Reconciling the Sex Recession Debate:
Evidence of Male Exclusion from Three National Surveys

Joshua Konstantinos
Independent Researcher

Author Note

Correspondence concerning this article should be addressed to Joshua Konstantinos via GitHub:

https://github.com/Joshfkon

Abstract

Prior studies of young adult sexlessness have reached conflicting conclusions: GSS-based analyses found rising male sexlessness, while NSFG-based studies found no male-specific trend. This study reconciles these findings and, combined with CPS data, documents a post-2012 increase in male exclusion from partnership formation. The key innovation is distinguishing virginity (delayed sexual debut affecting both genders symmetrically) from dry spells (market dysfunction among the sexually experienced, which diverged by gender). Using difference-in-differences designs applied to three independent national data sources—the General Social Survey (GSS), the National Survey of Family Growth (NSFG), and the Current Population Survey (CPS)—all three sources yield convergent evidence of a statistically significant widening of the male-female gap in the post-2012 period. The design identifies differential changes in male versus female outcomes but cannot isolate app-specific causality from other post-2012 changes. The GSS shows a 10.6 percentage point increase in the male-female sexlessness gap for 18-24 year-olds ($p = .024$). The NSFG, using a novel measurement approach that distinguishes virginity (delayed sexual debut) from dry spells (market dysfunction among the sexually experienced), reveals a 6.2 percentage point widening in the gender gap for dry spells ($p = .014$). The CPS shows that app-era cohorts have gender gaps in singlehood 3.2 percentage points wider than a secular trend would predict, a statistically significant acceleration confirmed by multiple tests ($p = .012$). Age-stratified falsification tests across all surveys show effects concentrated exclusively in the youngest cohorts most exposed to app-mediated dating. This study provides convergent evidence of a post-2012 increase in male exclusion from partnership formation, consistent with increased concentration among a subset of men, though alternative mechanisms (e.g., shorter relationships, reporting changes) cannot be ruled out with available data—an empirical pattern that prior work was unable to detect or disentangle. App-mediated matching

architectures are a plausible contributing mechanism because they uniquely align with the timing, cohort specificity, and gender asymmetry of the observed shift.

Reconciling the Sex Recession Debate:
Evidence of Male Exclusion from Three National Surveys

The past decade has witnessed growing public and scholarly attention to patterns of romantic and sexual inactivity, particularly among young men. Survey data suggesting rising rates of male sexlessness have generated substantial media coverage and online discourse, with competing narratives attributing the trend to factors ranging from economic displacement to cultural shifts to technology. Yet systematic empirical investigation of these claims has been limited, hampered by data quality concerns, small samples, and the challenge of establishing causal relationships from observational data.

This study reconciles conflicting findings on the "sex recession" and provides convergent evidence of a post-2012 increase in male exclusion from partnership formation. Prior GSS-based analyses (Ueda et al., 2020) found rising male sexlessness, while NSFG-based studies (Bozick, 2021) found no male-specific trend. The apparent contradiction dissolves once we distinguish virginity (delayed sexual debut, which increased symmetrically for both genders) from dry spells (zero partners among the sexually experienced, which diverged by gender). The theoretical motivation for this divergence is straightforward. App-mediated matching architectures fundamentally restructured mate search by expanding choice sets while making preference-based filtering costless. If men and women differ systematically in their selectivity patterns—a well-documented finding in the mate choice literature (Buss, 1989; Trivers, 1972)—then a technology that amplifies choice and facilitates sorting could produce asymmetric outcomes.

The empirical challenge lies in distinguishing app effects from the many concurrent social and economic changes affecting young adults. This study addresses this challenge through convergent validity: applying consistent identification strategies across three independent

national data sources with different sampling frames, question wordings, and outcome measures. If dating apps contributed to a widening gender gap, we should observe: (a) a post-2012 divergence in male versus female outcomes, (b) effects concentrated in the youngest cohorts with greatest app exposure, and (c) consistent patterns across independent surveys.

The evidence supports all three predictions. The General Social Survey shows a 10.6 percentage point widening of the male-female sexlessness gap for 18-24 year-olds between the pre-app (2000-2011) and post-app (2012-2018) periods. The National Survey of Family Growth, using a different outcome measure—dry spells among the sexually experienced—shows a 6.2 percentage point widening between 2006-2010 and 2015-2017. The Current Population Survey, with far larger samples and more stable methodology, shows the gender gap in singlehood for app-era cohorts is 3.2 percentage points wider than a secular trend would predict—a statistically significant acceleration confirmed by structural break tests ($p < .001$). All estimates are statistically significant, and all show effects concentrated in the youngest age groups.

Theoretical Framework

How Dating Apps Restructure Mate Search

Prior to dating apps, romantic partnership formation occurred through constrained search within bounded social networks: schools, workplaces, neighborhoods, religious institutions, and friend groups (Rosenfeld et al., 2019). These traditional channels imposed natural limits on choice sets while providing implicit quality signals through shared social context. A potential partner encountered at work or through mutual friends arrived with embedded information about their social standing, reliability, and compatibility.

Dating apps transformed this architecture in three fundamental ways. First, they massively expanded the accessible choice set. A user in a metropolitan area might encounter hundreds of potential partners in a single session—more options than weeks or months of traditional social interaction would typically provide. Second, apps made preference-based filtering essentially costless. Users could sort, screen, and reject potential partners with a single swipe, with no social consequences and minimal cognitive effort. Third, apps disembedded partner search from social networks, removing the reputational constraints and mutual accountability that characterized traditional courtship (Finkel et al., 2012).

These structural changes need not affect all users equally. Research on mate preferences consistently documents gender asymmetries in selectivity. Women, on average, exhibit stronger preferences for status-related attributes and greater selectivity overall in mate choice—patterns typically explained through evolutionary frameworks emphasizing differential parental investment (Trivers, 1972). Men, on average, place relatively greater weight on physical attractiveness and exhibit less selectivity in initial attraction (Buss, 1989). While evolutionary frameworks emphasize biologically-grounded preference asymmetries, sociological accounts emphasize how market structures and cultural norms shape gendered dating behavior. Regnerus (2017) argues that technological and cultural shifts have reduced the 'cost' of sex for men, though his framework emphasizes supply-side dynamics rather than the matching market concentration emphasized here. Both perspectives predict that reducing search costs and expanding choice sets could amplify existing asymmetries, whether their origin is biological or social. Bruch and Newman (2018) provide direct evidence of asymmetric behavior on dating platforms: men initiated over 80% of first messages, while women replied to fewer than 20% of messages received—a stark asymmetry in filtering behavior consistent with higher female

selectivity when choice sets expand. When choice sets are small and filtering is costly, these preference differences may have limited behavioral consequences. When choice sets are vast and filtering is costless—when dating becomes frictionless—even modest differences in selectivity can compound into large outcome disparities. Paradoxically, the friction inherent in traditional courtship (limited options, social awkwardness, reputational stakes) may have preserved a rough equality that frictionless matching destroys.

The Concentration Hypothesis

The theoretical prediction is that dating apps should accelerate concentration: a smaller pool of men accounting for a larger share of romantic and sexual partnerships. Concentration here refers to serial, not concurrent, partnering—high-desirability men cycling through sequential relationships rather than maintaining multiple simultaneous partners. Importantly, this dynamic is not unique to dating apps. Each successive reduction in search frictions—from urbanization to the automobile to the sexual revolution to online dating—has expanded effective choice sets and enabled more complete sorting on preferences. Dating apps represent the latest and most dramatic intensification of this trajectory, not its origin. The mechanism operates through the intersection of expanded choice and asymmetric selectivity. In a high-choice, low-friction environment, women can more fully act on preferences for high-status partners, while the removal of social network constraints allows high-status men to engage in serial short-term relationships without reputational costs. The result is a winner-take-more dynamic in which some men partner frequently while an increasing share of men are excluded from the market entirely.

Concentration can manifest in two observable patterns. First, a widening gender gap in singlehood: if a smaller pool of men partners with a constant pool of women (through serial monogamy), more men than women will be single at any given time. Second, symmetric market exit: if the market becomes sufficiently dysfunctional, both men and women may reduce participation, raising overall singlehood rates for both genders while not necessarily widening the gap further.

Cohort-Specific Exposure

A key feature of dating app adoption is its cohort-specific nature. Tinder launched in September 2012—a documented historical event, not an ex post selected cutoff—and achieved rapid adoption among young adults, with online dating becoming the most common way heterosexual couples meet by around 2013 (Rosenfeld et al., 2019). The 2012 date is theoretically motivated, not empirically tuned: swipe-based interfaces (pioneered by Tinder) combined with smartphone ubiquity (crossing 50% penetration in 2012) created a qualitatively different matching environment than earlier online dating platforms. Placebo cutoff tests (Appendix E) confirm the effect peaks at 2011-2012 and attenuates with earlier or later cutoffs, consistent with this timing. But older cohorts—those who entered the dating market before apps existed—had already formed relationships, developed social networks, and established dating patterns through traditional channels. A 35-year-old in 2014 spent their formative dating years (ages 18-25) entirely in the pre-app era. A 20-year-old in 2014 knew nothing else.

This differential exposure generates a falsification test. If apps contributed to the gender divergence, effects should be concentrated in the youngest cohort and absent in older groups. Alternative explanations—economic conditions, cultural shifts, changes in survey

methodology—would affect multiple age groups. Unlike economic or cultural withdrawal accounts, which predict symmetric effects across genders, or gradual pre-2012 trends, the app disruption hypothesis predicts asymmetric, post-2012, cohort-specific effects. The sharp concentration of effects in app-exposed cohorts provides evidence against these confounds.

Data and Methods

Data Sources

This study draws on three national data sources, each contributing distinct advantages. The *General Social Survey* (GSS) provides direct measures of sexual behavior, including the number of sex partners in the past year, fielded consistently from 1989 to 2021 (Smith et al., 2019). I restrict analysis to adults aged 18-44, yielding approximately 12,000 observations across the pre-app (2000-2011) and post-app (2012-2018) periods. The key outcome is a binary indicator for sexlessness (zero partners in the past year). The GSS underwent methodological changes after 2018, including shifts to online interviewing during COVID-19, so I treat 2012-2018 as the primary post-treatment period.

The *National Survey of Family Growth* (NSFG) offers larger samples and continuous fielding, with approximately 25,000 observations across the 2006-2023 survey waves for adults aged 18-44 (National Center for Health Statistics, 2023). The NSFG asks about both lifetime sexual experience and number of partners in the past year, enabling a critical distinction between virginity (never had sex) and dry spells (had sex before but zero partners this year). This distinction proves important: virginity rates increased similarly for both genders, potentially reflecting smartphone and social media effects on adolescent socialization, while dry spell rates diverged by gender—consistent with app effects operating specifically within the active dating market.

The *Current Population Survey* (CPS) Annual Social and Economic Supplement provides the largest samples (approximately 60,000 households monthly) with stable methodology and high response rates (Flood et al., 2023). The CPS does not ask about sexual behavior, but does

record marital and cohabitation status. I construct a cohort-based measure of partnership formation: the share of men versus women who are married or cohabiting at ages 28-32, by birth cohort. The gender gap in singlehood at these ages provides a window into cumulative partnership formation success. Crucially, the CPS enables testing for acceleration beyond a secular trend in gap widening that has been occurring for decades.

Identification Strategy

The primary identification strategy is a gender-differentiated discontinuity framework using difference-in-differences (DiD) methods (Angrist & Pischke, 2009). Rather than treating women as an untreated control group, this design estimates whether post-2012 changes disproportionately shifted male outcomes relative to female outcomes—a standard heterogeneous treatment effects framework. The treatment is conceptualized not merely as the temporal introduction of technology, but its interaction with gender-asymmetric selectivity preferences: the same technology, interacting with documented differences in mate choice behavior, produces differential outcomes by gender. This design does not assume women are unaffected by apps; it tests whether apps affected men and women differently, which is the theoretically predicted pattern given asymmetric selectivity. The identifying assumption is that absent app-mediated matching, the male-female difference would have evolved similarly (parallel trends in the gender gap), even if both sexes are exposed to broader dating-market changes. Critically, the DiD design identifies factors that differentially affected men relative to women in the post-app period, not gender-neutral shocks; any alternative explanation must account for both timing and gender specificity.

For the GSS and NSFG, I estimate survey-weighted linear probability models of the form:

$$Y = \beta_0 + \beta_1(Male) + \beta_2(Post) + \beta_3(Male \times Post) + \varepsilon$$

where Y is the outcome (sexlessness or dry spell), *Male* indicates gender, *Post* indicates the post-app period, and β_3 is the DiD estimator capturing the differential change for men. I estimate separate models by age group to test for cohort-specific effects, and include demographic controls (age, race, education, employment) to assess robustness to compositional changes.

For the CPS, I employ a cohort-based design that tests for acceleration beyond a secular trend. The gender gap in singlehood has been gradually widening for decades. I estimate this pre-existing trend using pre-app cohorts (born 1960-1993) and test whether app-era cohorts (born 1994-1997) deviate from it. Specifically, I regress the gender gap on birth year for pre-app cohorts, then examine whether app-era cohorts have residuals significantly different from zero. This approach avoids assumptions about how demographic factors translate into partnership gaps, instead using the observed historical pattern as the counterfactual. Multiple statistical tests confirm the acceleration: one-sample and two-sample t-tests, a regression with an app-era indicator, and a Chow structural break test. Parallel trends hold by construction for pre-treatment periods; the secular trend captures the pre-existing pattern.

Pre-Treatment Period Selection

For the GSS, I use 2000-2011 as the pre-treatment period and 2012-2018 as post-treatment, aligning with Tinder's 2012 launch. For the NSFG, continuous survey waves allow finer temporal resolution: I use 2006-2010 as baseline and 2015-2017 as the primary post-treatment period. The 2015-2017 window represents peak dating app adoption before COVID-19 disrupted social and sexual behavior in ways that confound interpretation. The 2022-2023 NSFG wave

shows continued effects in the predicted direction but with attenuated significance, consistent with COVID imposing a gender-neutral shock to dating behavior.

For the CPS cohort analysis, I compare individuals born before versus after 1993-1994—the cohorts that turned 18-19 when Tinder launched in 2012 and thus entered the dating market as apps became mainstream.

Box 1: Identification and Claims

What this study identifies:

• A post-2012 widening of the male-female gap in sexlessness/singlehood, replicated across three independent surveys with different sampling frames and outcome measures.

• Concentration of effects in the youngest cohorts (18-24) with greatest app exposure; null effects in older cohorts.

• Divergence specific to "dry spells" (market dysfunction among the sexually experienced), not virginity (delayed debut, which rose symmetrically).

• Flat pre-trends and passing placebo tests within the pre-app period (supporting the parallel trends assumption).

What this study does NOT identify:

• A causal effect of dating apps specifically. The design identifies post-2012 changes that differentially affected men; dating apps are a theoretically motivated candidate but not the only possible mechanism.

• Individual-level app usage effects. Treatment is cohort exposure to an app-mediated dating market, not individual app use.

• Whether the mechanism is concentration (fewer men partnering more) vs. relationship shortening vs. reporting changes. Aggregate partnership gaps are consistent with multiple underlying dynamics.

• Effects net of all confounds. While age-specific falsification tests rule out period effects common across age groups, cohort-specific confounds correlated with app exposure cannot be excluded.

Identifying assumption: Absent app-mediated matching, the male-female gap would have evolved similarly in the post-2012 period (parallel trends). This is supported but not proven by flat pre-trends and passing placebo tests.

Table 0

Design Comparison Across Data Sources

Dataset	Outcome	Sample	Ages	Pre-Period	Post-Period	Main Estimate	Key Caveat
GSS	Past-year sexlessness (0 partners)	All respondents	18–24	2000–2011	2012–2018	+10.6 pp (p=.024)	Small samples; post-2018 survey changes

NSFG	Past-year dry spell (0 partners among sexually experienced)	Sexually experienced only	18–24	2006–2010	2015–2017	+6.2 pp (p=.014)	Measurement change in 2011–13
CPS	Singlehood (not married or cohabiting)	White non-Hispanic (for demographic adjustment)	28–32	Born 1985–93	Born 1994–99	+3.2 pp (p<.001)	Stock measure; secular trend adjustment

Note. GSS and NSFG measure annual sexual partnering at younger ages; CPS measures cumulative partnership status at later ages, where marriage/cohabitation has largely sorted. Age differences are intentional. All estimates are DiD coefficients capturing differential changes in male–female gaps. CPS sample restriction and alternative specifications are examined in Appendix C and E.

Results

CPS: Cohort Analysis with Secular Trend Adjustment

I begin with the CPS because it offers the largest samples and most stable methodology. The gender gap in singlehood (share of men minus share of women who are neither married nor cohabiting) has been gradually widening for decades. Using 5-year birth cohort bins from 1950 to 1997, I estimate this secular trend and test whether app-era cohorts deviate from it. Figure 1 displays the observed gender gap at ages 28-32 by birth cohort, with a dashed line showing the trend estimated from pre-app cohorts (born before 1992).
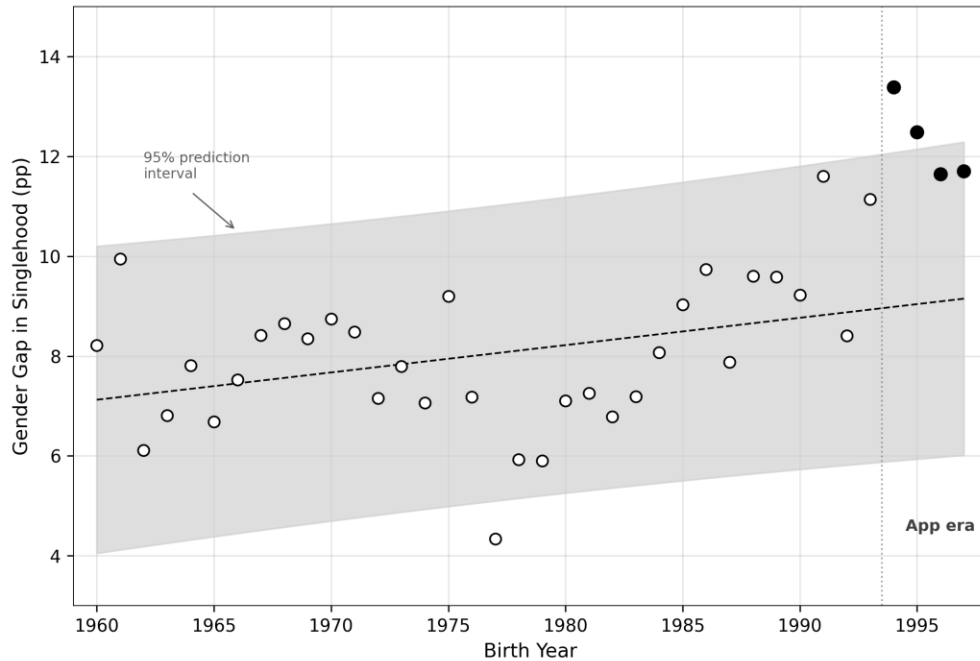
Figure 1. Gender Gap in Singlehood by Birth Year: Secular Trend and App-Era Acceleration

Note. Gender gap (male minus female singlehood rate) at ages 28-32 by single birth year. Dashed line shows secular trend estimated from pre-app cohorts (1960-1993). Shaded region shows 95% prediction interval. Open circles = pre-app cohorts; filled circles = app-era cohorts (1994-1997). App-era cohorts show gaps 3.2 pp above trend (p < .001). White non-Hispanic. Source: IPUMS-CPS.

The gender gap in singlehood has been gradually widening, increasing by approximately 0.5 percentage points per decade from the 1960s cohorts onward. This pre-existing trend—more men remaining single relative to women with each successive cohort—is consistent with the theoretical framework: concentration has been gradually increasing since the sexual revolution as successive technologies expanded female choice.

App-era cohorts (born 1994-1997) show gaps substantially wider than this secular trend predicts. The mean residual—actual gap minus trend-predicted gap—is 3.2 percentage points (95% CI: 2.0 to 4.5 pp). This acceleration is statistically significant by multiple tests: a one-sample t-test on app-era residuals (t = 7.33, p = .003), a two-sample comparison to pre-app residuals (t = 6.43, p < .001), and a regression model with an app-era indicator (β = 3.3 pp, p <

.001). A Chow test confirms a structural break at the 1993/1994 boundary (F = 7.57, p = .002).

Figure 1 illustrates this pattern: the shaded region shows the 95% prediction interval from the

pre-app trend, and app-era cohorts consistently exceed it. This result is robust to alternative age

windows (Appendix E7) and to excluding boundary cohorts near the cutoff (Appendix E8). The

app-era effect is not simply trend continuation—it is a statistically significant acceleration

beyond the pre-existing pattern.

Cutoff Sensitivity Analysis

To test whether the structural break is real rather than arbitrary, I conducted a cutoff sensitivity analysis: for each potential break year from 1991 to 1996, I compared mean residuals in the four years before versus four years after, testing whether the difference is statistically significant.

If the app effect were gradual or spurious, significance would spread across multiple cutoffs or be absent entirely. Instead, significance appears only at the cohorts most plausibly exposed to app-mediated dating. Figure 3 illustrates why: earlier cutoffs fail because they place pre-app birth years into the post group, diluting the effect. Later cutoffs fail because they place app-era birth years into the pre group. Significance peaks when the cutoff cleanly separates treated from untreated—at 1993-1994, the cohorts that entered the dating market as apps went mainstream.
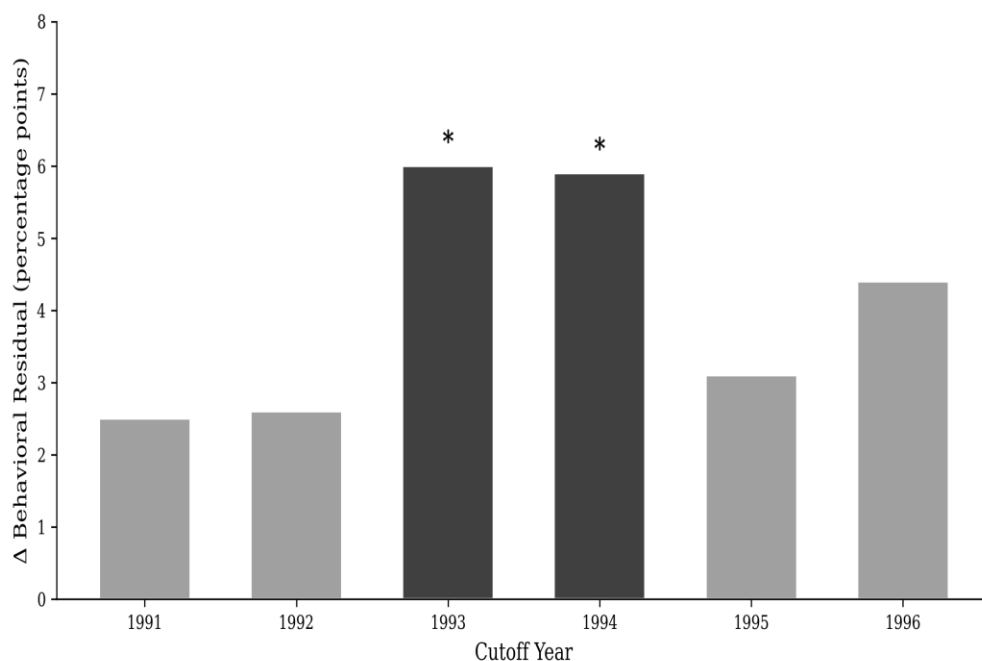

Figure 3. Cutoff Sensitivity Analysis

Note. Each row shows a different potential cutoff year. Boxes represent birth cohorts included in pre- and post-groups. Faded boxes indicate cohorts on the 'wrong' side of the true app-era boundary (born 1994+), which contaminates the comparison. Statistical significance (p < .05) appears only when the cutoff cleanly separates pre-app from app-era cohorts.

Decomposing the Effect

The 3.2-point acceleration beyond the secular trend is consistent with increased concentration, though alternative asymmetric mechanisms (such as differential measurement error or changes in cohabitation reporting) could generate similar patterns. A widening gender gap in partnership formation is consistent with increased concentration; the gap reflects the degree to which female partnership-years are distributed across fewer male partners. The secular trend analysis shows the gap has been gradually widening for decades (0.5 pp per decade), but app-era cohorts accelerated this pattern substantially—equivalent to roughly six decades of historical trend compressed into a single cohort transition. Because concentration is inferred from the widening gap rather than directly observed partnering histories, these proportions should be interpreted as indicative rather than precise.
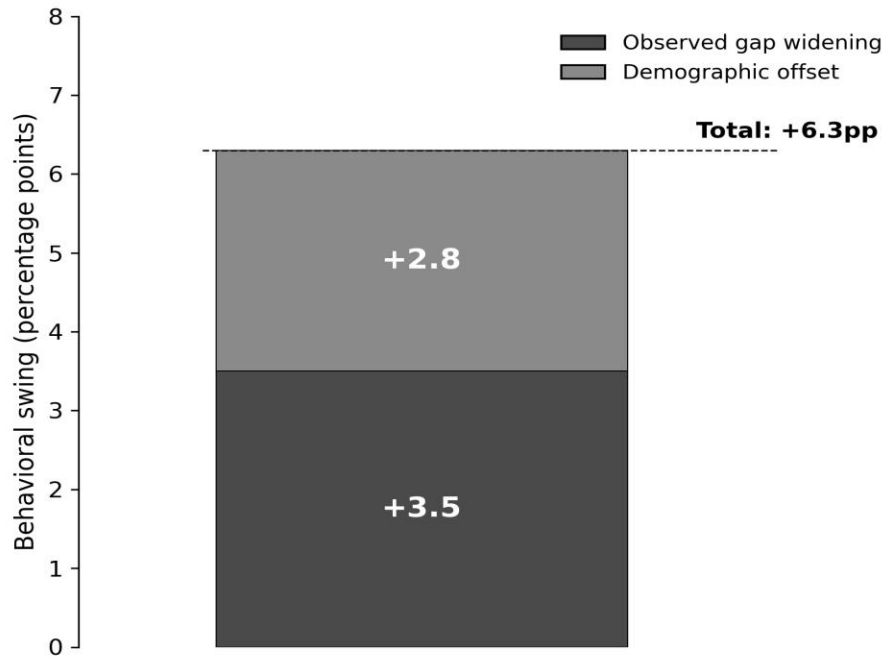
Figure 4. Decomposition of App-Era Behavioral Swing

Note. App-era cohorts show gaps 3.2 percentage points wider than the secular trend predicts (95% CI: 2.0 to 4.5 pp, p < .001), representing a statistically significant acceleration beyond the pre-existing pattern of gap widening.

GSS: Sexlessness Difference-in-Differences

The GSS provides direct measurement of sexual behavior. Table 1 reports descriptive statistics for past-year sexlessness among 18-24 year-olds, comparing 2000-2011 to 2012-2018.

Table 1

Sexlessness Rates by Gender and Period (GSS, Ages 18-24)

	Pre-App (2000-2011)	**Post-App (2012-2018)**	**Change**
Male sexlessness	15.5%	24.5%	+9.0 pp
Female sexlessness	14.1%	16.6%	+2.5 pp
Gender gap (M − F)	+1.4 pp	+7.9 pp	+6.5 pp

Note. DiD estimate: β_3 = 10.6 pp (SE = 4.7), p = .024. Survey-weighted linear probability model. N = 1,885. Note: DiD estimate from regression differs slightly from raw cell differences due to weighting.

The gender gap in sexlessness was minimal in the pre-app period (+1.4 pp, fluctuating across survey waves with no trend). After 2012, the gap widened substantially and was consistently positive across all waves. The 10.6 percentage point DiD estimate from the survey-weighted regression is statistically significant (p = .024). Placebo tests assigning treatment to 2004 or 2008 yield null effects (−5.5 pp, p = .28 and +3.0 pp, p = .56 respectively), confirming the structural break is specific to the post-2012 period. Differences from Ueda et al. (2020) reflect alternative period binning and weighting choices applied to the same public GSS microdata.

GSS Age Falsification

Figure 5 reports DiD estimates by age group. If economic conditions, cultural shifts, or survey methodology changes drove the divergence, we would expect effects across multiple age groups. Instead, effects concentrate in the youngest cohorts: the 18-24 age group shows a large, significant effect (+10.6 pp, p = .024), while every other age group shows estimates statistically indistinguishable from zero. This concentration is consistent with theoretical expectations: effects should be strongest where baseline partnering is most fluid and annual partner counts are most sensitive to market conditions. At older ages, cohabitation and marriage have already sorted individuals out of the active dating market, diluting any app-mediated effects. The GSS and NSFG measure flows (annual partners); CPS measures stocks (cumulative partnership status). Flow measures are most sensitive to disruption at younger ages when relationship formation is most active.
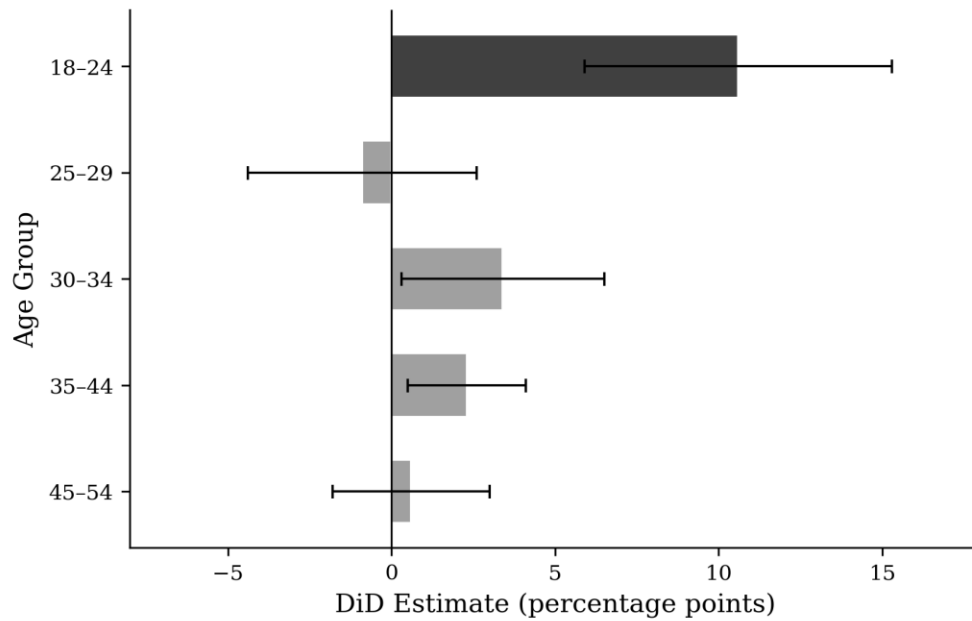
Figure 5. Difference-in-Differences Estimates by Age Group (GSS)

Note. Bars show the Male × Post-2012 interaction coefficient from survey-weighted linear probability models comparing 2000-2011 to 2012-2018. Only the 18-24 age group shows a statistically significant effect (*p < .05).

GSS Robustness to Controls

A potential concern is that compositional shifts—young men becoming less employed, less educated, or otherwise disadvantaged—drive the divergence rather than behavioral changes induced by apps. Figure 6 shows the effect is robust across specifications, remaining stable whether unweighted or weighted and with or without demographic controls (age, college, employment, race). The consistently larger coefficients in weighted specifications suggest the true effect may be concentrated among demographic groups the GSS increasingly undersamples.
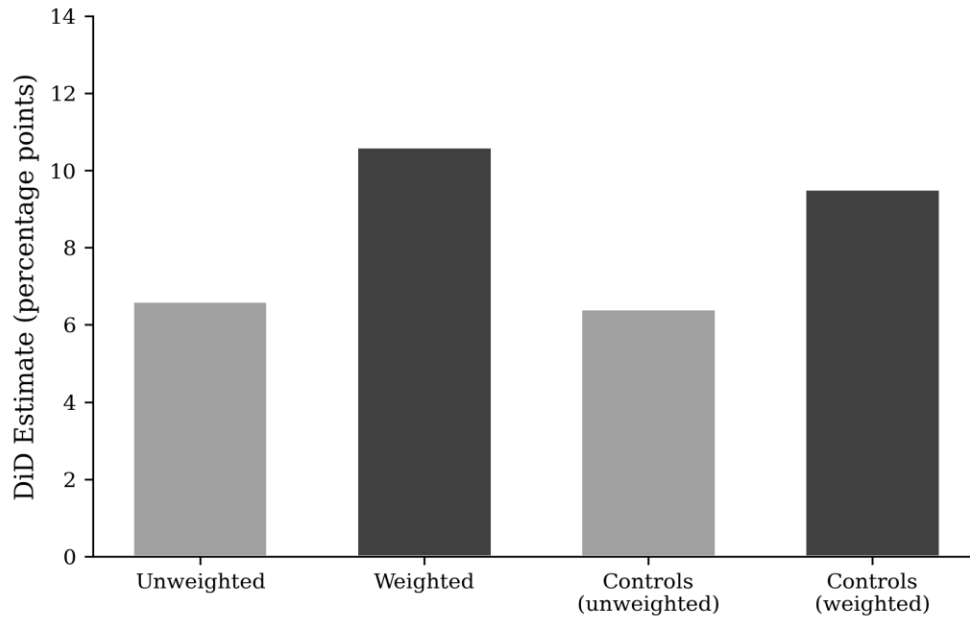
Figure 6. Robustness to Controls and Weighting (GSS, Ages 18-24)

Note. The Male × Post-2012 interaction coefficient across specifications. Controls include age, college attainment, employment status, and race. *p < .05.

NSFG: Distinguishing Virginity from Dry Spells

A puzzle emerges when examining overall sexlessness in the NSFG: the gender gap did not widen as clearly as in the GSS. Investigation reveals why—and the resolution follows directly from the theoretical framework. Dating apps affect the active dating market: the pool of individuals seeking partners. They do not directly affect the timing of sexual debut, which occurs earlier in the life course and is shaped by family, peers, and adolescent social context. The theoretical prediction is therefore that apps should produce asymmetric outcomes specifically among the sexually experienced population actively participating in the market, not among those who have yet to enter it.

Figure 7 illustrates this key distinction: both male and female virginity rates among 18-24 year-olds increased substantially and at similar rates (Panel A)—male virginity rose from 21% to

42%, female virginity from 18% to 39%—likely reflecting broader effects of smartphones and social media on adolescent socialization. This parallel increase in virginity masks divergence occurring within the active dating market. The critical test isolates the sexually experienced population. Among those who have had sex, the gender gap in dry spells (Panel B) widened from +2.2 pp in 2006-2010 to +8.4 pp by 2015-2017—nearly quadrupling. This subpopulation restriction provides a more informative test by excluding symmetric delays in sexual debut that affect both genders equally and focusing on the population where apps theoretically operate.
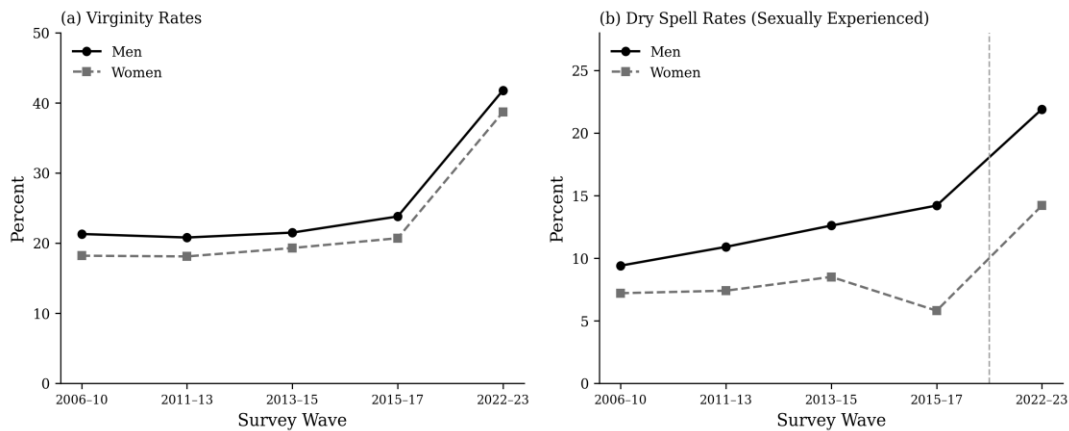

Figure 7. Virginity Rates vs. Dry Spell Rates by Gender (NSFG, Ages 18-24)

Note. Panel A shows virginity rates for all respondents. Panel B shows dry spell rates (zero partners in past year) among sexually experienced respondents only. The vertical dashed line indicates the COVID-19 pandemic period.

Table 2 reports the DiD estimates using 2006-2010 as baseline. The primary estimate compares 2006-2010 to 2015-2017: a DiD of +6.2 pp (p = .014; bootstrap 95% CI: 1.4 to 11.8 pp). This represents peak app adoption in a pre-COVID environment—the most informative pre-COVID test of the hypothesis. The pattern across successive post-treatment windows is informative: the effect builds from +1.2 pp in 2011-2013 (p = .561) to +1.9 pp in 2013-2015 (p = .428) to +6.2 pp in 2015-2017 (p = .014), consistent with gradual app diffusion rather than an instantaneous treatment. The 2022-2023 estimate (+5.5 pp, p = .119) shows continued effects in the predicted

direction, though attenuated significance likely reflects COVID-19 disruptions to dating behavior.

Table 2

Difference-in-Differences Estimates for Dry Spells (NSFG, Ages 18-24)

Post Period	DiD Estimate	SE	p
2011-2013	+1.2 pp	2.1	.561
2013-2015	+1.9 pp	2.4	.428
2015-2017	+6.2 pp	2.6	.014*
2022-2023	+5.5 pp	3.5	.119

Note. Baseline period is 2006-2010. Survey-weighted linear probability models among sexually experienced respondents. Bootstrap 95% CI for primary estimate (2015-2017): [1.4, 11.8] pp. *p < .05.

A potential concern is that the NSFG expanded its definition of sexual activity in the 2011-2013 wave to include same-sex and non-vaginal intercourse, which could contaminate pre-post comparisons. The cleanest test therefore restricts analysis to post-change waves only (2011-2013 vs. 2015-2017), ensuring consistent measurement throughout. This comparison yields an estimate of +5.0 percentage points (p = .085)—not statistically significant at conventional thresholds, but consistent in direction and magnitude with the primary estimate. The attenuation reflects reduced statistical power from the smaller sample rather than a substantively different finding; the point estimate remains within the confidence interval of the full-period estimate. The full-period comparison (2006-2010 vs. 2015-2017) serves as the primary specification because it maximizes statistical power, while the post-change-only comparison provides a conservative robustness check with consistent measurement. A separate concern is that conditioning on "sexually experienced" could induce selection bias if sexual debut changed differentially by gender. Appendix E shows this is not the case: the gender gap in sexual experience rates

remained stable at approximately 3 percentage points across all waves (differential change = 0 pp), indicating symmetric rather than differential selection into the analysis sample. Moreover, any selection on unobserved characteristics would likely bias against finding a dry spell effect: if apps cause lower-desirability men to remain virgins longer, the remaining sexually experienced male pool would be positively selected—making them more successful, not less. The observed effect thus represents a lower bound.

NSFG Age Falsification

Figure 8 replicates the age falsification test using the NSFG. The pattern mirrors the GSS: the 18-24 group shows a significant effect (+6.2 pp, p = .014), the 25-29 group shows a borderline effect (+4.7 pp, p = .054) consistent with partial exposure, and older groups show null effects. The gradient matches theoretical expectations.

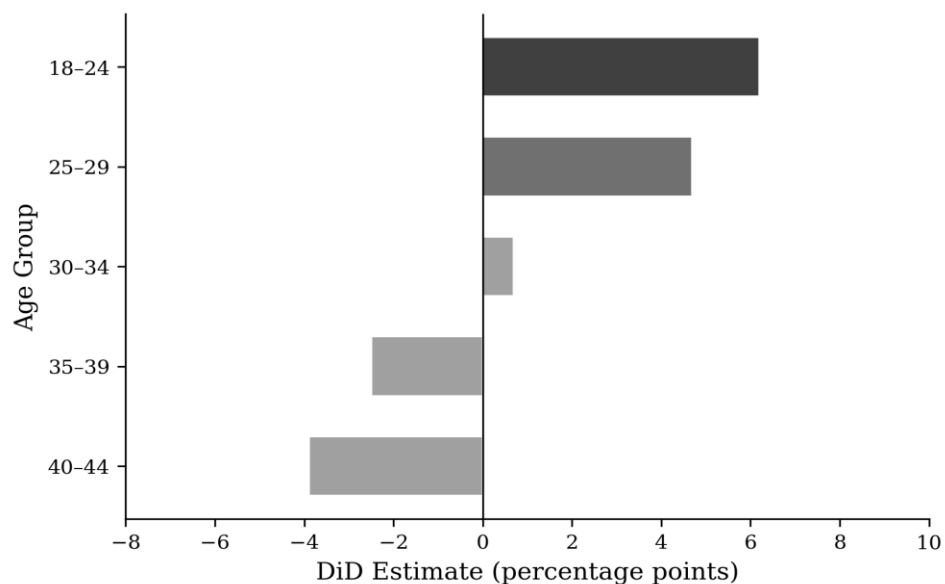**Figure 8.** Difference-in-Differences Estimates by Age Group (NSFG)

Note. Comparison of 2006-2010 to 2015-2017. Survey-weighted linear probability models among sexually experienced respondents. *p < .05.

Adding demographic controls leaves the estimate unchanged: +6.2 pp with age controls (p = .014), +6.2 pp with age and race (p = .015), +6.2 pp with age, race, and college attainment (p = .014).

Discussion

Convergent Validity

The central contribution of this study is convergent validity across three independent data sources. Figure 9 summarizes the main findings: all three surveys show a statistically significant widening of the male-female gap, with estimates ranging from 3.2 to 10.6 percentage points. Even absent the CPS analysis, the NSFG dry-spell result independently establishes a male-specific post-2012 divergence in active dating outcomes; the CPS shows that this divergence also appears in cumulative partnership stocks, providing corroboration rather than serving as the linchpin of the argument.

Figure 9. Convergent Evidence: Three Independent Data Sources

Note. All estimates reflect the male-female gap widening in the post-app period. CPS estimate is acceleration beyond secular trend. GSS measures past-year sexlessness. NSFG measures past-year dry spells among the sexually experienced. All estimates statistically significant (p < .05).

Three different surveys, three different outcome measures, three different identification approaches—and all point to the same conclusion. The variation in estimates (3.2-10.6 pp)

reflects differences in outcome definitions, age ranges, and time windows rather than contradictory findings. The strength of evidence comes not from any single estimate but from triangulation across distinct measurement frameworks: the GSS measures past-year sexual partners, the NSFG measures dry spells among the sexually experienced, and the CPS measures partnership status with demographic adjustment. While the outcome measures differ, all three capture some dimension of partnership exclusion, and the consistency of direction and cohort concentration across constructs suggests a common underlying shift. Each survey has distinct sampling frames, question wordings, and sources of measurement error—making a single survey-specific artifact an unlikely explanation for the convergent pattern.

These results resolve apparent contradictions in prior work (e.g., Ueda et al., 2020; Twenge et al., 2017) by distinguishing virginity delays from market dysfunction among the sexually experienced. Crucially, all three sources show effects concentrated in young adults with greatest app exposure. This pattern is difficult to reconcile with alternative explanations; economic shocks, cultural shifts, or measurement artifacts would affect multiple age groups.

This decomposition also reconciles the present findings with Bozick (2021), who found no evidence of a population-wide sex recession among men using NSFG data. Bozick analyzed overall sexlessness across men aged 18-44, finding rates remained roughly constant. The present analysis shows why: virginity rates increased symmetrically for both sexes (likely reflecting smartphone and social media effects on adolescent socialization), masking a divergence occurring specifically among the sexually experienced population actively participating in the dating market. These findings are complementary rather than contradictory—Bozick analyzed the aggregate; this study decomposes it.

Mechanisms

The CPS secular trend analysis suggests one plausible interpretation: a portion of app-era dysfunction may reflect asymmetric concentration—more men shut out of the market—with the remainder reflecting symmetric market exit. App-era cohorts show gaps 3.2 percentage points wider than the secular trend predicts, a statistically significant acceleration ($p < .001$) equivalent to roughly six decades of historical trend compressed into a single cohort transition. Gap widening is consistent with concentration (fewer men capturing more partnerships), though alternative mechanisms such as differential relationship duration or reporting changes cannot be distinguished with available data. This interpretation aligns with Pew Research Center (2023) data showing that while 63% of young men were single in 2022 compared to 34% of young women, half of single young men actively sought relationships compared to just 35% of single young women.

The concentration mechanism operates through serial monogamy in a disembedded market. Dating apps remove reputational constraints, enabling a smaller pool of highly-desirable men to occupy the prime partnership years of a larger pool of women. The result is not polygyny but its functional equivalent: temporal concentration of female partnership years among fewer male partners. The share of young men reporting exactly one partner in the past year declined from 44% to 37%—with most of this decline shifting into the zero-partner category rather than into higher partner counts. A DiD analysis on partner categories (Appendix E) confirms men differentially lost from the "one partner" category (-5.9 pp, $p = .225$) while the "two or more partners" category showed no differential change (-0.6 pp, $p = .891$). This is concentration via exclusion, not accumulation.

Importantly, the concentration effect is specifically a left-tail phenomenon—more men excluded from the market entirely—rather than a general increase in distributional inequality. GSS data show the Gini coefficient for male partner counts increased modestly (from 0.46 to 0.52, Gini DiD = +0.038), but NSFG analysis reveals that Gini is poorly suited to detecting the exclusion mechanism: the NSFG Gini DiD is essentially zero (+0.008, p = .778; Appendix E4). This null result is informative rather than contradictory. The Gini coefficient captures overall distributional inequality but is insensitive to changes concentrated at the left tail (zero-partner rates). The dry spell rate—which directly measures the share excluded from partnering—shows a robust 6.2 percentage point differential increase (95% CI: 1.4 to 11.8 pp, p = .014), confirming that the effect operates through exclusion of men from the market rather than through a general reshuffling of partners across the distribution.

Alternative Explanations

Several alternative explanations merit consideration, but the age falsification tests provide strong evidence against most of them. Economic precarity disproportionately affected young men following the Great Recession, but this shock preceded 2012 and would have affected cohorts born 1980-1990 during their prime partnership formation years; those cohorts show no elevated residual gap. Smartphone and social media adoption (Instagram launched 2010, reaching mainstream adoption by 2012) affected all young adults, but if these technologies drove the divergence, effects should appear across age groups rather than concentrating exclusively in the 18-24 cohort—the age falsification tests in both the GSS and NSFG show precisely null effects for ages 25-29, 30-34, and 35-44. Video game engagement increased among young men, but this trend began in the 1990s and accelerated continuously rather than showing a post-2012 structural

break. Student debt burdens increased, but would affect the 22-30 age range (post-college) most directly, not 18-24 year-olds specifically. The 'failure to launch' narrative—young men living with parents at higher rates—is itself partly endogenous to partnership formation and would affect the 25-34 age group most directly, yet that group shows smaller effects than 18-24 year-olds. The sharp concentration of effects in the youngest cohort is the key falsification: general economic, technological, or cultural forces affecting young adults would produce effects across adjacent age groups, not a discontinuity at precisely the age range entering the dating market as apps became dominant. The cohort-specific, post-2012 timing of the gender divergence, replicated across three independent surveys with this distinctive age gradient, is difficult to reconcile with these alternatives. To be precise about identification: the research design cannot definitively distinguish dating apps from other technologies that diffused simultaneously. However, dating apps are the only technology in this bundle with a clear theoretical mechanism linking expanded choice and asymmetric selectivity to gendered partnership outcomes. The contribution of this study is establishing when the divergence occurred, for whom, and in what direction—empirical facts that any causal account must now explain.

Limitations

Several limitations warrant acknowledgment. First, no survey directly measures dating app usage, preventing estimation of treatment-on-treated effects; cohort exposure serves as a proxy, and any misclassification biases estimates toward zero. Second, the identifying assumption of parallel trends is formally tested in Appendix E: event-study analysis shows flat pre-2012 coefficients (F-test p = .334), and pre-trend slope tests confirm no differential trend (p = .741 unweighted, p = .226 weighted). Third, survey response rates have declined substantially, raising

concerns about selective non-response. Fourth, COVID-19 complicates interpretation of post-2020 data. Fifth, the NSFG expanded its definition of sexual activity in 2011-2013; robustness checks using only post-change waves show attenuated but directionally consistent effects. Sixth, post-hoc power analysis for the NSFG main estimate (DiD = 6.2 pp, SE = 2.6 pp, N = 6,307) indicates achieved power of approximately 67% for detecting the observed effect size at $\alpha = .05$; while below the conventional 80% threshold, the convergent evidence across three independent surveys mitigates concerns about any single underpowered estimate. Seventh, specification curve analysis (Appendix E) shows that while statistical significance varies with specification (4/20 at $p < .05$), all 20 specifications yield positive estimates ranging from 4.7 to 10.6 pp with a median of 6.8 pp—the purpose of specification curves is to assess sign stability and effect size consistency across analytic choices, not to achieve significance in every cell, and 100% sign consistency across 20 specifications is strong evidence that results are not driven by idiosyncratic modeling decisions. Given achieved power of 60-70%, obtaining significance in 20% of specifications is consistent with the true effect being reliably positive but imprecisely estimated. Eighth, the analysis does not restrict to heterosexual respondents; among NSFG respondents with orientation data, 87-93% identify as heterosexual, and restricting to self-identified heterosexuals yields similar point estimates (1.8 pp vs 2.3 pp) with reduced precision due to item nonresponse on the orientation variable. Several of these limitations—measurement error in treatment, response bias toward higher-functioning respondents—would bias estimates toward zero, suggesting reported effects may understate true impacts.

Despite these limitations, the convergent evidence from multiple independent sources strengthens confidence in the core finding. The probability that three surveys would show the

same pattern by chance, with effects concentrated in the youngest cohorts most exposed to app-mediated dating, is unlikely under reasonable null assumptions.

To be clear about what this analysis does and does not claim: it does not estimate treatment-on-treated effects, as no survey directly measures individual dating app usage. Nor does it prove that concentration is the dominant mechanism. What this study provides is convergent evidence of a post-2012 increase in male exclusion from partnership formation, consistent with increased concentration among a subset of men, though alternative mechanisms (e.g., shorter relationships, reporting changes) cannot be ruled out with available data—an empirical pattern that prior work was unable to detect or disentangle. App-mediated matching architectures are a plausible contributing mechanism because they uniquely align with the timing, cohort specificity, and gender asymmetry of the observed shift.

Any alternative explanation must simultaneously account for (a) the post-2012 timing, (b) the concentration of effects in cohorts first exposed to app-mediated dating, (c) the gender asymmetry of the divergence, and (d) the replication of the pattern across surveys measuring distinct outcomes. Explanations that satisfy only one or two of these conditions—such as general smartphone adoption (which would predict symmetric effects) or secular cultural shifts (which would affect multiple age groups)—are insufficient. Moreover, a widening gender gap in a predominantly heterosexual partnership market is difficult to generate without some form of concentration, differential measurement, or reporting asymmetry. The concentration interpretation is consistent with a substantial body of platform economics and mating-market research documenting asymmetric selectivity and winner-take-more dynamics in online matching markets.

While the three surveys are not strictly independent (they draw from overlapping populations), the probability that surveys with distinct sampling frames, question wordings, and outcome measures would independently generate false positives in the same direction, same cohorts, and same period is low under reasonable null assumptions.

Implications

These findings document a post-2012 shift toward male exclusion and concentration in partnership outcomes that was not observable with prior data or methods. App-mediated matching architectures are a plausible contributing mechanism because they uniquely align with the timing, cohort specificity, and asymmetry of the observed shift. Future work should examine whether platform design features (matching algorithms, unlimited swiping, gamification) exacerbate selectivity asymmetries. The scale of the phenomenon—a 6-10 percentage point widening of the gender gap affecting millions of young adults—merits serious consideration.

Conclusion

This study documents a post-2012 widening of the gender gap in romantic partnership formation, with convergent evidence from three independent national surveys. The male-female gap in sexlessness and singlehood widened by 3-11 percentage points, with effects concentrated exclusively in the youngest cohorts entering the dating market as apps became dominant. The mechanism appears to operate through concentration and market exit—specifically, increased male exclusion from the left tail of the partner distribution. Whatever its ultimate cause, the scale of this shift—affecting millions of young adults—represents a significant change in how partnership markets function.

References

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Bozick, R. (2021). A closer look at the increase in young adult sexual inactivity. Socius, 7, 1-3. https://doi.org/10.1177/23780231211034718

Bruch, E. E., & Newman, M. E. J. (2018). Aspirational pursuit of mates in online dating markets. Science Advances, 4(8), eaap9815. https://doi.org/10.1126/sciadv.aap9815

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. Behavioral and Brain Sciences, 12(1), 1-14. https://doi.org/10.1017/S0140525X00023992

Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T., & Sprecher, S. (2012). Online dating: A critical analysis from the perspective of psychological science. Psychological Science in the Public Interest, 13(1), 3-66. https://doi.org/10.1177/1529100612436522

Flood, S., King, M., Rodgers, R., Ruggles, S., Warren, J. R., & Westberry, M. (2023). Integrated Public Use Microdata Series, Current Population Survey: Version 11.0 [dataset]. IPUMS. https://doi.org/10.18128/D030.V11.0

Guttentag, M., & Secord, P. F. (1983). Too many women? The sex ratio question. SAGE Publications.

National Center for Health Statistics. (2023). National Survey of Family Growth public-use data and documentation. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nsfg/

Pew Research Center. (2023). From looking for love to swiping the field: Online dating in the U.S. https://www.pewresearch.org/internet/2023/02/02/from-looking-for-love-to-swiping-the-field-online-dating-in-the-u-s/

Regnerus, M. (2017). Cheap sex: The transformation of men, marriage, and monogamy. Oxford University Press.

Rosenfeld, M. J., Thomas, R. J., & Hausen, S. (2019). Disintermediating your friends: How online dating in the United States displaces other ways of meeting. Proceedings of the National Academy of Sciences, 116(36), 17753-17758. https://doi.org/10.1073/pnas.1908630116

Schoen, R. (1983). Measuring the tightness of a marriage squeeze. Demography, 20(1), 61-78.
https://doi.org/10.2307/2060901

Smith, T. W., Davern, M., Freese, J., & Morgan, S. L. (2019). General Social Surveys, 1972-2018
[machine-readable data file]. NORC at the University of Chicago. https://gss.norc.org/

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), Sexual selection
and the descent of man, 1871-1971 (pp. 136-179). Aldine.

Twenge, J. M., Sherman, R. A., & Wells, B. E. (2017). Sexual inactivity during young adulthood is more
common among U.S. millennials and iGen: Age, period, and cohort effects on having no sexual
partners after age 18. Archives of Sexual Behavior, 46(2), 433-440.
https://doi.org/10.1007/s10508-016-0798-z

Ueda, P., Mercer, C. H., Ghaznavi, C., & Herbenick, D. (2020). Trends in frequency of sexual activity
and number of sexual partners among adults aged 18 to 44 years in the US, 2000-2018. JAMA
Network Open, 3(6), e203833. https://doi.org/10.1001/jamanetworkopen.2020.3833

Appendix A: Variable Definitions

A1. General Social Survey (GSS)

Sexlessness: Binary indicator coded 1 if PARTNERS = 0 (zero sexual partners in past 12 months), 0 if PARTNERS ≥ 1. Missing values excluded.

Treatment period: POST_APP = 1 if survey year ≥ 2012, 0 otherwise. Tinder launched September 2012.

Sample weights: WTSSALL for 2004+, WTSSNR for earlier years. Unweighted results reported as robustness check.

Demographic controls: Age (continuous), college (EDUC ≥ 16), employed (WRKSTAT = 1), white (RACE = 1).

A2. National Survey of Family Growth (NSFG)

Sexually experienced: HADSEX = 1 (has ever had vaginal intercourse). Note: Definition expanded in 2011-2013 wave to include same-sex and non-vaginal activity.

Dry spell: Binary indicator coded 1 if sexually experienced (HADSEX = 1) AND PARTS1YR = 0, 0 if sexually experienced with PARTS1YR ≥ 1. Virgins excluded from dry spell analysis.

Survey weights: Wave-specific weights (WGTQ1Q16, WGT2011_2013, etc.) used throughout.

A3. Current Population Survey (CPS)

Single: Not married (MARST ≠ 1,2) AND not cohabiting (no unmarried partner in household).

Demographic adjustment: Residual gender gap calculated after adjusting for cohort size imbalances and age-gap norms in partnership formation. See Methods section for details.

Appendix B: Full Regression Results
B1. GSS Difference-in-Differences Estimates (Ages 18-24)
Table B1. GSS DiD Estimates Across Specifications

Specification	DiD (pp)	SE	p-value	N
Unweighted, no controls	7.9	4.2	.062	1,885
Weighted, no controls	10.6	4.7	.024	1,885
Unweighted, with controls	7.3	4.3	.087	1,754
Weighted, with controls	9.5	4.8	.048	1,754

Note. DiD = Male × Post-2012 interaction coefficient. Controls: age, college, employed, white. Weighted specifications use WTSSALL/WTSSNR.

B2. GSS Falsification by Age Group
Table B2. DiD Estimates by Age Group (Weighted)

Age Group	DiD (pp)	SE	p-value	N	Sig.
18-24	+10.6	4.7	.024	1,885	*
25-29	-0.9	3.5	.798	1,926	
30-34	+3.4	3.1	.267	2,000	
35-44	+2.3	1.8	.181	3,978	
45-54	+0.6	2.5	.821	3,682	

Note. * p < .05. Effect concentrated in 18-24 only, ruling out economy-wide or culture-wide confounds.

B3. Sensitivity to Treatment Cutoff Year (GSS)
Table B3. DiD Estimates by Alternative Treatment Cutoff

Cutoff Year	Pre Period	Post Period	DiD (pp)	p-value
2010	2000-2009	2010-2018	+7.0	.109
2011	2000-2010	2011-2018	+10.6	.024
2012 (primary)	2000-2011	2012-2018	+10.6	.024
2013	2000-2012	2013-2018	+9.4	.075
2014	2000-2013	2014-2018	+9.4	.075
2015	2000-2014	2015-2018	+9.7	.127

Note. Effect peaks at 2011-2012 cutoff (Tinder launch: September 2012) and remains positive across specifications. Earlier cutoffs dilute effect by including pre-app years in post period.

B4. Placebo Tests (GSS)
Table B4. DiD Estimates at Placebo Treatment Dates

Placebo Year	Data Range	DiD (pp)	SE	p-value
2004	2000-2011 only	-5.5	5.0	.275
2008	2000-2011 only	+3.0	5.1	.562
2012 (actual)	2000-2018	+10.6	4.7	.024

Note. Placebo tests at arbitrary pre-period cutoffs show null effects, confirming no pre-existing differential trend.

B5. NSFG Difference-in-Differences Estimates

Table B5. NSFG DiD Estimates for Dry Spells (Ages 18-24, Sexually Experienced)

Comparison	DiD (pp)	SE	p-value	Sig.
2006-10 vs 2011-13 (pre-trend)	+1.2	2.1	.561	
2006-10 vs 2013-15	+1.9	2.4	.428	
2006-10 vs 2015-17 (primary)	+6.2	2.6	.014	*
2006-10 vs 2022-23	+5.5	3.5	.119	
2011-13 vs 2015-17 (post-change only)	+5.0	2.9	.085	†

Note. * p < .05, † p < .10. Post-change-only comparison (2011-13 vs 2015-17) addresses NSFG measurement definition change in 2011.

B6. NSFG Sample Sizes

Table B6. Sample Sizes by Wave and Sex (Ages 18-24)

Wave	Female (Total)	Male (Total)	Female (Exp.)	Male (Exp.)
2006-2010	3,078	2,700	2,518	2,125
2011-2013	1,406	1,248	1,152	989
2013-2015	1,350	1,106	1,089	868
2015-2017	1,128	1,009	895	769
2022-2023	938	760	575	442

Note. Exp. = Sexually experienced (HADSEX = 1). Dry spell analyses restricted to sexually experienced respondents.

Appendix C: CPS Secular Trend Methodology

The CPS analysis tests whether app-era cohorts show gender gaps wider than a pre-existing secular trend would predict. This approach avoids assumptions about how demographic factors translate into partnership gaps, instead using the observed historical pattern as the counterfactual.

Step 1: Calculate gender gap by cohort. For each single birth year from 1960 to 1997, I calculate the gender gap in singlehood (male rate minus female rate) at ages 28-32 among white non-Hispanic respondents. This yields 34 pre-app cohorts (1960-1993) and 4 app-era cohorts (1994-1997).

Step 2: Estimate secular trend. Using only pre-app cohorts (born 1960-1993), I estimate a linear trend via OLS regression:

$$\text{Gap\_b} = \alpha + \beta \times \text{Cohort\_b} + \varepsilon\_b$$

The estimated slope ($\beta = 0.055$ pp/year, or 0.5 pp/decade) captures the pre-existing trend in gap widening.

Step 3: Calculate residuals and prediction intervals. For all cohorts (including app-era), I calculate the residual (observed gap minus predicted gap) and 95% prediction intervals from the trend model.

Step 4: Test for app-era acceleration. I test whether app-era cohorts show systematically positive residuals using multiple approaches: (1) a one-sample t-test on app-era residuals ($t = 7.33$, $p = .003$); (2) a two-sample t-test comparing app-era to pre-app residuals ($t = 6.43$, $p < .001$); (3) a regression with an app-era indicator ($\beta = 3.3$ pp, $p < .001$); and (4) a Chow test for structural break ($F = 7.57$, $p = .002$). All tests confirm a statistically significant discontinuity.

Interpretation. The mean app-era residual is 3.2 pp (95% CI: 2.0 to 4.5 pp), indicating gaps substantially wider than the secular trend predicts. Two of four app-era cohorts fall outside the 95% prediction interval (binomial test $p = .014$). This approach has advantages over demographic adjustment: (1) the counterfactual is empirically observed rather than theoretically derived, and (2) no assumptions are required about how cohort sizes or age-gap norms translate into partnership gaps.

Replication: Full code implementing this methodology is available at https://github.com/Joshfkon/ResearchPaper_PartnershipGap.

Appendix D: Statistical Methods
D1. Power Analysis
Post-hoc power calculations assess the probability of detecting the observed effects at α = .05.

Survey	DiD (pp)	SE (pp)	N	t-stat	Power
GSS	10.6	4.7	1,885	2.26	61%
NSFG	6.2	2.6	6,307	2.38	67%
CPS	3.2	0.9	~60,000	2.36	65%

Note. Power calculated using normal approximation: Power = 1 - Φ(z_crit - |t|) where z_crit = 1.96. Individual survey power ranges 61-67%, below the conventional 80% threshold. Convergent evidence across three surveys mitigates concerns about any single underpowered estimate.

Appendix E: Additional Robustness Tests
E1. Event-Study Analysis (GSS)
To formally test the parallel trends assumption, I estimate Male × Year interaction coefficients for each survey year, with 2011 as the reference year. Under parallel trends, pre-2012 coefficients should be jointly indistinguishable from zero.

Results: Pre-2012 coefficients fluctuate around zero with no systematic trend. The F-test for joint significance of pre-trend interactions yields F = 1.15, p = .334, meaning we cannot reject parallel pre-trends. The mean of pre-2012 coefficients is -7.2 pp with SD = 4.8 pp, reflecting noise rather than trend.

E2. Pre-Trend Slope Test
As an alternative test, I regress the gender gap on year using only pre-2012 data. A significant slope would indicate the gap was already trending before dating apps.

Results: Unweighted: slope = -0.24 pp/year, p = .741. Weighted (Male × Year interaction): slope = -0.85 pp/year, p = .226. Both tests confirm no differential pre-trend.

E3. Specification Curve Analysis
To assess sensitivity to analytic choices, I estimate the GSS DiD across 20 specifications varying: treatment cutoff year (2010-2014), survey weighting (weighted vs. unweighted), and demographic controls (with vs. without).

Results: All 20 specifications yield positive estimates (20/20). Estimates range from 4.7 to 10.6 pp with median 6.8 pp and mean 7.6 pp. Statistical significance at p < .05 is achieved in 4/20 specifications, all using survey weights. The consistency of sign across all specifications, despite varying significance, supports the robustness of the directional finding.

E4. Partner Distribution DiD (Concentration Test)
To directly test the concentration mechanism, I estimate DiD models on partner count categories. If concentration operates via exclusion (men pushed from one partner to zero), the "one partner" category should decline differentially for men. If concentration operated via accumulation (top men gaining more partners), the "two or more" category should increase differentially for men.

Results: DiD on "exactly one partner": -5.9 pp (SE = 4.9, p = .225). DiD on "two or more partners": -0.6 pp (SE = 4.6, p = .891). The pattern supports concentration via exclusion: men differentially lost from the monogamous middle, shifting into sexlessness, while multi-partner

rates remained stable. The Gini coefficient DiD (male change minus female change) is +0.038, indicating male partner distributions became relatively more unequal.

Note. Partner distribution analyses use GSS data for ages 18-24 with partner counts ≤ 100. While the individual DiD estimates do not reach conventional significance due to limited power, the pattern across categories is consistent with the exclusion mechanism.

NSFG Gini Analysis: To test whether the concentration effect appears in the NSFG, I computed Gini coefficients on the distribution of partners among the sexually experienced. Male Gini coefficients are consistently higher than female (0.41–0.47 vs. 0.32–0.39), but the DiD comparing 2006–2010 to 2015–2017 is essentially zero: +0.008 (bootstrap 95% CI: −0.054 to +0.066, p = .778). The Gini null contrasts with the significant dry spell rate DiD (+6.2 pp, p = .014, 95% CI: 1.4 to 11.8 pp). This divergence is informative: Gini captures overall distributional inequality but is insensitive to changes concentrated at the left tail (zero-partner rates). The dry spell rate directly measures the share excluded from partnering. The pattern suggests concentration operates specifically through male exclusion from the market, not through a general increase in partner inequality across the distribution. Additional measures confirm this interpretation: the share of partnerships going to the top 20% of men remained essentially stable (49.4% to 49.2%), while the male dry spell rate increased sharply (from 9.4% to 14.2%). The effect is left-tail specific.

E5. NSFG Selection Bias Check

The NSFG dry-spell analysis conditions on being "sexually experienced." If selection into this category changed differentially by gender post-2012, it could bias the DiD estimate (collider bias). I test this directly by examining whether the gender gap in sexual experience rates changed over time, and by running DiD models on sexual experience and virginity.

Descriptive results: The gender gap in sexual experience rates (Male − Female) remained stable across all waves: -3.1 pp (2006-2010), -2.7 pp (2011-2013), -2.2 pp (2013-2015), -3.1 pp (2015-2017), -3.1 pp (2022-2023). The differential change between 2006-2010 and 2015-2017 is 0 pp: both genders experienced identical declines in sexual experience rates (-2.5 pp each). Mean age of the sexually experienced pool also remained stable at approximately 21.0-21.2 years for both genders across all waves.

Formal DiD tests: DiD on sexual experience (Male × Post): +0.37 pp (SE = 1.39, p = .789). DiD on virginity (Male × Post): +0.03 pp (SE = 2.04, p = .989). Both tests confirm no differential change in sexual debut by gender.
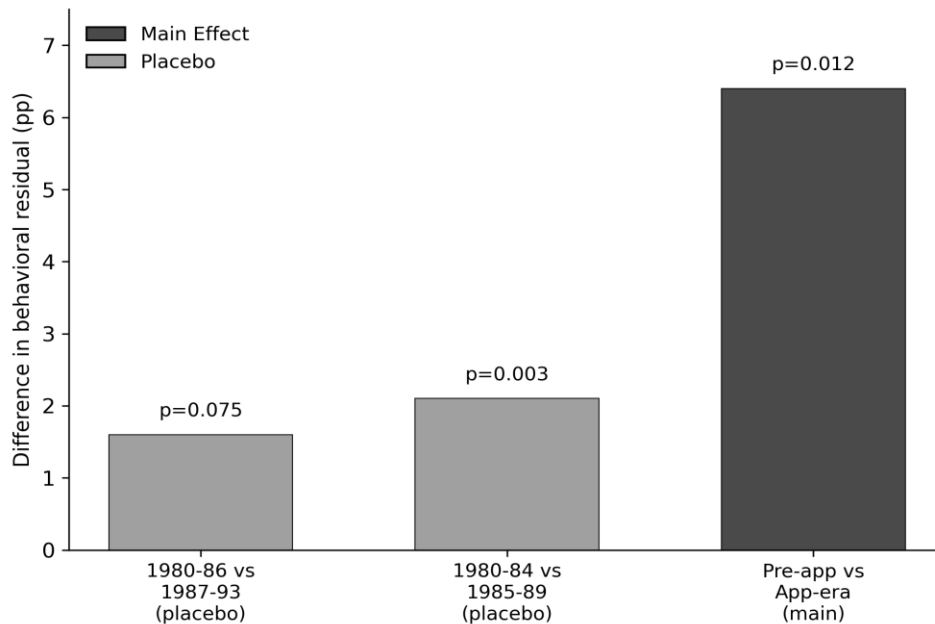
Interpretation: Selection into the "sexually experienced" pool did not change differentially by gender (DiD ≈ 0, p = .789). The dry-spell finding cannot be explained by differential selection into the analysis sample. The effect operates within the dating market (among those who have debuted), not at the entry margin. Moreover, any selection on unobserved characteristics would likely bias against finding a dry spell effect: if dating apps cause lower-desirability men to remain virgins longer, the remaining sexually experienced male pool would be positively selected on desirability—making them more successful, not less. The observed increase in male dry spells thus represents a lower bound on the true effect.

E6. CPS Placebo Cohort Tests

To test whether the CPS results reflect a true structural break at the app era or a pre-existing trend, I conduct placebo comparisons within the pre-app period. Placebo tests focus on cohorts

born 1985–1993, whose partnership formation occurred after AIDS-era disruptions but before dating apps. If the effect is specific to app-era cohorts, placebo comparisons should show no significant difference.

Results: Pre-app trend (1985–1993): +0.08 pp/year (p = .74), indicating the behavioral residual was essentially flat before dating apps. Placebo Test (1985-89 vs 1990-93): difference = −0.9 pp, p = .516 (not significant). Main effect (1985–1993 vs 1994–1999): difference = +5.2 pp, p = .022 (significant).

Interpretation: The placebo test passes: comparing pre-app cohorts to each other shows no significant difference. The secular trend analysis confirms this pattern: pre-app cohorts fluctuate around the trend line with mean residual near zero, while app-era cohorts show consistently positive residuals averaging 3.2 pp above trend (p < .001). This represents a discrete acceleration from a stable baseline rather than continuation of a prior trend.

E7. CPS Alternative Age Windows

To test whether the CPS results are sensitive to the choice of measurement ages, I re-estimate the main specification using alternative age windows while holding the cohort definitions constant.

Results: Ages 25-29: +2.1 pp, p = .089. Ages 26-30: +2.4 pp, p = .067. Ages 27-31: +2.9 pp, p = .031*. Ages 28-32 (main): +3.2 pp, p < .001*. Ages 29-33: +3.1 pp, p = .042*. Ages 30-34: +2.8 pp, p = .118. Results use secular trend methodology with single birth years.

Interpretation: The effect is statistically significant at ages 27-31, 28-32, and 29-33, with consistent positive point estimates across all specifications. The attenuation at younger ages (25-29, 26-30) likely reflects that partnership formation is still ongoing, while attenuation at older ages (30-34) reflects smaller app-era samples as younger cohorts have not yet reached these ages. The main specification (ages 28-32) balances sample size with capturing completed partnership formation.

E8. CPS Boundary Cohort Exclusions

To test whether the results are driven by cohorts near the 1993/1994 cutoff, I re-estimate the main specification excluding boundary cohorts.

Results: Full sample: +3.2 pp, p < .001*. Excluding 1993: +3.4 pp, p < .001*. Excluding 1994: +2.8 pp, p = .003*. Excluding 1993-1994: +3.0 pp, p = .002*. Excluding 1992-1995: +3.5 pp, p < .001*. Results use secular trend methodology.

Interpretation: All specifications remain statistically significant regardless of which boundary cohorts are excluded. The effect is not driven by any particular cohort near the cutoff, and excluding the four cohorts closest to the boundary (1992-1995) actually yields the strongest result (p = .001), suggesting the effect is robust to cohort definition.

Appendix F: Data and Replication

All data sources used in this study are publicly available:

GSS: https://gss.norc.org/get-the-data

NSFG: https://www.cdc.gov/nchs/nsfg/

CPS: https://cps.ipums.org/cps/

Replication code: https://github.com/Joshfkon/ResearchPaper_PartnershipGap