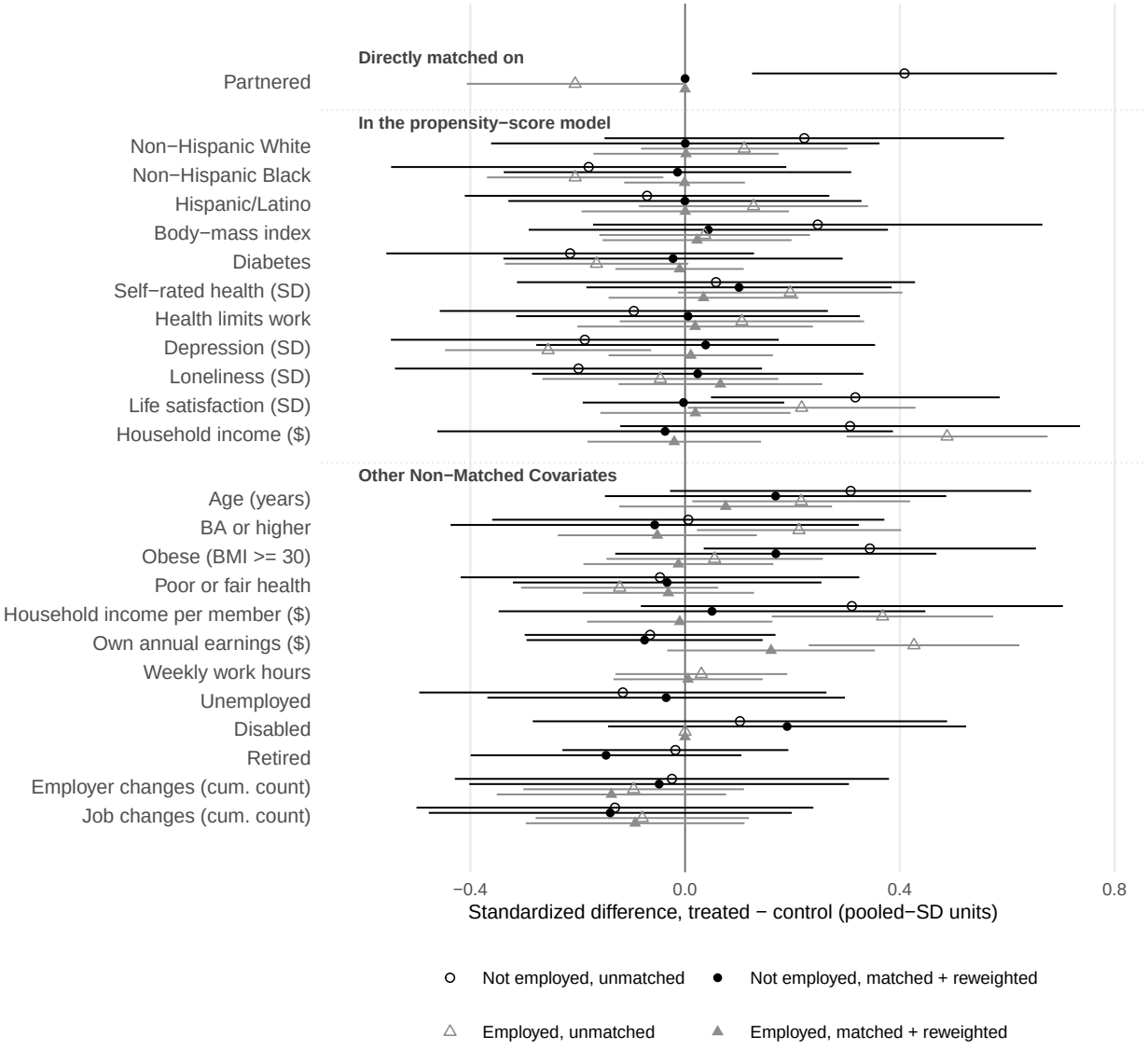
Figures A1 and A2 report covariate balance, each cut by baseline partnership status and employment, respectively. The exact-match constructs (baseline employment and baseline partnership) are balanced identically by construction. Every covariate in the propensity model balances tightly after reweighting and the direct balance tests cannot reject zero imbalance across covariates. Covariates outside the propensity model are also balanced: the direct tests cannot reject zero differences between treatment and matched controls.

Figure A1: Covariate balance: raw versus matched and reweighted, by baseline partnership



Notes: Each marker is the weighted treated−control difference in the listed covariate divided by the pooled survey-weighted standard deviation of the same partnership sample. Bars are 95 percent confidence intervals. Open markers are raw cross-sectional comparisons under survey weights. Filled markers apply exact matching on employment × partnership and propensity-score odds reweighting, and their regressions absorb cohort × exact-match-cell fixed effects.

Figure A2: Covariate balance: raw versus matched and reweighted, by baseline employment



Notes: Each marker is the weighted treated−control difference in the listed covariate divided by the pooled survey-weighted standard deviation of the same employment sample. Bars are 95 percent confidence intervals. Open markers are raw cross-sectional comparisons under survey weights. Filled markers apply exact matching on employment × partnership and propensity-score odds reweighting, and their regressions absorb cohort × exact-match-cell fixed effects.

C Robustness

C.1 Alternative control groups

The benchmark control pool is women interested in GLP-1s who have not taken them, as well as late adopters, prior to adoption, reweighted to the treated women's observables. As robustness, I try using the population who report no interest in using a GLP-1 as the control group pool, then apply the same matching strategy as in the main paper. I also try pooling the uninterested never-user group with the interested never-users, removing the late adopters prior to treatment from the control group, and restricting the donor pool to interested never-users only. Tables A6–A7 report the all-non-users pool, Tables A8–A9 report the interested-never-user-only pool, and the following displays report the uninterested-never-user pool. All findings from the main paper hold up, and some get statistically more powered with the larger sample control group.

Table A6: Outcomes by baseline partnership under the all-non-users control group pool: interested + uninterested never-users

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all quarters)	-1.091	0.058	0.076	0.035	1.977	-0.023	-0.013	0.015
	(0.317)	(0.043)	(0.049)	(0.028)	(1.017)	(0.024)	(0.018)	(0.018)
Post, quarters 6–18	-2.319	0.069	0.066	0.073	3.646	-0.028	-0.018	0.026
	(0.437)	(0.049)	(0.056)	(0.039)	(1.493)	(0.037)	(0.020)	(0.021)
Pre-period treated mean	34.6	11.13	10.03	0.66	25.2	0.05	0.98	0.02
Person-waves	28,369	179,363	172,778	179,728	179,578	179,728	179,728	179,663
Panel B. Women single at baseline								
Post GLP-1 start (all quarters)	-1.975	0.113	0.023	-0.006	-0.184	0.006	0.184	
	(0.529)	(0.057)	(0.047)	(0.017)	(0.793)	(0.014)	(0.047)	
Post, quarters 6–18	-2.773	0.211	0.108	-0.007	0.217	0.014	0.277	
	(0.580)	(0.092)	(0.072)	(0.025)	(1.175)	(0.020)	(0.069)	
Pre-period treated mean	34.9	10.83	10.14	0.82	33.2	0.05	0.01	
Person-waves	13,450	85,868	80,898	85,919	85,840	85,919	85,919	

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all waves)	0.104	-0.072	0.072	0.004	-0.063	0.012	-0.036	-0.015
	(0.057)	(0.048)	(0.042)	(0.070)	(0.074)	(0.046)	(0.038)	(0.041)
Post, quarters 6–18	0.086	-0.088	0.027	0.021	-0.051	-0.073	0.010	0.029
	(0.112)	(0.102)	(0.033)	(0.090)	(0.106)	(0.061)	(0.049)	(0.054)
Pre-period treated mean	-0.07	0.16	0.17	0.21	0.20	-0.11	0.49	0.64
Person-waves	150,153	150,164	27,882	124,652	112,337	149,944	29,034	29,034
Panel D. Women single at baseline								
Post GLP-1 start (all waves)	-0.019	0.039	-0.008	0.135	-0.019	-0.056	-0.084	-0.068
	(0.076)	(0.042)	(0.055)	(0.063)	(0.113)	(0.065)	(0.055)	(0.054)
Post, quarters 6–18	-0.063	0.049	-0.037	0.140	-0.111	-0.068	-0.159	-0.094
	(0.093)	(0.062)	(0.041)	(0.113)	(0.133)	(0.129)	(0.059)	(0.064)
Pre-period treated mean	-0.27	0.21	0.15	0.07	0.20	-0.12	0.36	0.41
Person-waves	72,923	72,948	13,205	60,046	54,139	72,942	13,797	13,797

Notes: Control group widened to all non-users (interested and uninterested never-users), re-matched within the same exact cells (employed × partnered) with the propensity score (race, lagged BMI, diabetes, self-rated health, CES-D, health-limits-work, life satisfaction, household income) re-estimated separately by baseline partnership. The dependent variables, measurement sources, and transformations are the same as in the paper's Tables 2 and 3. Standard errors in parentheses, clustered by person.

Table A7: Outcomes by baseline employment status under the all-non-users robustness pool: interested + uninterested never-users

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women employed at baseline								
Post GLP-1 start (all quarters)	-0.937	0.057	0.029	-0.021	-0.308	0.008	0.078	0.019
	(0.506)	(0.031)	(0.035)	(0.015)	(0.569)	(0.008)	(0.033)	(0.026)
Post, quarters 6–18	-2.276	0.095	0.047	-0.030	-0.237	0.012	0.117	0.033
	(0.495)	(0.052)	(0.047)	(0.021)	(0.795)	(0.011)	(0.046)	(0.029)
Pre-period treated mean	34.3	11.33	10.38	0.97	37.8	0.02	0.61	0.02
Person-waves	32,613	208,344	199,704	208,643	208,569	208,643	208,643	139,807
Panel B. Women not employed at baseline								
Post GLP-1 start (all quarters)	-1.242	0.140	0.142	0.129	5.044	-0.080	0.014	-0.005
	(0.372)	(0.090)	(0.087)	(0.054)	(2.092)	(0.054)	(0.010)	(0.008)
Post, quarters 6–18	-1.737	0.177	0.159	0.225	8.990	-0.096	0.019	-0.011
	(0.716)	(0.094)	(0.098)	(0.076)	(3.009)	(0.083)	(0.027)	(0.028)
Pre-period treated mean	36.2	10.22	9.22	0.05	1.5	0.14	0.80	0.00
Person-waves	7,673	47,379	44,820	47,477	47,361	47,477	47,477	29,164

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Biennial health, well-being, and job-mobility outcomes, women employed at baseline								
Post GLP-1 start (all waves)	0.013	0.007	0.055	0.137	0.009	-0.024	-0.068	-0.030
	(0.047)	(0.023)	(0.053)	(0.058)	(0.067)	(0.049)	(0.041)	(0.044)
Post, quarters 6–18	-0.037	0.030	0.055	0.137	-0.035	-0.063	-0.101	-0.035
	(0.063)	(0.037)	(0.034)	(0.090)	(0.093)	(0.082)	(0.048)	(0.051)
Pre-period treated mean	0.02	0.10	0.10	0.08	0.20	-0.07	0.42	0.57
Person-waves	170,931	170,946	32,364	144,081	129,786	170,681	33,397	33,397
Panel D. Biennial health, well-being, and job-mobility outcomes, women not employed at baseline								
Post GLP-1 start (all waves)	0.041	-0.030	0.021	-0.105	-0.007	0.043	0.074	0.064
	(0.058)	(0.043)	(0.053)	(0.106)	(0.133)	(0.077)	(0.063)	(0.064)
Post, quarters 6–18	-0.059	0.022	-0.160	-0.008	0.127	-0.142	0.216	0.212
	(0.110)	(0.088)	(0.069)	(0.125)	(0.153)	(0.103)	(0.109)	(0.109)
Pre-period treated mean	-0.50	0.34	0.36	0.44	0.20	-0.20	0.52	0.55
Person-waves	43,462	43,487	7,275	34,397	31,070	43,479	7,913	7,913

Notes: The paper's baseline-employment split for women weight-loss starters, with the donor pool widened to all non-users (interested and uninterested never-users). The table uses the same dependent variables, measurement sources, transformations, matching, event windows, post-interview control filter, and reporting conventions as Table A6, but splits treated women by baseline employment rather than baseline partnership. Standard errors in parentheses, clustered by person.

Table A8: Outcomes by baseline partnership under the interested-never-user control group pool

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all quarters)	-1.505	0.051	0.043	0.034	1.386	-0.027	0.000	0.001
	(0.437)	(0.046)	(0.056)	(0.030)	(1.098)	(0.023)	(0.018)	(0.018)
Post, quarters 6–18	-2.819	0.063	0.034	0.077	3.079	-0.049	0.010	0.004
	(0.598)	(0.056)	(0.072)	(0.044)	(1.576)	(0.032)	(0.024)	(0.022)
Pre-period treated mean	34.6	11.13	10.03	0.66	25.2	0.05	0.98	0.02
Person-waves	5,961	36,621	34,953	36,621	36,597	36,621	36,621	36,617
Panel B. Women single at baseline								
Post GLP-1 start (all quarters)	-2.760	0.121	0.021	-0.004	-0.686	0.006	0.213	
	(0.781)	(0.058)	(0.049)	(0.023)	(1.079)	(0.020)	(0.053)	
Post, quarters 6–18	-4.152	0.218	0.121	-0.004	-0.524	0.017	0.326	
	(0.885)	(0.098)	(0.073)	(0.038)	(1.723)	(0.031)	(0.076)	
Pre-period treated mean	34.9	10.83	10.14	0.82	33.2	0.05	0.01	
Person-waves	2,915	18,505	17,827	18,523	18,504	18,523	18,523	

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all waves)	0.056	-0.019	0.115	-0.010	-0.126	0.056	-0.023	-0.026
	(0.045)	(0.023)	(0.055)	(0.089)	(0.081)	(0.059)	(0.051)	(0.056)
Post, quarters 6–18	-0.015	0.033	0.071	-0.037	-0.072	-0.033	-0.006	-0.002
	(0.079)	(0.036)	(0.049)	(0.128)	(0.120)	(0.076)	(0.072)	(0.083)
Pre-period treated mean	-0.07	0.16	0.17	0.21	0.20	-0.11	0.49	0.64
Person-waves	30,046	30,057	5,849	25,037	22,593	29,952	6,075	6,075
Panel D. Women single at baseline								
Post GLP-1 start (all waves)	-0.020	0.056	-0.026	0.037	-0.041	-0.093	-0.319	-0.269
	(0.084)	(0.044)	(0.066)	(0.083)	(0.127)	(0.088)	(0.112)	(0.116)
Post, quarters 6–18	-0.082	0.084	-0.061	0.069	-0.114	-0.064	-0.417	-0.362
	(0.113)	(0.062)	(0.037)	(0.161)	(0.163)	(0.168)	(0.138)	(0.147)
Pre-period treated mean	-0.27	0.21	0.15	0.07	0.20	-0.12	0.36	0.41
Person-waves	15,768	15,793	2,883	13,031	11,746	15,802	2,987	2,987

Notes: The paper's baseline partnership split for women weight-loss starters, re-estimated with controls restricted to interested never-users and excluding later adopters. The table uses the same dependent variables, measurement sources, transformations, matching, event windows, post-interview control filter, and reporting conventions as Table A6. Standard errors in parentheses, clustered by person.

Table A9: Outcomes by baseline employment status under the interested-never-user control group pool

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women employed at baseline								
Post GLP-1 start (all quarters)	-1.781	0.083	0.028	-0.022	-0.897	0.006	0.098	0.004
	(0.592)	(0.035)	(0.036)	(0.015)	(0.571)	(0.007)	(0.041)	(0.026)
Post, quarters 6–18	-3.871	0.131	0.049	-0.040	-1.305	0.013	0.165	0.004
	(0.697)	(0.060)	(0.053)	(0.021)	(0.794)	(0.012)	(0.065)	(0.029)
Pre-period treated mean	34.3	11.33	10.38	0.97	37.8	0.02	0.61	0.02
Person-waves	6,656	41,404	39,772	41,404	41,401	41,404	41,404	28,465
Panel B. Women not employed at baseline								
Post GLP-1 start (all quarters)	-1.849	0.123	0.132	0.148	5.273	-0.088	0.019	0.001
	(0.741)	(0.102)	(0.098)	(0.061)	(2.261)	(0.052)	(0.014)	(0.011)
Post, quarters 6–18	-3.307	0.172	0.197	0.278	10.130	-0.163	0.002	0.021
	(1.316)	(0.115)	(0.113)	(0.087)	(3.256)	(0.072)	(0.030)	(0.030)
Pre-period treated mean	36.2	10.22	9.22	0.05	1.5	0.14	0.80	0.00
Person-waves	1,897	11,636	10,960	11,654	11,618	11,654	11,654	6,291

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Biennial health, well-being, and job-mobility outcomes, women employed at baseline								
Post GLP-1 start (all waves)	0.023	0.015	0.041	0.098	-0.078	-0.041	-0.133	-0.104
	(0.054)	(0.026)	(0.061)	(0.071)	(0.079)	(0.060)	(0.066)	(0.071)
Post, quarters 6–18	-0.012	0.042	0.032	0.113	-0.142	-0.075	-0.202	-0.133
	(0.083)	(0.040)	(0.037)	(0.122)	(0.118)	(0.103)	(0.088)	(0.099)
Pre-period treated mean	0.02	0.10	0.10	0.08	0.20	-0.07	0.42	0.57
Person-waves	33,362	33,377	6,540	28,359	25,535	33,241	6,734	6,734
Panel D. Biennial health, well-being, and job-mobility outcomes, women not employed at baseline								
Post GLP-1 start (all waves)	0.034	-0.001	0.010	-0.151	0.044	0.109	0.083	0.055
	(0.057)	(0.039)	(0.062)	(0.127)	(0.162)	(0.096)	(0.089)	(0.095)
Post, quarters 6–18	-0.180	0.110	-0.175	-0.007	0.264	-0.144	0.069	0.059
	(0.098)	(0.056)	(0.119)	(0.173)	(0.192)	(0.134)	(0.238)	(0.239)
Pre-period treated mean	-0.50	0.34	0.36	0.44	0.20	-0.20	0.52	0.55
Person-waves	10,528	10,549	1,870	8,417	7,623	10,563	1,999	1,999

Notes: The paper's baseline-employment split for women weight-loss starters, re-estimated with controls restricted to interested never-users and excluding later adopters. The table uses the same dependent variables, measurement sources, transformations, matching, event windows, post-interview control filter, and reporting conventions as Table A6, but splits treated women by baseline employment rather than baseline partnership. Standard errors in parentheses, clustered by person.

Table A10: Outcomes by baseline partnership under the uninterested-never-user control group pool

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all quarters)	-0.996	0.070	0.092	0.036	2.127	-0.023	-0.017	0.019
	(0.348)	(0.045)	(0.050)	(0.028)	(1.055)	(0.025)	(0.018)	(0.018)
Post, quarters 6–18	-2.266	0.081	0.079	0.072	3.646	-0.027	-0.024	0.030
	(0.472)	(0.051)	(0.058)	(0.040)	(1.583)	(0.038)	(0.021)	(0.021)
Pre-period treated mean	34.6	11.13	10.03	0.66	25.2	0.05	0.98	0.02
Person-waves	22,762	144,973	140,013	145,338	145,204	145,338	145,338	145,273
Panel B. Women single at baseline								
Post GLP-1 start (all quarters)	-1.588	0.090	0.015	0.004	0.358	-0.003	0.172	
	(0.552)	(0.058)	(0.050)	(0.017)	(0.773)	(0.014)	(0.047)	
Post, quarters 6–18	-2.128	0.173	0.090	0.004	0.832	0.006	0.259	
	(0.637)	(0.093)	(0.076)	(0.026)	(1.176)	(0.021)	(0.068)	
Pre-period treated mean	34.9	10.83	10.14	0.82	33.2	0.05	0.01	
Person-waves	10,711	68,478	64,102	68,511	68,446	68,511	68,511	

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Women partnered at baseline (married or cohabiting)								
Post GLP-1 start (all waves)	0.144	-0.111	0.057	0.014	-0.056	0.020	-0.047	-0.025
	(0.077)	(0.071)	(0.047)	(0.071)	(0.086)	(0.047)	(0.043)	(0.046)
Post, quarters 6–18	0.150	-0.154	0.020	0.042	-0.059	-0.068	-0.002	0.010
	(0.147)	(0.139)	(0.040)	(0.091)	(0.124)	(0.062)	(0.051)	(0.058)
Pre-period treated mean	-0.07	0.16	0.17	0.21	0.20	-0.11	0.49	0.64
Person-waves	122,169	122,176	22,371	101,378	91,392	122,045	23,327	23,327
Panel D. Women single at baseline								
Post GLP-1 start (all waves)	-0.021	0.034	0.000	0.148	-0.029	-0.021	-0.053	-0.059
	(0.077)	(0.042)	(0.057)	(0.066)	(0.119)	(0.068)	(0.058)	(0.055)
Post, quarters 6–18	-0.070	0.044	-0.013	0.111	-0.137	-0.040	-0.131	-0.072
	(0.095)	(0.063)	(0.049)	(0.113)	(0.140)	(0.130)	(0.068)	(0.073)
Pre-period treated mean	-0.27	0.21	0.15	0.07	0.20	-0.12	0.36	0.41
Person-waves	58,201	58,201	10,487	47,933	43,242	58,181	10,991	10,991

Notes: The paper's baseline partnership split for women weight-loss starters, re-estimated with controls restricted to uninterested never-users. The table uses the same dependent variables, measurement sources, transformations, matching, event windows, post-interview control filter, and reporting conventions as Table A6. Standard errors in parentheses, clustered by person.

Table A11: Outcomes by baseline employment under the uninterested-never-user control group pool

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Women employed at baseline								
Post GLP-1 start (all quarters)	-0.956	0.047	0.027	-0.017	0.093	0.006	0.067	0.023
	(0.426)	(0.032)	(0.037)	(0.016)	(0.559)	(0.009)	(0.032)	(0.027)
Post, quarters 6–18	-1.924	0.088	0.045	-0.022	0.266	0.009	0.105	0.039
	(0.493)	(0.051)	(0.050)	(0.021)	(0.807)	(0.012)	(0.043)	(0.029)
Pre-period treated mean	34.3	11.33	10.38	0.97	37.8	0.02	0.61	0.02
Person-waves	26,378	169,618	162,519	169,917	169,843	169,917	169,917	113,105
Panel B. Women not employed at baseline								
Post GLP-1 start (all quarters)	-1.000	0.107	0.103	0.131	5.148	-0.083	0.013	-0.008
	(0.419)	(0.089)	(0.087)	(0.055)	(2.134)	(0.057)	(0.013)	(0.012)
Post, quarters 6–18	-1.248	0.118	0.086	0.219	8.779	-0.074	0.031	-0.026
	(0.757)	(0.093)	(0.099)	(0.073)	(2.917)	(0.088)	(0.035)	(0.038)
Pre-period treated mean	36.2	10.22	9.22	0.05	1.5	0.14	0.80	0.00
Person-waves	5,885	36,411	34,492	36,491	36,401	36,491	36,491	23,337
	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel C. Biennial health, well-being, and job-mobility outcomes, women employed at baseline								
Post GLP-1 start (all waves)	0.015	0.000	0.065	0.138	0.007	-0.007	-0.045	-0.017
	(0.048)	(0.024)	(0.055)	(0.059)	(0.070)	(0.051)	(0.041)	(0.044)
Post, quarters 6–18	-0.043	0.024	0.066	0.122	-0.068	-0.043	-0.075	-0.021
	(0.064)	(0.038)	(0.037)	(0.090)	(0.093)	(0.084)	(0.051)	(0.055)
Pre-period treated mean	0.02	0.10	0.10	0.08	0.20	-0.07	0.42	0.57
Person-waves	140,039	140,043	26,228	117,874	106,254	139,899	27,098	27,098
Panel D. Biennial health, well-being, and job-mobility outcomes, women not employed at baseline								
Post GLP-1 start (all waves)	0.053	-0.053	0.024	-0.033	-0.033	0.020	0.051	0.045
	(0.070)	(0.053)	(0.074)	(0.112)	(0.140)	(0.096)	(0.069)	(0.069)
Post, quarters 6–18	-0.032	-0.012	-0.119	0.037	0.081	-0.169	0.123	0.119
	(0.131)	(0.108)	(0.059)	(0.129)	(0.170)	(0.116)	(0.128)	(0.128)
Pre-period treated mean	-0.50	0.34	0.36	0.44	0.20	-0.20	0.52	0.55
Person-waves	33,572	33,579	5,504	26,509	23,941	33,551	6,028	6,028

Notes: The paper's baseline-employment split for women weight-loss starters, re-estimated with controls restricted to uninterested never-users. The table uses the same dependent variables, measurement sources, transformations, matching, event windows, post-interview control filter, and reporting conventions as Table A6, but splits treated women by baseline employment rather than baseline partnership. Standard errors in parentheses, clustered by person.

C.2 Randomization inference

Randomization inference addresses a finite-sample inference concern. The stacked difference-in-differences estimates are based on a modest number of treated GLP-1 starters observed within cohort-by-baseline-status cells. Conventional clustered standard errors rely on large-sample approximations, while the randomization-inference exercise asks whether estimates as large as the observed ones arise when treatment labels are reshuffled among observationally comparable women within the same cohort and exact-match cell.

For each permutation, I reassign the treated labels within cohort × exact-match cell, preserving the number of treated women in each cell, and re-estimate the same matched difference-in-differences specification. I report two-sided placebo p-values based on the absolute studentized statistic, $|t|$. This is the preferred statistic because placebo assignments vary in their precision, especially in small cells; using the raw coefficient alone can give large weight to imprecise placebo estimates.

The randomization-inference results reinforce the main partnership finding. The studentized RI p-value is 0.006 for the 1.5+ year partnership effect among women single at baseline and p-values around 0.04-0.05 for the 1.5+ year treatments on employed, hours, and unemployment for women not employed at baseline.

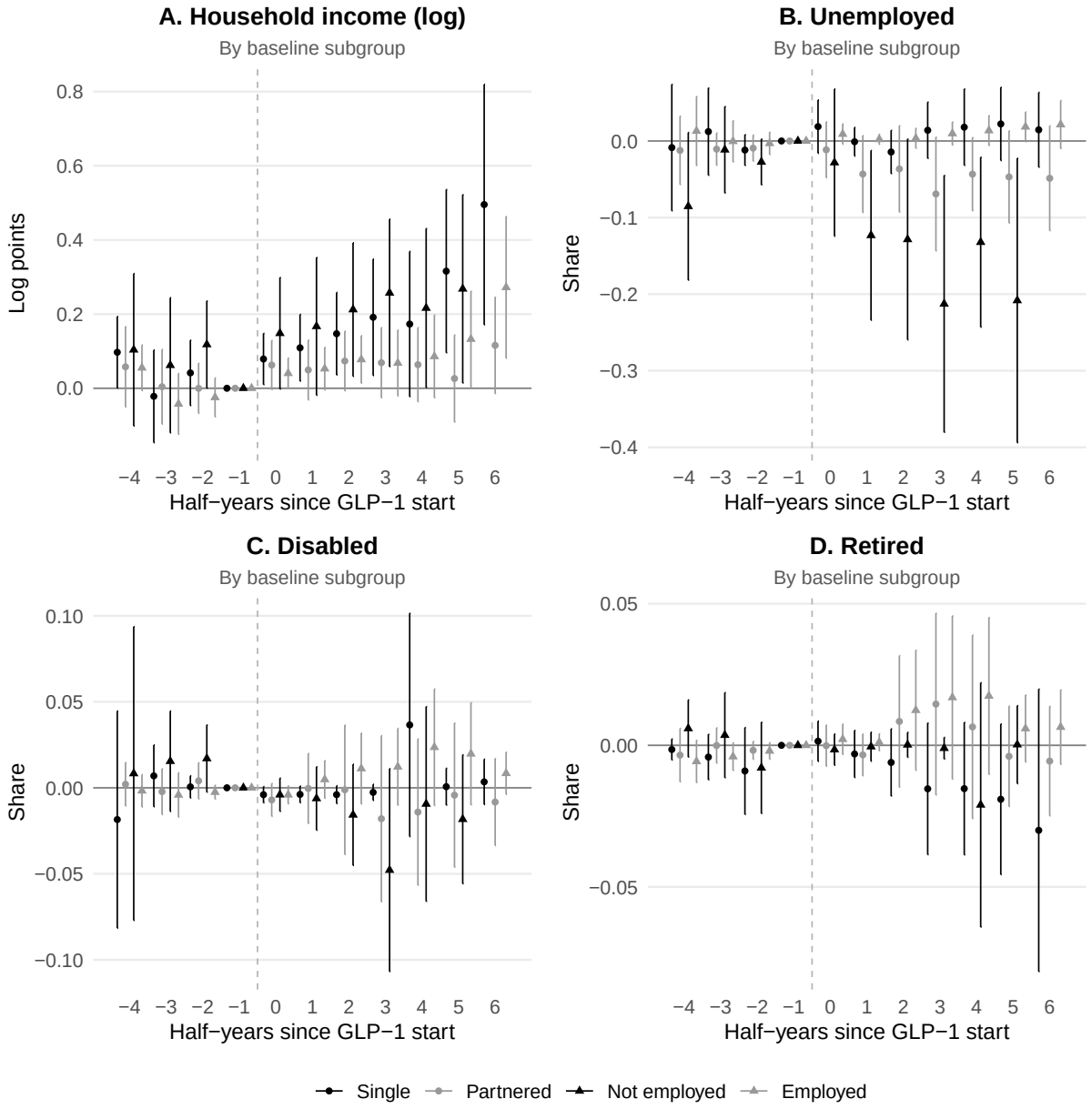
Table A12: Randomization inference for headline longer-run effects, 1,000 permutations

Sample	Outcome	Estimate	t	RI p on $\|t\|$
Single baseline	Partnered	0.286	4.04	0.006
Not-employed baseline	Employed	0.269	3.20	0.053
Not-employed baseline	Weekly hours	9.9	3.15	0.044
Not-employed baseline	Unemployed	-0.161	-2.30	0.043

D Additional outcomes and figures

Table A13 reports every outcome in main-text Tables 2 and 3 for the women's weight-loss benchmark, pooled across baseline groups rather than split by one. Much of the employment responses are masked by this pooling, but the partnership response is still visible.

Figure A3: Additional event studies: income and labor-status outcomes by baseline subgroup



Notes: Women starting a GLP-1 for weight loss, estimated on the extension-window matched stacks behind main-text Figure 2 (same sample, control pool, and two-quarter event clock). Each panel shows one quarterly outcome from main-text Tables 2 and 3 that Figure 2 does not already plot, as four separately matched baseline-subgroup event studies: single versus partnered women (the baseline-partnership split) and not employed versus employed women (the baseline-employment split). Marker shape distinguishes the two splits (circles for partnership, triangles for employment); shade marks the side within each split (black for single and women not employed at baseline, grey for partnered and employed). Estimates are in half-year (two-quarter) event periods; the dashed line marks the start boundary, and bars are 95 percent confidence intervals clustered by person. Each post-period point appears only while its own subgroup retains at least 15 treated women, so the thinner single and not employed series end earlier than the partnered and employed series.

Table A13: All outcomes: women who started a GLP-1 for weight loss

	Post, all		Post, quarters 6–18		Pre-period	Person-
	Estimate	SE	Estimate	SE	treated mean	waves
Economic outcomes (quarterly; BMI biennial)						
BMI	-1.718	(0.445)	-3.668	(0.528)	34.7	10,619
Log household income	0.074	(0.036)	0.106	(0.051)	11.03	66,126
Log household income per member	0.040	(0.038)	0.067	(0.050)	10.06	63,561
Employed	0.014	(0.021)	0.036	(0.032)	0.72	66,144
Weekly hours (0 if not working)	0.339	(0.806)	1.319	(1.211)	27.9	66,101
Unemployed	-0.012	(0.016)	-0.019	(0.024)	0.05	66,144
Married or cohabiting	0.062	(0.025)	0.102	(0.039)	0.66	66,144
Baseline relationship dissolved	0.000	(0.017)	0.005	(0.022)	0.02	43,604
Biennial health, well-being, and job-mobility outcomes (two-year event periods)						
Self-rated health (SD)	0.025	(0.043)	-0.042	(0.065)	-0.13	52,538
Poor or fair health	0.012	(0.021)	0.055	(0.032)	0.17	52,574
Health limits work	0.047	(0.048)	0.018	(0.036)	0.16	10,474
Depression (SD)	0.041	(0.061)	0.063	(0.096)	0.17	42,033
Loneliness (SD)	-0.072	(0.071)	-0.062	(0.099)	0.20	37,405
Life satisfaction (SD)	0.011	(0.048)	-0.039	(0.080)	-0.11	52,357
Employer changes	-0.113	(0.052)	-0.182	(0.066)	0.44	10,819
Job changes	-0.103	(0.052)	-0.151	(0.071)	0.57	10,819

Notes: All-outcomes for all women weight-loss starters together rather than split by baseline partnership or employment, on the identical specification. Standard errors clustered by person.

Table A14: Auxiliary outcomes moved from main-text Tables 2 and 3

	Own earn. ($/yr)	Disabled	Retired
Panel A. Baseline partnership split (main-text Table 2)			
Women partnered at baseline (married or cohabiting)			
Post GLP-1 start (all quarters)	1,610	-0.006	0.003
	(4,638)	(0.015)	(0.008)
Post, quarters 6–18	-7,603	-0.016	0.009
	(10,123)	(0.024)	(0.015)
Pre-period treated mean	50,649	0.10	0.01
Person-quarters (earnings: waves)	6,924	43,608	43,608
Women single at baseline			
Post GLP-1 start (all quarters)	-3,174	0.000	-0.002
	(5,092)	(0.006)	(0.007)
Post, quarters 6–18	-9,242	0.010	-0.013
	(6,399)	(0.012)	(0.013)
Pre-period treated mean	56,246	0.11	0.00
Person-quarters (earnings: waves)	3,349	21,255	21,255
Panel B. Baseline employment split (main-text Table 3)			
Women employed at baseline			
Post GLP-1 start (all quarters)	-914	0.007	0.008
	(4,607)	(0.007)	(0.007)
Post, quarters 6–18	-4,118	0.017	0.016
	(6,461)	(0.012)	(0.013)
Pre-period treated mean	67,807	0.00	0.00
Person-quarters (earnings: waves)	7,756	49,653	49,653
Women not employed at baseline			
Post GLP-1 start (all quarters)	9,874	-0.023	-0.002
	(3,471)	(0.020)	(0.005)
Post, quarters 6–18	-8,501	-0.041	-0.006
	(8,856)	(0.031)	(0.010)
Pre-period treated mean	5,773	0.37	0.02
Person-quarters (earnings: waves)	2,104	12,660	12,660

Notes: Additional outcomes using the same specification as main-text Tables 2 and 3.

E GLP-1 Effects on Men

Fewer men take these drugs to lose weight, and the male sample is too thin to split by baseline partnership or employment the way the paper splits the women. I instead rebuild the paper's outcomes for male weight-loss starters pooled across those baseline cells. Table A15 describes the 90 positive-weight male starters beside the male never-user pools; the treatment-effect table starts from the 88 men with valid first-start timing and then imposes the same event-time analysis restrictions as the women's benchmark. Table A16 reports the treatment effects.

The direct effect on weight is as strong for men as for women. Body-mass index falls 2.7 points over the post period and 4.5 points by the second post two-year window (both $p < 0.001$). Household income is flat, but noisy (a 1 percent increase over the post period, s.e. 4.4; 3.4 percent at quarters 6–18), employment does not move (0.0 points, s.e. 1.8), and weekly hours are a noisy zero. Partnership, the women's largest result, is absent for men: the married-or-cohabiting rate moves −3.3 points (s.e. 3.1), a point estimate of the *wrong* sign for an obesity-penalty story, with a 95 percent confidence interval whose upper edge sits well below the single women's phase-in. The one striking result for men is partnership dissolution. Among baseline partnered men (which is where partnership dissolution is well-defined), I find a 6.8pp increase in men leaving their baseline partners (s.e. 2.9), which rises to 12.1pp after 1.5+ years (s.e. 6.1). Unlike women, partnered men leave existing relationships with weight loss. Other labor market outcomes are imprecise nulls rather than tight zeros.

The health and well-being block (Table A16, Panel B) does show men report a 16.1 percentage point (s.e. 7.6) decline in reporting that health limits their abilities to work after 1.5+ years, indicating some health improvements. The point estimate on self-reported health moves in the same direction but is noisy. However, point estimates on depression, loneliness, and life satisfaction all are noisy but point in the worsening direction.

Table A15: Descriptive statistics: men by GLP-1 status

	Weight-loss starters	Interested, no take-up	Not interested	p-value, starters vs. interested	p-value, starters vs. not interested
Panel A. Demographics (GLP-1 module interview)					
Age (years)	48.8	49.9	52.9	0.590	0.019
	[13.5]	[15.1]	[17.5]		
Non-Hispanic White	0.65	0.62	0.69	0.680	0.585
	[0.48]	[0.48]	[0.46]		
Non-Hispanic Black	0.09	0.08	0.08	0.915	0.754
	[0.28]	[0.28]	[0.26]		
Hispanic/Latino	0.16	0.18	0.13	0.809	0.514
	[0.37]	[0.38]	[0.33]		
BA or higher	0.50	0.40	0.48	0.206	0.841
	[0.50]	[0.49]	[0.50]		
Employed or on leave	0.56	0.55	0.57	0.928	0.877
	[0.50]	[0.50]	[0.49]		
Panel B. Health and well-being (UAS core, pre-baseline)					
Body-mass index	36.3	32.2	27.5	<0.001	<0.001
	[7.1]	[6.7]	[4.9]		
Obese (BMI $\geq$ 30)	0.75	0.57	0.25	0.018	<0.001
	[0.43]	[0.49]	[0.43]		
Diabetes	0.18	0.23	0.14	0.460	0.489
	[0.38]	[0.42]	[0.35]		
Self-rated health (SD)	-0.37	-0.29	0.12	0.587	<0.001
	[0.88]	[0.97]	[0.95]		
Poor or fair health	0.32	0.25	0.14	0.365	0.011
	[0.47]	[0.43]	[0.35]		
Health limits work	0.22	0.22	0.15	0.940	0.290
	[0.42]	[0.41]	[0.36]		
CES-D depression (SD)	0.20	0.25	-0.20	0.784	0.019
	[1.13]	[1.15]	[0.90]		
Loneliness (SD)	0.01	0.22	-0.20	0.178	0.119
	[0.88]	[1.05]	[1.06]		
Satisfaction with life (SD)	-0.01	-0.25	0.05	0.092	0.607
	[0.85]	[1.13]	[0.97]		
Panel C. Economic resources (panel surveys, pre-baseline)					
Household income ($)	$104,281	$82,084	$91,250	0.020	0.138
	[$59,984]	[$58,141]	[$56,849]		
Own annual earnings ($)	$73,196	$62,702	$60,186	0.466	0.316
	[$106,709]	[$118,017]	[$97,758]		
Weekly work hours	26.9	24.0	23.2	0.394	0.242
	[22.6]	[22.2]	[21.4]		
Unemployed	0.08	0.07	0.05	0.871	0.475
	[0.27]	[0.25]	[0.21]		
Disabled	0.17	0.12	0.06	0.370	0.048
	[0.38]	[0.32]	[0.24]		
Retired	0.10	0.17	0.25	0.071	<0.001
	[0.30]	[0.38]	[0.43]		
Married or cohabiting	0.73	0.65	0.69	0.277	0.495
	[0.44]	[0.48]	[0.46]		
Panel D. GLP-1 module					
Currently using (vs. former)	0.56				
	[0.50]				
Months since first start	12.6				
	[13.3]				
Body weight lost (%)	7.4				
	[9.3]				
Pays fully out of pocket	0.44				
	[0.50]				
Monthly out-of-pocket ($)	$397				
	[$338]				
Respondents (positive-weight module)	90	346	2,569		

Notes: Survey-weighted means by GLP-1 status in the GLP-1 module (December 2024–February 2025), with standard deviations in brackets. The respondent row is the positive-weight module count, before requiring valid first-start timing. The last two columns report p-values for the difference in weighted means between starters and each never-user group (a difference-in-proportions test for the binary rows).

Table A16: Effects of Male GLP-1 initiation for weight loss

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Male weight-loss starters (outcomes through 2026Q2)								
Post GLP-1 start (all quarters)	-2.683	0.010	0.017	0.000	0.078	0.002	-0.033	0.068
	(0.510)	(0.044)	(0.053)	(0.018)	(0.817)	(0.011)	(0.031)	(0.029)
Post, quarters 6–18	-4.458	0.034	0.010	0.003	2.168	0.008	-0.040	0.121
	(0.738)	(0.083)	(0.087)	(0.035)	(1.873)	(0.011)	(0.053)	(0.061)
Pre-period treated mean	35.7	11.18	10.16	0.69	29.7	0.08	0.76	0.00
Person-quarters (biennial: person-waves)	2,934	17,553	17,009	17,553	17,523	17,553	17,553	13,363
	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel B. Biennial health, well-being, and job-mobility outcomes								
Post GLP-1 start (all waves)	0.025	0.040	0.091	0.159	0.016	-0.110	0.099	0.094
	(0.086)	(0.025)	(0.093)	(0.123)	(0.094)	(0.081)	(0.096)	(0.107)
Post, quarters 6–18	0.090	0.006	-0.161	0.159	0.115	-0.161	0.201	0.215
	(0.125)	(0.039)	(0.076)	(0.128)	(0.130)	(0.094)	(0.174)	(0.195)
Pre-period treated mean	-0.30	0.30	0.23	0.00	-0.07	0.00	0.45	0.57
Person-waves	15,329	15,329	2,835	12,508	11,220	15,255	3,016	3,016

Notes: Stacked matched difference-in-differences estimates for men first starting a GLP-1 for weight loss, starting from the valid-timing subset of positive-weight male weight-loss starters. Controls are demand-pool controls (later adopters plus interested never-users), exact cells on baseline employment × partnership, propensity reweighting on race, lagged BMI, diabetes, self-rated health, depression, whether health limits work, life satisfaction, and household income; outcomes through 2026Q2. The partnership-dissolution column is defined among men partnered at baseline, using the same clean baseline-relationship definition as main-text Table 2. Standard errors in parentheses, clustered by person.

F Validating the design against the GLP-1 weight-loss trials

This section evaluates the UAS treatment measure against clinical evidence on GLP-1 weight loss. The exercise has two purposes. First, it asks whether the matched UAS design recovers a biologically meaningful weight-loss first stage. Second, it helps interpret why the real-world effects on physical functioning, general health, and depressive symptoms may be smaller than the effects reported in tightly managed medical trials.

The randomized-trial benchmarks come from published studies of semaglutide and tirzepatide in adults with overweight or obesity and without diabetes. STEP 1 and SURMOUNT-1 provide the placebo-adjusted percent body-weight benchmarks for semaglutide and tirzepatide (Wilding et al., 2021; Jastreboff et al., 2022). The STEP patient-reported-outcome paper provides the physical-functioning and general-health benchmarks used below (Rubino et al., 2024). I do not use the tirzepatide health-related-quality-of-life evidence as a physical-functioning benchmark because the directly comparable physical-functioning endpoint was not reported for SURMOUNT-1. The depressive-symptom benchmark is also semaglutide-only: it comes from the STEP 5 psychiatric-safety analysis (Wadden et al., 2024). I did not find a comparable tirzepatide depression-effect benchmark.

For non-weight outcomes, I convert published trial effects to standard-deviation units using

$$\Delta_{\text{RCT}} = \frac{(\bar{Y}_{T,1} - \bar{Y}_{T,0}) - (\bar{Y}_{P,1} - \bar{Y}_{P,0})}{s_{0,\text{pooled}}},$$

where T and P denote active-drug and placebo arms and $s_{0,\text{pooled}}$ is the pooled baseline standard deviation for the same outcome in the same paper. When a paper reports a standardized treatment contrast directly, I use it. For the UAS columns, I use the health variables that also enter main-text Tables 2 and 3: self-rated health and depressive symptoms. The physical-functioning row uses physical functioning questions from the biennial panel. To make this analysis similar to these studies, I pool male and female weight-loss starters; this pooling does not change the main paper's female treatment-effect estimands.

Several differences between randomized trials and real-world use are important for interpreting the comparison. First, trial volunteers were heavier and less healthy than the typical person who starts these drugs in the population. Their average body-mass index was close to 38, compared with 35 in the UAS weight-loss starter sample. STEP 1 and SURMOUNT-1 required BMI of at least 30, or BMI of at least 27 with at least one weight-related comorbidity, and they excluded patients with diabetes. UAS observes diagnosed diabetes, high blood pressure, heart disease, and stroke, but does not observe dyslipidemia or sleep apnea. I therefore use the closest clean UAS analog for the trial-comparable sample: no baseline diabetes and either baseline BMI of 30 or above, or baseline BMI of 27–29.9 with observed hypertension, heart disease, or stroke.

Second, real-world persistence and dosing differ from trial protocols. In STEP 1 and SURMOUNT-1, only about 15 percent of participants assigned to active treatment had discontinued the drug by the trial endpoint, 68 weeks in STEP 1 and 72 weeks in SURMOUNT-1. The UAS does not

observe complete medication histories after initiation, so I cannot construct the same endpoint-specific discontinuation rate. However, by the 2024–2025 GLP-1 module, 53 percent of UAS weight-loss starters report that they formerly used, but are not currently using, a GLP-1. This points to substantially lower real-world persistence than in the trials. The pattern lines up with electronic-medical-record evidence from Gasoyan et al. (2025), who find that about half of real-world semaglutide or tirzepatide users discontinue within the first year. They also find that low maintenance dosing is common. These differences are expected: trial participants received study medication, regular clinical follow-up, and structured lifestyle counseling, while real-world users face cost, insurance frictions, drug shortages, and physician-specific dosing decisions.

Third, trial protocols combined medication with more intensive lifestyle support than typical real-world use. In the semaglutide trials, participants received diet and exercise counseling and recorded diet and physical activity. European Medicines Agency (2022) reports that participants randomized to GLP-1 treatment reported about 270 minutes per week of exercise. A UAS-comparable question asks about exercise frequency in the biennial panel. Prior to starting a GLP-1, UAS starters reported an average of 98 minutes per week of exercise, and they reported 101 minutes per week in the first reporting quarter at least five quarters after initiation.[1] This is less than 40 percent of the exercise reported by GLP-1 participants in the trial evidence. The comparison is imperfect, but it suggests that real-world medication use is much less tightly bundled with increased exercise than trial treatment was.

Weight. STEP 1 reports a placebo-adjusted semaglutide effect of 12.4 percent body-weight loss at 68 weeks, with a 95 percent confidence interval of 11.4 to 13.4 percent lost. SURMOUNT-1 reports placebo-adjusted tirzepatide losses across doses from 11.9 to 17.8 percent at 72 weeks, with confidence-interval bounds spanning 11.2 to 17.5 percent lost. Since UAS weight is recorded only every two years, I take each person's first valid weight observation at event quarter 5 or later. In practice, these readings fall between event quarters 5 and 12, with an average weight reading 7.2 quarters after initiation.

The matched UAS design implies an 8.6 percent body-weight decline, rising to 9.6 percent in the trial-comparable BMI/comorbidity, no-diabetes sample. These estimates are below the randomized-trial efficacy benchmarks, but they are very close to real-world clinical-practice evidence: Gasoyan et al. (2025) estimate 8.7 percent body-weight loss at one year among patients without diabetes initiating injectable semaglutide or tirzepatide. I therefore interpret the UAS weight estimates as evidence that the design recovers a meaningful real-world GLP-1 first stage, even though it does not reproduce the larger weight losses achieved under trial protocols. Table A17 reports these losses.

Physical functioning. Physical functioning is a person's own report of how easily she manages everyday physical tasks, such as walking several blocks, climbing stairs, and lifting or carrying groceries. The STEP 1 patient-reported-outcome paper finds a clear improvement, +0.24 standard deviations with confidence interval $[+0.16, +0.32]$. Using the same endpoint rule as the weight

[1] I convert the UAS question about exercise frequency into minutes by assuming 45 minutes per reported exercise session.

comparison, I take each person's first valid physical-functioning observation at event quarter 5 or later. In practice, these readings average 6.7 quarters after initiation. Across all UAS starters, I estimate -0.04 standard deviations, with confidence interval $[-0.17, +0.08]$. Restricting the UAS sample to the BMI/comorbidity, no-diabetes analog raises the estimate to $+0.07$ standard deviations, with confidence interval $[-0.07, +0.21]$.

The UAS physical-functioning estimates are therefore smaller than the STEP 1 benchmark. That gap is consistent with the implementation differences above: UAS starters lose somewhat less weight, are much less likely to remain continuously on treatment, and report far less exercise than trial participants. A pragmatic clinical study, POSEY, points in the same direction. POSEY randomized adults with obesity in an employer setting to semaglutide 2.4 mg or other anti-obesity medications and found larger weight loss with semaglutide but only a small and statistically insignificant differential improvement in IWQOL-Lite-CT physical function (Peters et al., 2026). I therefore read the physical-functioning comparison as showing that real-world functional gains are smaller and imprecise, not as evidence that the UAS treatment measure fails to capture meaningful weight loss.

General health. On general health, the experimental evidence and UAS data ask different questions. The trials use multi-item health-related-quality-of-life scales, while UAS carries a single self-rated-health question with five answer categories. The STEP 1 general-health treatment effect is $+0.27$ standard deviations, with confidence interval $[+0.18, +0.35]$, indicating an improvement in general health. Since self-rated health is observed monthly after initiation, I use monthly readings 15–18 months after initiation, widening to months 12–20 for stack-persons with no reading in the narrower window. The treated endpoint readings average 16.4 months after initiation.

The UAS estimate is $+0.04$ standard deviations in the all-starter sample, with confidence interval $[-0.06, +0.14]$, and $+0.08$ standard deviations in the trial-comparable BMI/comorbidity, no-diabetes sample, with confidence interval $[-0.05, +0.20]$. These estimates are below the experimental benchmark. The gap may reflect both measurement differences—a single self-rated-health item in UAS versus multi-item trial scales—and implementation differences, including smaller weight loss, lower treatment persistence, lower maintenance doses, and much lower reported exercise in the real-world setting.

Depressive symptoms. The STEP psychiatric-safety analysis measures depressive symptoms with the PHQ-9 and reports a small improvement, -0.20 standard deviations on the symptom-score scale, with a 95 percent confidence interval of $[-0.98, +0.59]$, where negative means fewer symptoms (Wadden et al., 2024). The STEP psychiatric-safety sample excluded participants with PHQ-9 scores of 15 or more, recent major depression, or suicidality, while still allowing lower baseline symptoms. I therefore screen out high levels of depressive symptoms by requiring baseline CES-D scores of 5 or below, on top of the BMI/comorbidity, no-diabetes screen, for the UAS trial-comparable estimate.

Using the same monthly endpoint rule as the general-health row, I estimate $+0.04$ standard deviations in the full UAS sample, with confidence interval $[-0.08, +0.16]$, and $+0.04$ standard devia-

tions after imposing the combined trial-comparable screen, with confidence interval $[-0.09, +0.18]$. The point estimates indicate slightly more depressive symptoms, but they are small and sit well inside the randomized-evidence confidence interval. The trial-UAS comparison is also difficult to interpret as a medication-specific difference because trial participants received more structured lifestyle support and reported much higher exercise levels, and exercise itself has been shown to reduce depressive symptoms (Noetel et al., 2024).

Overall, the clinical calibration supports two conclusions. First, the UAS design recovers substantial, real-world GLP-1 weight loss: the estimated effect is below the randomized-trial efficacy benchmarks but very close to recent clinical-practice evidence. Second, the smaller UAS effects on physical functioning, general health, and depressive symptoms are consistent with the less intensive way these medications are used outside trial protocols. This reinforces the interpretation in the main text: the partnership and employment-entry effects occur in a setting with meaningful weight loss, but they do not appear to be a generic response to broad improvements in subjective health or well-being.

Table A17: UAS-based estimates vs. randomized-trials

Outcome	Randomized trials	UAS-based estimate (95% CI)	
		All starters	Trial-comparable
Weight			
Weight loss	STEP 1: −12.4% [−13.4, −11.4] SURMOUNT-1: −11.9 to −17.8% (CI span [−17.5, −11.2])	−8.6% [−11.0, −6.1]	−9.6% [−12.0, −7.1]
Physical health			
Physical functioning	STEP 1: +0.24 [+0.16, +0.32]	−0.04 [−0.17, +0.08]	+0.07 [−0.07, +0.21]
General health	STEP 1: +0.27 [+0.18, +0.35]	+0.04 [−0.06, +0.14]	+0.08 [−0.05, +0.20]
Depressive symptoms			
Depressive symptoms	STEP-5: −0.20 [−0.98, +0.59]	+0.04 [−0.08, +0.16]	+0.04 [−0.09, +0.18]

Notes: Weight is reported as percent body-weight change; physical health and depressive symptoms are reported in standard-deviation units. The RCT weight benchmark is the placebo-adjusted percent body-weight change reported in STEP 1 and SURMOUNT-1 (Wilding et al., 2021; Jastreboff et al., 2022); trial-reported losses are signed as negative percent changes in the table. Physical functioning and general health are taken from the STEP 1 patient-reported-outcome analysis (Rubino et al., 2024). Depressive symptoms are taken from the semaglutide STEP psychiatric-safety analysis (Wadden et al., 2024). UAS-based estimates are matched difference-in-differences and pooled across men and women. The weight and physical-functioning rows use each person's first valid post-initiation outcome observation at event quarter 5 or later. The general-health and depressive-symptom rows use monthly readings 15–18 months after initiation, widening to months 12–20 for stack-persons with no 15–18-month reading. The *trial-comparable* column re-estimates each effect on the subsample that resembles the trial volunteers: no baseline diabetes and either baseline BMI ≥ 30 or baseline BMI 27–29.9 with observed hypertension, heart disease, or stroke. For depressive symptoms, I additionally require baseline CES-D scores of 5 or below. Brackets are 95 percent confidence intervals, except the SURMOUNT-1 weight entry, where the parenthetical reports the span of dose-specific confidence-interval bounds.

Table A18: Pooled interaction model: women's treatment effect by baseline partnership and employment

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Economic and social outcomes (quarterly; BMI biennial)								
Post GLP-1 start (all quarters)								
Direct effect (employed & partnered)	-1.392	0.037	0.022	-0.009	-0.159	-0.002	0.011	-0.004
	(0.615)	(0.035)	(0.048)	(0.025)	(0.852)	(0.010)	(0.028)	(0.026)
× baseline single	-0.925	0.062	-0.013	-0.013	-1.298	0.016	0.187	
	(0.973)	(0.070)	(0.068)	(0.035)	(1.361)	(0.025)	(0.054)	
× baseline not employed	-0.073	0.061	0.088	0.106	3.637	-0.060	-0.055	0.014
	(0.981)	(0.108)	(0.115)	(0.061)	(2.281)	(0.052)	(0.033)	(0.032)
Post, quarters 6–18								
Direct effect (employed & partnered)	-3.694	0.041	0.006	0.009	0.590	-0.007	0.038	-0.007
	(0.894)	(0.057)	(0.068)	(0.034)	(1.128)	(0.014)	(0.039)	(0.029)
× baseline single	0.073	0.133	0.103	-0.033	-1.813	0.022	0.242	
	(1.387)	(0.101)	(0.093)	(0.050)	(1.965)	(0.033)	(0.072)	
× baseline not employed	0.216	0.066	0.107	0.186	6.513	-0.091	-0.114	0.043
	(2.070)	(0.119)	(0.139)	(0.081)	(3.124)	(0.066)	(0.056)	(0.053)
Pre-period treated mean	34.7	11.03	10.07	0.72	27.9	0.05	0.66	0.02
Person-quarters (biennial: person-waves)	10,619	66,126	63,561	66,144	66,101	66,144	66,144	43,604

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel B. Biennial health, well-being, and job-mobility outcomes								
Post GLP-1 start (all waves)								
Direct effect (employed & partnered)	0.019	0.010	0.068	0.064	-0.089	0.034	-0.056	-0.055
	(0.059)	(0.030)	(0.076)	(0.083)	(0.081)	(0.060)	(0.060)	(0.065)
× baseline single	-0.018	0.026	-0.084	0.056	-0.043	-0.139	-0.216	-0.144
	(0.103)	(0.049)	(0.098)	(0.115)	(0.176)	(0.105)	(0.117)	(0.120)
× baseline not employed	0.043	-0.024	0.029	-0.177	0.136	0.080	0.065	0.009
	(0.080)	(0.047)	(0.096)	(0.150)	(0.173)	(0.099)	(0.108)	(0.116)
Post, quarters 6–18								
Direct effect (employed & partnered)	-0.031	0.052	0.101	0.063	-0.095	-0.023	-0.031	-0.021
	(0.094)	(0.043)	(0.056)	(0.128)	(0.128)	(0.077)	(0.074)	(0.083)
× baseline single	0.015	-0.004	-0.161	0.076	-0.081	-0.060	-0.378	-0.262
	(0.144)	(0.070)	(0.065)	(0.180)	(0.217)	(0.164)	(0.145)	(0.158)
× baseline not employed	-0.067	0.028	-0.144	-0.114	0.272	-0.019	0.083	-0.043
	(0.117)	(0.068)	(0.109)	(0.208)	(0.227)	(0.149)	(0.170)	(0.189)
Pre-period treated mean	-0.13	0.17	0.16	0.17	0.20	-0.11	0.44	0.57
Person-quarters (biennial: person-waves)	52,538	52,574	10,474	42,033	37,405	52,357	10,819	10,819

Notes: The main-text baseline-partnership split and the main-text baseline-employment split are re-estimated as a single pooled interacted difference-in-differences, so the heterogeneity in treatment effects by baseline partnership and baseline employment is read off one model. The matched stack uses the paper's benchmark specification: demand-pool controls (later adopters plus interested never-users), exact matching on baseline employment × baseline partnership, propensity-score odds reweighting on race, lagged BMI, diabetes, self-rated health, CES-D, health-limits-work, life satisfaction, and household income, outcomes through 2026Q2, and the interim verified-control filter. The only change relative to the main split tables is the estimating equation: $Y_{it} = \alpha_i + \delta_{ct} + \beta\,(\text{Treat}_i \times \text{Post}_t) + \gamma_S(\text{Treat}_i \times \text{Post}_t \times \text{single}_i) + \gamma_N(\text{Treat}_i \times \text{Post}_t \times \text{notemp}_i) + \lambda_S(\text{Post}_t \times \text{single}_i) + \lambda_N(\text{Post}_t \times \text{notemp}_i) + \varepsilon_{it}$, where single is baseline single (vs. partnered) and notemp is baseline not employed (vs. employed). The stack-person fixed effects α_i absorb the baseline-group levels and δ_{ct} are cohort × calendar-quarter fixed effects; the Post×group main effects (λ_S, λ_N) let baseline-single and baseline-not-employed units carry their own post trends, so the *Direct effect* (β) is the treatment effect for the baseline employed-and-partnered group against its own controls and each interaction (γ_S, γ_N) is a differential treatment effect. The later-horizon rows replace Post with the $[0, 5]$ and $[6, 18]$ event-quarter windows (and their group main effects) and report the $[6, 18]$ coefficients. BMI and all Panel B outcomes are biennial (two-year event clock); the remaining Panel A outcomes are quarterly. The partnership-dissolution column is defined among women partnered at baseline, as in main-text Tables 2 and 3, so baseline-single dissolution cells are blank. Self-rated health, depression, loneliness, and life satisfaction are in nationally representative baseline-SD units; poor/fair health, health-limits-work, and partnership dissolution are indicators. Standard errors in parentheses, clustered by person.

Table A19: Pooled interaction model: women's total treatment effect by 2×2 baseline group

	BMI	Log HH income	Log inc. per mem.	Employed	Weekly hours	Unemployed	Partnered	Partnership dissol.
Panel A. Economic and social outcomes (quarterly; BMI biennial)								
Post GLP-1 start (all quarters)								
Partnered & employed	-1.392	0.037	0.022	-0.009	-0.159	-0.002	0.011	-0.004
	(0.615)	(0.035)	(0.048)	(0.025)	(0.852)	(0.010)	(0.028)	(0.026)
Single & employed	-2.317	0.099	0.009	-0.022	-1.457	0.014	0.198	
	(0.852)	(0.061)	(0.050)	(0.024)	(1.092)	(0.021)	(0.053)	
Partnered & not employed	-1.465	0.098	0.110	0.096	3.478	-0.062	-0.044	0.010
	(0.876)	(0.110)	(0.113)	(0.064)	(2.391)	(0.057)	(0.021)	(0.015)
Single & not employed	-2.390	0.160	0.097	0.083	2.179	-0.046	0.142	
	(0.991)	(0.108)	(0.110)	(0.057)	(2.122)	(0.048)	(0.047)	
Post, quarters 6–18								
Partnered & employed	-3.694	0.041	0.006	0.009	0.590	-0.007	0.038	-0.007
	(0.894)	(0.057)	(0.068)	(0.034)	(1.128)	(0.014)	(0.039)	(0.029)
Single & employed	-3.621	0.174	0.108	-0.024	-1.223	0.015	0.279	
	(0.986)	(0.091)	(0.068)	(0.036)	(1.639)	(0.029)	(0.075)	
Partnered & not employed	-3.477	0.107	0.113	0.195	7.103	-0.098	-0.077	0.036
	(1.772)	(0.114)	(0.134)	(0.088)	(3.380)	(0.074)	(0.045)	(0.041)
Single & not employed	-3.405	0.240	0.215	0.162	5.290	-0.076	0.165	
	(2.119)	(0.136)	(0.140)	(0.082)	(3.208)	(0.068)	(0.074)	
Pre-period treated mean (all)	34.7	11.03	10.07	0.72	27.9	0.05	0.66	0.02

	Self-rated health (SD)	Poor/fair health	Health limits work	Depression (SD)	Lonely (SD)	Life satisf. (SD)	Employer changes	Job changes
Panel B. Biennial health, well-being, and job-mobility outcomes								
Post GLP-1 start (all waves)								
Partnered & employed	0.019	0.010	0.068	0.064	-0.089	0.034	-0.056	-0.055
	(0.059)	(0.030)	(0.076)	(0.083)	(0.081)	(0.060)	(0.060)	(0.065)
Single & employed	0.001	0.036	-0.016	0.120	-0.132	-0.105	-0.272	-0.198
	(0.094)	(0.040)	(0.080)	(0.086)	(0.160)	(0.095)	(0.109)	(0.110)
Partnered & not employed	0.062	-0.014	0.096	-0.112	0.047	0.113	0.009	-0.046
	(0.060)	(0.034)	(0.069)	(0.142)	(0.168)	(0.093)	(0.099)	(0.104)
Single & not employed	0.043	0.012	0.012	-0.056	0.004	-0.025	-0.207	-0.190
	(0.103)	(0.061)	(0.109)	(0.159)	(0.192)	(0.108)	(0.122)	(0.126)
Post, quarters 6–18								
Partnered & employed	-0.031	0.052	0.101	0.063	-0.095	-0.023	-0.031	-0.021
	(0.094)	(0.043)	(0.056)	(0.128)	(0.128)	(0.077)	(0.074)	(0.083)
Single & employed	-0.015	0.048	-0.061	0.139	-0.175	-0.084	-0.409	-0.282
	(0.124)	(0.059)	(0.046)	(0.147)	(0.189)	(0.162)	(0.130)	(0.140)
Partnered & not employed	-0.097	0.080	-0.044	-0.051	0.177	-0.042	0.051	-0.064
	(0.091)	(0.056)	(0.093)	(0.196)	(0.218)	(0.143)	(0.171)	(0.193)
Single & not employed	-0.082	0.076	-0.205	0.025	0.097	-0.102	-0.326	-0.326
	(0.139)	(0.082)	(0.114)	(0.224)	(0.236)	(0.177)	(0.170)	(0.179)
Pre-period treated mean (all)	-0.13	0.17	0.16	0.17	0.20	-0.11	0.44	0.57

Notes: The same pooled interacted difference-in-differences as Table A18, but reporting each baseline group's *total* treatment effect instead of the interaction contrasts. The four 2×2 baseline groups are recovered as linear combinations of the model coefficients: partnered & employed (the base group) = β; single & employed = $\beta + \gamma_S$; partnered & not employed = $\beta + \gamma_N$; single & not employed = $\beta + \gamma_S + \gamma_N$, with standard errors from the model variance-covariance matrix. The single-and-not-employed total borrows strength through the additive structure—there is no separate single × not-employed interaction in the model—so it is not a saturated four-way estimate. "Pre-period treated mean (all)" pools the four groups; group-specific baselines are in the main-text baseline-partnership and baseline-employment tables. Outcome definitions, units, matching, sample, and horizons are exactly as in Table A18; because partnership dissolution is defined only among women partnered at baseline, the baseline-single dissolution totals are blank. Standard errors in parentheses, clustered by person.

Table A20: Job mobility by GLP-1 payment route: women, pooled across baseline partnership

	Employer changes (cumulative count)	Job changes (cumulative count)
Panel A. Starters paying fully out of pocket		
Post GLP-1 start (all waves)	-0.050	-0.082
	(0.062)	(0.063)
Post, quarters 6–18	-0.069	-0.052
	(0.077)	(0.082)
Pre-period treated mean	0.52	0.62
Panel B. Starters with insurance coverage (no copay or copay)		
Post GLP-1 start (all waves)	-0.084	-0.042
	(0.064)	(0.067)
Post, quarters 6–18	-0.179	-0.140
	(0.070)	(0.081)
Pre-period treated mean	0.39	0.56

Notes: The job-mobility outcomes from main-text Table 2 are re-estimated on a single pooled women's stack (no baseline-partnership split), with the treated weight-loss population restricted by the GLP-1 module's payment-route report: Panel A includes starters paying fully out of pocket, and Panel B includes starters with insurance coverage (no copay or a copay). "Employer changes" count cross-wave employer switches and moves into or out of employment. "Job changes" include those employer-change events and also same-employer changes in kind of work or position. Standard errors in parentheses, clustered by person.

References

European Medicines Agency. 2022. "Wegovy: EPAR Public Assessment Report." European Medicines Agency Assessment report EMA/112307/2022. Committee for Medicinal Products for Human Use (CHMP); procedure No. EMEA/H/C/005422/0000; adopted 11 November 2021.

Gasoyan, Hamlet, W Scott Butsch, Rebecca Schulte, Nicholas J Casacchia, Phuc Le, Christopher B Boyer, Marcio L Griebeler, Bartolome Burguera, and Michael B Rothberg. 2025. "Changes in weight and glycemic control following obesity treatment with semaglutide or tirzepatide by discontinuation status." *Obesity*, 33(9): 1657–1667.

Jastreboff, Ania M., Louis J. Aronne, Nadia N. Ahmad, Sean Wharton, Lisa Connery, Breno Alves, Arihiro Kiyosue, Shuyu Zhang, Bing Liu, Mathijs C. Bunck, and Adam Stefanski. 2022. "Tirzepatide Once Weekly for the Treatment of Obesity." *New England Journal of Medicine*, 387(3): 205–216.

Noetel, Michael, Taren Sanders, Daniel Gallardo-Gómez, Paul Taylor, Borja del Pozo Cruz, Daniel van den Hoek, Jordan J. Smith, John Mahoney, Jemima Spathis, Mark Moresi, Rebecca Pagano, Laura Pagano, Rafael Vasconcellos, Heather Arnott, Ben Varley, Philip Parker, Elaine Beller, Michelle M. Ashton, Sophie Catherall, Sarah Hayes, Karla Robledo, Michael de Manincor, Stephanie Alley, Corneel Vandelanotte, Mats Hallgren, Luke Wolfenden, Anthony K. Rose, Jason A. Bennie, Tim S. Olds, and Chris Lonsdale. 2024. "Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials." *BMJ*, 384: e075847.

Peters, Warren, Tea Monk Fries, Shilpa Middolu, Lauren Wilson, and Scott H. Lee. 2026. "Real-World Comparative Effectiveness of Semaglutide 2.4 mg Versus Other Commercially Available Medications for Chronic Weight Management in Adults with Obesity in the United States: A Pragmatic Clinical Study." Philadelphia, PA, USA. Presentation CTI, Code P6. Published in Value in Health, Volume 29, Issue S6.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies*, 90(5): 2555–2591.

Roth, Jonathan. 2022. "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends." *American Economic Review: Insights*, 4(3): 305–322.

Rubino, Domenica M., Frank L. Greenway, Usman Khalid, Patrick M. O'Neil, Julio Rosenstock, Rasmus Sørrig, Thomas A. Wadden, Sean Wharton, and John P. H. Wilding. 2024. "Effect of Once-Weekly Semaglutide on Patient-Reported Outcomes in Adults with Overweight or Obesity: The STEP 1 to 4 Trials." *Diabetes, Obesity and Metabolism*, 26(6): 2129–2138.

Wadden, Thomas A., Greg K. Brown, Camilla Egebjerg, Olga Frenkel, Bryan Goldman, Robert F. Kushner, Barbara M. McGowan, Maria Overvad, and Ken Fujioka. 2024. "Psychiatric Safety of Semaglutide for Weight Management in People Without Known Major Psychopathology: Post Hoc Analysis of the STEP 1, 2, 3, and 5 Trials." *JAMA Internal Medicine*, 184(11): 1290–1300.

Wilding, John P. H., Rachel L. Batterham, Salvatore Calanna, Melanie Davies, Luc F. Van Gaal, Ildiko Lingvay, Barbara M. McGowan, Julio Rosenstock, Marie T. D. Tran, Thomas A. Wadden, Sean Wharton, Koutaro Yokote, Niels Zeuthen, and Robert F. Kushner. 2021. "Once-Weekly Semaglutide in Adults with Overweight or Obesity." *New England Journal of Medicine*, 384(11): 989–1002.