Changes in gendered social position and the depression gap over time in the United States

Jonathan M. Platt

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

Changes in gendered social position and the depression gap over time in the United States

Jonathan M. Platt

Introduction

There is a large literature across disciplines aimed at understanding the causes of the depression gap, defined as an excess of depression among women compared with men. Based on the totality of evidence to date, social stress appears to be an important explanation for the depression gap. Social stress theory highlights women's disadvantaged social position relative to men, positioning gender differences in socio-economic opportunities as social stressors, while also acknowledging how gender socialization teaches women to respond to stressors in depressogenic ways from an early age. This dissertation applied social stress theory to better understand the social causes of the depression gap with three related aims. Aim 1 summarized the evidence for variation or stability in the depression gap in recent decades, through a systematic review and meta-regression of depression gap studies over time and by age. Aim 2 examined the evidence for a changing depression gap across birth cohorts, and tested the extent to which any changes over time were mediated by changing gender differences in education, employment, and housework rates, three indicators of broader trends in gendered social position through the 21st Century. Aim 3 examined whether women in the workforce with competing domestic labor roles were at increased risk of depression, and whether pro-family workplace benefits buffered the effects of competing roles.

Methods

In aim 1, depression gap estimates were extracted through a systematic review of published literature (from 1982-present). Analytic datasets were comprised of 76 diagnostic-based estimates and 68 symptom-based estimates. For each dataset, meta-regression models estimated time and age variation in the depression gap, as well as the interaction between time and age group, to estimate the variation in the

gap over time by age. Data from the National Longitudinal Surveys were utilized for aims 2 and 3. Depression was measured with the Center for Epidemiologic Studies Depression Scale (CESD), and the depression gap was defined as differences in mean CESD scores for women vs. men. The aim 2 sample included 13,666 respondents interviewed from 1992-2014. Hierarchical mixed models estimated the magnitude of the gender depression gap over time, and its relationship with 10-year birth cohort (range: 1957-1994) and whether any variation was mediated by gender differences in: those with a college degree or more, those who were employed full-time, and the average number of hours spent doing housework per week. The sample in aim 3 was limited to employed women ages 17-57 (n=3993). Generalized estimating equations estimated the relationship between competing roles and depression, and the interaction between competing roles and pro-family employee benefits on depression. Interaction results were compared to models estimating the effect of non-family-related benefits on the relationship between competing roles and depression.

Results

In aim 1, there was no evidence of change in the depression gap over time. Compared with the reference group (i.e., respondents ages 60+), the age effect was appreciable among the youngest age group (age 10-19) (RR=1.44; 95% CI=1.19, 1.74), but did not differ for any other age groups. The age by time interaction was elevated for youngest age group (RR=1.27; 95% CI=1.0, 1.61), suggesting that, compared to the oldest age group, the diagnostic depression gap had increased among the youngest ages from 1982 to 2017. There was no evidence of time changes among any other age group. Results were similar for symptom-based studies.

In aim 2, there was a linear decrease in the depression gap by 0.18 points across birth cohort (95% CI= -0.26, -0.10). The results of the mediation analysis estimated that an increasing ratio of college degree attainment mediated 39% of the gender depression gap across cohorts (95% CI= 0.18, 0.78). There was no evidence of mediation due to changing employment or housework ratios.

In aim 3, there was evidence that women in competing roles reported a 0.56-point higher CESD score (95% CI= 0.15, 0.97), compared with women not in competing roles. The interaction between pro-family benefits and competing roles was associated with CESD scores (B=-0.44, p=0.023). More specifically, among women *without* access to pro-family benefits, those in competing roles reported a 6.1 point higher CESD score (95% CI=1.14, 11.1), compared with those not in competing roles, however, among women *with* access to these benefits, there was no association between competing roles and CESD scores (difference=0.44; 95% CI=-0.2, 1.0). Results were similar for non-family-related benefits. Women in competing roles *without* non-family-related benefits reported a 3.59 point higher CESD score than those not in competing roles (95% CI=1.24, 5.95) while among women *with* access to these benefits, there was no association between competing roles and CESD symptoms.

Conclusion

This dissertation provided evidence to partially support the hypothesis that the depression gap is changing over time and is meaningfully related to the social environment, through which gender roles, responsibilities, and opportunities available to women and men are defined and reinforced. The results of these studies suggest that the depression gap may be expanding and contracting over time for different age groups. Understanding the social causes of the depression gap is important to reduce the present and future burden of the depression gap, and to understand the fundamental processes through which depression disparities may be perpetuate or attenuated in adolescence and beyond.

Table of contents

I. List of tables and figures

Supplementary Tables and Figures

Section V.

Section VI.

II. Acknowledgements

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank Dr. Kerry Keyes, whose expertise, guidance, and friendship has been an invaluable part of my dissertation and training since day one at Columbia. I am also extremely grateful to Dr. Sharon Schwartz, who instilled a strong moral framework to understand the implications of scientific inquiry, both within the academy and in the broader world, and Dr. Lisa Bates, who taught me that theory was really, really important…really.

In addition, I am grateful for the support of the NIH Psychiatric Epidemiology T32 training grant, which provided the financial support to allow me to be immersed as a full-time student while completing my degree. This training was supplemented by a weekly seminar, in which Dr. Bruce Dohrenwend, Dr. Sharon Schwartz, Dr. Ezra Susser, and Dr. Kerry Keyes and many others have created a space where all voices are heard, and trainees are welcome to discuss and debate the work of inspiring scholars, week after week. It has left an indelible mark on my academic engagement and scholarship.

I would like to thank all of the staff in the Columbia University Department of Epidemiology who keep the department running on a daily basis, and help make student's lives more organized, especially Brenda Scariff, Liliane Zaretsky, Elizabeth Ferrari, and Anthony Diaz. I must also express my gratitude to my classmates, especially Danielle Crookes, Patrick Dawson, Eleanor Hayes-Larson, John Pamplin, and Luis Segura, for their friendship and collegial academic support. They have been an instrumental part of my learning and happiness as a student.

Finally, I would like to thank my family for the love and support throughout my training at Columbia. To my best friend and frenemy, Claire, thank you for your support, your feedback, and for taking on more than your fair share of the domestic labor while I finished my dissertation. I would also like to thank my mother, father, and sister for their wisdom, support, and sympathetic ear. Growing up with exposure to the moral clarity with which you live and work has instilled me with the courage to seek a career in service to others, that is both professionally and personally satisfying. Finally, I must acknowledge my grandmother, Betty Manion, who lived much of her life during the period of historical and social change underlying this dissertation. She almost singlehandedly raised six children while working full-time, running her own business for over 30 years. She has been a source of inspiration both in her life and her memory.

III. Dedication

To Claire, who sees past the glass ceiling.

IV. Introduction

The term depression refers generally to a period of low mood or sadness that can affect an individual's thoughts, behavior, feelings, and sense of well-being. In clinical settings, depression is often diagnosed as major depressive disorder,[1] though in epidemiologic research, depression is measured in a variety of ways, which may reflect distinct but related constructs such as demoralization or distress. Despite these measurement differences, studies have consistently found that depression incidence, prevalence, and symptom levels are higher among women than men (hereafter referred to as the *depression gap*). The depression gap was first reported in the 1970s,[2] and has since yielded a large body of research to test a diverse set of mechanisms as causes of gender patterns.

Part of this research has included attempts to explain the depression gap as spurious and an artifact of measurement error. For example, some have suggested that women are more likely than men to recall depression symptoms.[3,4] However, gender differences persist in studies that have assessed depression both prospectively and retrospectively.[5] Others have suggested that women are more likely to seek help for depression symptoms,[6] however, the depression gap has been reported in both clinical and community samples.[7] The stigma surrounding depression may also be gendered, where women feel less stigma in reporting depressive symptoms and be more likely to endorse survey items measuring depression.[8] However, the depression gap has been reported based on both self- and informant-report,[9] as well as studies that account for social desirability bias.[10] At the symptom level, if social desirability bias were a significant driver of gender differences, one would expect to see greater gender differences in those symptoms that are more stigmatized (e.g., 'I feel sad most of the time') than others (e.g., 'I have less energy than usual'). However, there is no consistent symptom-level pattern to support this hypothesis.[3] While there is some evidence that gender-dimorphic items in rating scales may influence estimates,[11] differences in rating scales and tools do not account for the depression gap.[12] Overall, gender differences in major depression appear to be genuine and not due to bias.

To date, most research into the causes of the depression gap have sought to test whether known causes of inter-individual variation in depression are more prevalent among women than men. For example, if childhood adversity increases depression risk generally, and women experience more childhood adversity than men, then childhood adversity may also explain the depression gap.

The emergence of depression gap is approximately correlated with puberty,[13,14] which has led many to study hormonal mechanisms as a cause. For these hormones to explain the depression gap, studies should identify consistent patterns where changes in hormone levels are associated with changes in depression risk differentially among women and men. Overall, while hormones may partially explain inter-individual depression risk, they do not appear to explain the between-group risk that would underlie the depression gap.[15,16]

Across the many studies of depression, one of the most consistent risk factors is a family history of the disorder,[13,17] which has led researchers to search for genetic causes of depression.[18,19] Research examining genetic causes of the depression gap has hypothesized that a genetic risk factor for depression may be X-chromosome-linked, leading to a differential risk of depression by gender, though studies that identify gender differences are limited and inconsistent with this hypothesis.[20–22] Overall, evidence across multiple genetic study designs has offered no support for genetic mechanisms as causes of the depression gap.[23–25]

Based on the totality of evidence to date, social stress appears to be the most likely explanation for the depression gap. Social stress theory suggests that gender differences in social position may influence stress exposures and responses.[26,27] Women traditionally have had fewer opportunities for educational attainment than men, lower professional prestige and income, and more domestic responsibilities (e.g. housework, childcare), all of which may act as social stressors.[28] Women are also socialized, through gender norms, to respond to stressors in depressogenic ways.[29–34] These may increase women's depression risk, and thus explain gender differences in depression.[28,35,36] However, if these social differences underlie depression, changes over time in women's social position in the United States suggest that the depression gap may change as well, but three areas of evidence remain underexplored.

First, the variation in the depression gap over time is unknown. The gap is typically cited as approximately twice as high among women than men,[37] however, recent evidence suggests that it may be changing.[38–40] A systematic review of published research on the depression gap will document any changes over time, while accounting for study characteristics that are not related to temporal variation.

Second, while social stress theory has previously been used to understand the depression gap, there is limited research leveraging change over time to test how gendered social position might affect the depression gap. Since the mid-20th century, the relative social position of women vs. men has improved in recent birth cohorts. Traditional gender norms[34,41,42] have become less restrictive, and as a result, women have become more likely to get a college degree,[43,44] and advance to high-skill occupations,[45] all while delaying having children.[46] If social position is protective against depression, then changes in women's social position should predict a decline in depression prevalence over time. Concurrently, the effects of these social changes on men's depression remain underexplored. Therefore, the second aim of this dissertation will use a prospective study of employment, education, domestic experiences, and health of respondents from successive birth cohorts to assess changes over time among women and men, and explore whether changes in women's social position are associated with changes in the depression gap over time.

Third, women's rising social position may lead to additional unintended sources of stress. Specifically, women in dual workplace and domestic labor roles may be at greater risk of experiencing multiple role strain, which could increase their depression risk.[47] That is, working women who have children are well documented to suffer professionally and economically, relative to working women without children. In light of this, companies have increasingly offered pro-family benefits to their employees, such as paid family-leave, childcare subsidies, and flexible working schedules, intended to retain employees after the birth of children.[48] If the depression gap is partially explained by multiple role strain, then these benefits should buffer the depression risk among women in the work force. To test this hypothesis, I will examine whether the availability of pro-family employee benefits reduces depression risk among women in multiple roles.

In summary, the depression gap is well-documented, but changes in the hypothesized causes suggest that it may be changing. This dissertation will examine those changes, in three aims:

Aim 1 – Systematically review the literature to summarize the variation in the depression gap in the US population over time and by age, accounting for differences across studies based on the use of diagnostic vs. symptom scales.

Aim 2 – Examine whether temporal variation in the depression gap is associated with changes in gendered social position, as indicated by changes in education and employment rates among women versus men, as well as decreasing housework divisions between women and men. Change over time will be tested by estimating gender differences in depression prevalence in a longitudinal US population-representative sample of men and women grouped in sequential birth cohorts.

Aim 3 – Test if the burden of competing workplace and domestic roles among women increases their risk of depression, compared to women without competing roles, and the extent to which that risk is buffered by access to pro-family benefits in the workplace.

V. Is the US gender gap in depression changing over time? A meta-regression

INTRODUCTION

Depression, the gender depression gap, and gendered social position

Major depressive disorder is a persistent[i] state of low mood and apathy that can affect the thoughts, behavior, feelings, and sense of well-being of the afflicted.[1] It is the leading cause of disability among Americans ages 15-44.[49] Compared with men, women report more depressive symptoms, and a higher incidence and prevalence of depression, throughout the world.[50] In the United States, studies generally find that the prevalence of depression among women is twice that of men.[37] This pattern is not solely an artifact of gender differences in reporting mental health symptoms or seeking treatment, rather evidence to date indicates that the differences reflect meaningful differences in depression.[9,10,51]

In studies of the depression gap, depression is often defined as meeting a diagnostic threshold (e.g. based on the DSM-IV);[14] however, there is considerable heterogeneity in the definition and operationalization of depression. In addition to diagnostic criteria, studies of the depression gap also define depression as psychological distress[52] and more general measures of somatic or internalizing behaviors.[53] While these measures represent related constructs, differences may potentially influence the magnitude of the gap; yet, existing studies generally converge on higher mean symptoms among women compared with men.[54]

The depression gap emerges in early adolescence, remains relatively stable throughout adulthood, then decreases at later ages.[13,14] Biological[55–57] and social[58] mechanisms have been explored to explain the gap, with the most robust evidence to date supporting social stress. As applied to gender, social stress theory suggests that gender differences in social position may influence stress exposures and responses.[26,27] Women traditionally have had fewer opportunities for educational attainment, lower professional prestige

[i] Symptoms must be present for at least two weeks to meet current criteria for major depressive disorder

and income, and more domestic responsibilities (e.g. housework, childcare) than men, all of which may act as social stressors.[28] From an early age, women are also socialized, through gender norms, to respond to stressors in depressogenic ways.[29–34] These factors may increase women's depression risk, and explain gender differences in depression.[28,35,36] If so, changes in women's social position, and therefore changes in these factors, should change the depression gap in turn.

Since the mid-20th century, education and employment opportunities have become increasingly available to women. Women surpassed men in earning a college degree by the early 1990's.[43,44] Employment rates among working-age women doubled to 70%,[59] and today, women are nearly as likely to be employed as men.[60] The time women spend on housework has also decreased.[61,62] Over the same period, women have waited longer to marry,[63] increasingly used birth control,[64,65] had fewer children,[66] and had children at later ages.[46]

The effect of these particular changing opportunities would likely be clearest for women at the ages when they are directly engaged in formal and domestic labor roles. However, to the extent that changes in gendered social position reflect broader changes in norms and the process of gender socialization, these changes may decrease the gap for girls and women of all ages.

Some studies have suggested that the depression gap may be changing,[38–40] but overall evidence is inconclusive, in part due to three limitations. First, follow-up periods in single longitudinal studies are often too short to identify temporal trends in depression. Second, while existing studies suggest that the depression gap may vary over time,[54] there may also be variation by age across time. Examining variation by both age and time is necessary to identify any temporal variation due to social change. Among reviews that have directly accounted for age in assessing temporal variation in the depression gap, most have focused on a single age group or developmental period.[67,68] A wide time span with age groups across the lifecourse is necessary in order to fully characterize variation in the depression gap by both age and time, and potentially illuminate the role of gendered social position, which would be particularly relevant only at some ages. Third, less attention has been paid to gender differences in levels of depressive symptoms, compared with the gender gap in diagnostic depression. Examining gender differences in symptom-based

assessments of depression may shed important light on gender patterns in levels of depressive symptoms.[69] Additionally, as noted above, symptom-based tools may also measure a more generalized construct (e.g., demoralization),[70] which may be related to diagnostic depression, yet examination of potential variation in the magnitude of the gap across time and age by measures may reveal meaningful trends in particular constructs related to depression.

Given these limitations of individual studies, a systematic review and meta-regression were conducted to characterize changes in the depression gap over time. First, studies of gender differences in the depression gap in recent decades were identified and summarized. Second, data from the systematic review were extracted to form the analytic sample of the meta-regression, which estimated the variation in the gap over time by age, and accounting for other potential sources of variation between studies. Trends in the depression gap were considered separately based on diagnostic vs. symptom-based depression tools, to explore whether variation has been different at a diagnostic threshold vs. across a range of depression symptoms.

METHODS

Individual studies of temporal trends in the depression gap are limited. Therefore, in order to characterize variation in the depression gap, the review was structured to estimate cross-study variation over time with meta-regression models. For each depression gap estimate, the baseline study year formed the main independent variable in the meta-regression model. That is, each study year represented a cross-sectional estimate of the depression gap. The study year regression coefficient corresponded to change in the depression gap over time, accounting for differences in age and other potential sources of variation. The following sections describe the study identification, data abstraction, and analytical procedures in more detail.

Identification of studies

The initial literature search focused on peer-reviewed research published in English language journals between January 1980 and May 2019. The year 1980 was chosen as the lower limit because it coincided with changes to women's social position in the US that had been ongoing since the mid 20th Century. The year 1980 also represented the introduction of version three of the Diagnostic and Statistical Manual (DSM-III) that was integrated into instruments used in community-based psychiatric epidemiological surveys to estimate the US population prevalence of psychiatric disorders including depression.[71,72] The DSM measures were informed by related constructs such as demoralization and distress.[73–75] Only studies of the US population were included, given the background of changing social position for women in the US potentially influencing the gender gap in depression. Finally, the search focused on studies based on nationally-representative sampling frames for several reasons. First, the effects of changes in gendered social position are widespread, and therefore, should be characterized at the US-population level; second, estimates from non-population-representative community samples may reflect gender-specific selection factors (e.g., clinical samples[6]); and third, US-population-level samples typically have large sample sizes that maximize statistical precision of depression gap estimates.

The literature search and study selection flow chart is detailed in Figure 1.1. The initial search included five electronic databases: PubMed, JSTOR, Embase, PsychInfo, and Scopus. The following general search strategy was used to identify all potential articles and datasets: ('gender' OR 'sex') AND (('male' AND 'female') AND ('depress*[ii]' OR 'distress' OR 'demoraliz*' OR 'internaliz*'). Search terms were optimized using MeSH terms and adapted for each database. The initial search yielded 1007 potential abstracts. Bibliographies of related reviews and meta-analyses were also searched, which yielded 20 additional estimates, primarily from one review that included previously unpublished data obtained through personal contact with authors.[54]

[ii] * denotes a stem that may encompass various forms a word, e.g., depress* = depressive, depression, depressed, etc.

Study titles and abstracts were imported into a reference management system,[76] 218 duplicate studies were removed, and 809 abstracts were screened in more detail and additional studies were removed based on the following exclusion criteria:

- The sample population was not based on a nationally representative sampling strategy (e.g., school-based and other convenience samples, clinical populations)
- The sample was from a non-US population
- The paper presented no quantitative data (e.g., qualitative study, narrative review)
- The sample included non-human subjects
- Gender-specific data (e.g., prevalences, risk ratios) were not presented
- Depression measures were not based on a symptom-level interview (e.g., self-reported doctor diagnosed depression)

A second reviewer independently screened the 809 abstracts. Agreement between the two reviewers was very good (kappa= 0.827, 95% CI=0.788, 0.867),[77] and the reviewers further discussed any conflicting judgments to reach consensus. As a result, 452 studies met exclusion criteria and were removed.

The full text of the 357 remaining studies was reviewed, and studies were screened again in more detail. One additional restriction was made in the full-text review to ensure the independence between depression gap estimates. In the instance that the same dataset was used for multiple studies, only the study with the most complete sample was included (i.e., the fewest stated restrictions to derive the analytic sample from the full study sample).

The reasons for exclusion of the full-text reviewed studies were: study design (e.g., case control, sampling based on depression status) (k=158, 50%), a non-nationally representative sampling strategy (k=123, 39%) and duplicate data source (k=35, 11%) (see Figure 1.1).

Based on all inclusion and exclusion criteria, 41 studies were included. Several of these studies included multiple estimates for different age groups, and each group was considered as an independent

estimate (range:1-17 estimates per study). Also, in the instance of longitudinal studies with multiple waves of follow-up, only baseline data were included to avoid issues of within-sample correlation of depression gap estimates and potential selection bias from attrition. Several studies did not measure depression at baseline, but included it in later interviews. Estimates from the first follow-up interview where depression was measured were included, and attrition data (proportion lost-to-follow-up) were extracted to consider the potential for selection bias. The full meta-analytic dataset contained 144 independent estimates from nationally representative samples. The total sample size was 813,189 (52% women). The study selection process is summarized in Figure 1.1.

Data Abstraction

For each estimate, the following information was collected: author names, year published, year at study baseline, sample size by gender, age range of respondents, the depression effect measure, effect estimate, variance of the effect estimate, the instrument used to measure depression (e.g., DSM-IV, CESD), and the period of recall of depression symptoms (e.g., past 30 days, past-year, lifetime).

Effect measures

The depression gap was summarized as a prevalence ratio (PR) among studies that reported depression based on a diagnostic threshold (i.e., diagnostic studies), and as a standardized mean difference (SMD) among studies that reported mean depression symptoms (i.e., symptom-based studies). All analyses were done separately for diagnostic and symptom-based studies.

Prevalence ratios were calculated as the risk of depression among female vs. male respondents, therefore a PR greater than one indicated that women reported excess risk of depression. Among studies where gender stratified depression prevalences were presented, the PR was calculated from a contingency

table. The PRs were then log-transformed to estimate standard errors. Each study was weighted by the inverse of the standard error of the log prevalence ratio.[78]

Standardized mean differences were calculated as the mean depression score among women minus the mean score among men, divided by the pooled standard deviation (a weighted sum of within-group standard deviations).[79] A positive estimate indicated an excess of depression symptoms among women vs. men, and null, small, medium, and large effect sizes were conventionally defined as SMD=0-0.19, SMD=0.2-0.49, SMD=0.5-0.79, and SMD=0.8-1.0 respectively.[79] Each study was weighted by the inverse of the standard error of the SMD.[78]

Independent variables

Time was considered as the main independent variable, to estimate variation in the depression gap across studies. Time was defined based on study year at baseline, and was grouped as: before 1990; 1990-1994; 1995-1999; 2000-2004; 2005-2009; 2010 or after. The time coefficients were approximately linear when modeled as indicator variables, so for simplicity in interpretation time was modeled as a continuous six-level variable in all models. Time was also considered as an ungrouped continuous variable (range 1982-2017) to examine the sensitivity of the analysis to the groupings.

Age was considered as an effect modifier of time. The age group variable was created among samples where the depression gap was estimated for discrete age ranges. Four age groups were defined with indicator variables: ages 10-19 representing childhood/adolescence; 20-39 representing early adulthood; 40-59 representing middle adulthood; and 60 or older representing older ages. Groupings were chosen in order to capture meaningful life periods, while also ensuring large enough samples within each group. Studies with wider age ranges (e.g., ages 18-65) were included in the descriptive analysis but not the meta-regression models (3 diagnostic studies (3.9%) and 6 symptom studies (8.8%)).

Confounding variables

The depression instrument was also considered as a confounding variable of the association between time and the depression gap. The instrument used to assess symptoms or diagnostic thresholds was categorized separately for symptom scales (CDI, PHQ-9, other vs. CESD) and diagnostic tools (DSM-III or III-R vs. DSM-IV or IV-R, other). Only one study presented estimates using the DSM-5 diagnostic algorithm, so it was grouped with DSM-IV-based studies.

Publication bias

Publication bias arises when studies with null findings are systematically less likely to be published, and can bias meta-regression parameters toward the null.[80] The likelihood of publication bias in the present analysis was low, however, as many studies presented the depression gap as descriptive data. The depression gap was often not the main focus of the article, suggesting that the magnitude or significance of the depression gap would have little influence over whether a study was published. Nonetheless, to explore potential publication bias, a funnel plot was estimated for each set of studies. Minimal publication bias was indicated by a symmetrical distribution of studies around the pooled effect size. The degree of this symmetry was tested using Egger's test, which tested whether the magnitude of effect estimates differed by the study precision.[81] An intercept that crossed 0 indicated no statistical evidence of publication bias. In addition, the trim-and-fill procedure was used to estimate what the actual effect size would have been in the absence of any publication bias by imputing additional studies until funnel plot symmetry was established.[82]

Analysis

First, a descriptive analysis summarized the data sources, study designs, sampling procedures, and depression measures of all included studies. Additionally, a pooled depression gap was estimated to

summarize the depression gap across all studies in the analytic sample. Prediction intervals were also estimated to provide a range of estimates that would be expected in future studies, based on the observed data.[83] While generating one summary estimate of the depression gap was not a primary goal of this study, it was nonetheless estimated in a supplementary analysis in order to quantitatively summarize the depression gap literature and compare the pooled magnitude of the depression gap with estimates that are typically cited in individual studies in the US population (i.e., risk ratio=2.0).

The primary analysis was to estimate variation in the depression gap over time, implemented with two meta-regression models. The first model regressed the depression gap on time and age, to estimate the conditionally independent temporal and age variation in the depression gap. The second model regressed the depression gap on the interaction between time and age group, to estimate the variation in the gap over time by age. Models were also adjusted for depression instrument. All meta-regression models used maximum likelihood estimation with robust standard errors. All analyses were implemented with 'meta'[84] and 'metafor'[85] packages in R, version 3.5.1.

Multiple Imputation

For 27 estimates of the diagnostic depression gap (40%), data needed to compute the standard error of the effect estimate were not reported (i.e., only an unadjusted PR was reported). Compared with non-missing studies, studies with missing data were published earlier (mean year at baseline (SD) = 2000 (3.8) vs. 2007 (9.2)) and had smaller samples (n total (SD) = 3953 (2343) vs. 9546 (7094)), but did not differ by age, instrument, or reported depression gap. To minimize the amount of information lost due to missing data, the meta-regression model was estimated with imputed variance parameters from 20 imputed datasets using chained equations, combined with corrected standard errors.[86] A pooled depression gap was also estimated with imputed study data. Imputed model estimates were compared to complete case models to examine the degree of their robustness to missing data.

RESULTS

Descriptive summary

Tables 1.1 and 1.2 provide the descriptive details of the diagnostic and symptom-based studies that comprised the analytic sample. Nearly all estimates utilized data from a secondary analysis of large studies capturing a broad array of health outcomes and risk factors of Americans. The data sources with the most estimates of the diagnostic depression gap were: the National Survey of Drug Use and Health (NSDUH),[87–96] and the National Epidemiologic Study of Alcoholism and Related disorders (NESARC).[97] The most reported symptom-based estimates came from the National Health and Nutrition Examination Survey (NHANES),[98–103] and the National Longitudinal Survey of Youth (NLSY), 1979[104] and 1997[105] cohorts. Two additional (symptom-based) estimates were based on primary data collection and analysis.[106,107]

Sampling strategies were similar across diagnostic and symptom-based studies. Studies of respondents ages 18 and older recruited participants using a multi-stage probability sampling strategy, with sampling at the household- and individual-level. Three studies of adolescents used a similar multi-stage sampling strategy, but sampled at the school- and individual-level.[109–111] Four studies recruited participants through random-digit dialing,[106,112–115] and two recent studies of adolescents were recruited through online methods.[116,117] All sample estimates were weighted to reflect the US-Census based demographic distributions (by age, race/ethnicity, and gender) at the time of enrollment.

All studies also utilized sampling weights to adjust for non-response in the recruitment process. In three estimates from longitudinal studies, the depression gap was not measured at baseline. Rather, the gap was derived from a follow-up interview.[104,105,118] Of them, the NLSY studies reported high retention rates of 88%[105] and 91%.[104] The third study was from a National Sample of Older Adults, which utilized a steady-state sampling design, supplementing the aging cohort with a new sample of participants ages 51-56 every six years. These estimates did not differ from other estimates using baseline and cross-sectional data.

Table 1.3 summarizes the distributions of all analytic variables. Of the 144 total estimates, 76 measured the depression gap with a diagnostic instrument and 68 measured the gap with symptom scores. Overall, the study year at baseline ranged from 1982 to 2017. The respondent ages ranged from 10 to 99 years old. Estimates from samples of ages 10-19 were most frequently reported, representing 35.5% of diagnostic and 48.6% of symptom-based estimates. Depression was assessed using DSM-IV/IV-R criteria in 71 diagnostic studies (93.4%), and the CESD scale was used to measure depression in 42 symptom-based studies (61.7%). Among diagnostic studies, 97.4% of studies assessed past-year depression (two studies assessed lifetime depression[97,102]), so symptom period was not included as an independent variable; a sensitivity analysis included only studies of past-year depression to determine whether the meta-regression estimates were biased by the few studies with a longer recall period.

The effect sizes of all diagnostic depression gap estimates and a pooled summary depression gap is presented in supplementary figure 1.1. Among these studies, 95% reported a significant or nearly significant depression gap. Of the three diagnostic studies that reported no gap, confidence intervals of all estimates had lower-limits of 0.92 or more.[95,96] Prevalence ratios ranged from 1.26 (95% CI= 0.99, 1.59) to 4.23 (95% CI=3.37, 5.31), and the pooled summary PR was 2.01 (95% CI=1.88, 2.14). The prediction interval ranged from 1.17 to 3.44. The effect sizes of all symptom-based depression gap studies and a pooled summary depression gap is presented in supplementary figure 1.2. Among these studies, 82.4% of studies reported a significant depression gap. SMDs ranged from -0.12 (95% CI= -0.4, 0.16) to 0.59 (95% CI=0.51, 0.67), the pooled summary SMD was 0.22 (95% CI= 0.19, 0.25), indicating a medium effect. The prediction interval ranged from -0.02 to 0.46 (supplementary figure 1.2).

Meta-regression

Meta-regression models estimated the average effects of time, age, the interaction between time and age, and instrument in depression gap estimates. Model results are presented in table 1.4. Main effects among diagnostic studies were estimated in model 1a. The depression gap with all model variables at their

reference levels was 2.35 (95% CI=1.51, 3.68). Overall, there was no evidence of change in the depression gap over time. The age effect was appreciable among those age 10-19 (PR=1.44; 95% CI=1.19, 1.74), compared with the reference group (i.e., respondents ages 60+). Based on the exponentiated combined intercept and age 10-19 coefficients, the depression gap was 3.38 among this age group. The depression gap did not differ for any other age groups vs. the referent. Model 2a tested the interaction between age group and study year at baseline. The interaction term for youngest age group was elevated (PR=1.27; 95% CI=1.0, 1.61), suggesting that, compared to the oldest age group, the diagnostic depression gap had increased among the youngest ages over the study period. There was no evidence of time changes among any other age group vs. the referent. To determine the robustness of the interaction models to the categorization of the time variable, study year was modeled as an ungrouped continuous variable (range: 1990-2017). The model results were generally consistent although the estimates were attenuated (interaction between time and age 10-19 PR=1.05; 95% CI=1.01, 1.09). The depression gap did not differ by diagnostic instrument. Model estimates were not appreciably different when limited to studies that assessed past-year depression status (k=74; 97.4% of included studies). Compared with age 60+ samples, the depression gap was greater among the age 10-19 group (PR=1.42; 95% CI=1.19, 1.69) (model 1a) and increased over time (PR=1.3; 95%=1.04, 1.61) (model 2a). The depression gap did not differ for any other age groups vs. the referent.

Main effects among symptom-based studies were estimated in model 1b (Table 1.4). In these studies, the depression gap with all variables at their reference levels was 0.27 (0.17, 0.38). There was no evidence for change over time overall. Compared to age 60+ samples, the depression gap was greater only among the youngest ages (age 10-19) (SMD=0.29, based on combined intercept and age 10-19 model coefficients). In model 2b, the interaction term for youngest age group was elevated (SMD=0.01; 95% CI=0.01, 0.09), suggesting that, compared to the oldest ages, the symptom-based depression gap increased over the study periods among the youngest ages. Compared to studies that measured depression with the CESD, the depression gap was significantly higher in the 7 studies that used the PHQ (SMD=0.15; 95% CI=0.04, 0.22) and other instruments (SMD=0.13; 95% CI=0.04, 0.22). When time was modeled as an ungrouped

continuous variable (range: 1982-2017), model estimates were attenuated but results were similar (interaction between time and age 10-19 PR=0.03; 95% CI=0.01, 0.06).

Multiple imputation

In a sensitivity analysis, missing variance information was multiply imputed for 27 diagnostic studies. The depression gap with all variables at the reference level was slightly larger than in the unimputed model (PR=2.47; 95% CI=1.25, 4.87), and the age by time interaction tests were similar to the unimputed estimates (ages 10-19 PR=1.16; 95%=1.01, 1.33, no other age differences vs. the referent). The imputed random effects model pooled PR was not appreciably different from the complete case analysis (PR=1.97; 95% CI=1.82, 2.14). Overall, these results suggested that the complete case analysis was not appreciably biased by missing data.

Publication bias

Funnel plots of the effect size of each study against its precision (the inverse of the standard error), to visually depict the potential for publication bias, are shown in figures 1.2 (diagnostic studies) and 1.3 (symptom-based studies). In the symptom-based model, Egger's test indicated no evidence of publication bias (intercept=-1.19 (95% CI=-3.5, 1.1), though the trim-and-fill procedure imputed 23 additional studies to achieve symmetry in the funnel plot. Imputing these studies yielded a wider prediction interval=-0.026, 0.57, and increased the pooled effect size from 0.22 to SMD=0.27 (95% CI=0.24; 0.30). In the diagnostic-based model, Egger's test indicated no evidence of publication bias (intercept=-0.266 (95% CI=-1.78, 1.24), The trim-and-fill procedure imputed no additional studies to achieve symmetry in the funnel plot.

DISCUSSION

The purpose of this systematic review and meta-regression was to review studies of the depression gap and characterize changes in the gap over time. To my knowledge, this is the largest study to examine changes in the depression gap over time by age in the United States. There were four central findings. First, women's depression risk was twice that of men overall, and the effect size was moderate among symptom-based studies. Second, there was no variation over time among adults ages 20 and older, which does not support the hypothesis that changing gendered social position is narrowing the depression gap. Third, the depression gap increased over time among respondents ages 10-19. Fourth, variation in the magnitude of the symptom-based depression gap was related to differences in depression instrument.

Concordant with nearly all of the depression gap literature, the present meta-analysis identified an appreciable depression gap between men and women. Findings were generally consistent between diagnostic depression and symptom-based depression measures. The overwhelming evidence highlights the consistency and robustness of depression disparities across studies. More variation in the depression gap was found in studies of depression symptom scales, emphasized by a wide prediction interval that ranged from zero to medium effect sizes. This variation was likely due in part to differences in the depression instrument across these studies. Symptom scales, like the CESD, which was the most commonly used instrument in these studies, correlate with diagnostic depression, but likely measure more general psychological distress and demoralization constructs.[119–121] Conflating all of these constructs as a single depression measure may introduce measurement error into depression gap estimates. Indeed, this study found evidence that the magnitude of the depression gap varied across measures in the symptom studies. In spite of this, the pooled SMD was precise and robust to instrument differences, however, this variation should be considered when measuring and interpreting the depression gap using symptom scales in future individual studies.

This study examined variation over time across studies using meta-regression models. While there was no evidence for change in the depression gap over time on average, there was heterogeneity in the time effects by age group. Among adults ages 20 and older, there was no variation over time in the depression gap. Evidence of changes in the adult depression gap to date has been mixed. Some have reported a

narrowing gender depression gap among younger adults over time.[39,122] In the Epidemiologic Catchment Area study, researchers found that gender differences in DSM-III depression risk had decreased among individuals born from 1905-1965.[71,72,123] A similar narrowing trend was reported from 2005-2014 in a study of young adults ages 18-25.[38] In contrast, other studies have reported no effects or an increasing depression gap over time. For example, the National Comorbidity Study Replication reported no changes in the gender gap in DSM-III-R depression among individuals born from 1936-1975.[40] Using longitudinal data from the Americans' Changing Lives study, Yang and Lee found evidence that the depression gap increased among individuals born from 1915 to 1955.[124]

The time period covered by the present study coincides with broad changes to women's social position in the US. It was hypothesized that these changes would narrow the depression gap, but the results do not support a clear effect on the depression gap among adults. While it could be that the depression gap is not influenced by social position, the lack of an effect could also reflect both positive and negative consequences of changing position on the depression gap. On one hand, changing social position is indicated by greater opportunities in the workplace and access to personal socioeconomic[45,60,125–128] and psychosocial resources among women.[129–132] Greater resources may reduce exposure to stress[32] and mitigate the effects of stressors[29–31] in ways that influence the risk of depression.[28,35,36,133] On the other hand, these changes may increase exposure to conflict- and overload- related stressors that could increase women's depression risk.[134–137]

Among the youngest respondents, however, the depression gap was appreciably larger than among respondents age 60+. This pattern has been reported by individual[39,50] and meta-analytic[54] studies of age effects in the depression gap, which suggest that the depression gap peaks around age 13-15, then decreases but remains significant throughout adulthood. This peaking corresponds with the onset of puberty, which marks significant changes neural and biological systems that influence attention, sensory experiences, motivation, and social behaviors.[138] Puberty also leads to substantial changes in the social context of adolescents, marked by increases in psychosocial stressors and interpersonal conflict among peers.[139] The development of secondary sex characteristics and other physical changes, such as acne or increased adipose

tissue, serve as additional sources of potential negative social interactions.[140] These changes have been shown to increase the risk of depression and anxiety, especially in adolescent girls,[141,142] whose experiences may be exacerbated by depressogenic coping strategies such as rumination.[143,144]

In addition to identifying age effects overall, the interaction between age and study year indicated that the depression gap has increased among the youngest respondents since 1982. These results align with previous studies showing that the adolescent depression gap has been increasing and emerging at earlier ages for several generations.[38,40,145] One large study reported that the emergence of the depression gap decreased from age 29, among women born between 1936-1945, to age 14, among women born in 1966-1975.[146] In the present study, studies with respondents younger than 13 did not report a significant depression gap,[110,117] suggesting that the depression gap did not appear to be emerging at earlier ages than previously reported.

The magnitude of the depression gap appears to be increasing among adolescents overall, concordant with previous findings.[147] Causes of these trends are not clear, though changes in the adolescent social environment have been hypothesized. The prevalence of online harassment and bullying has increased over the past 20 years, and is more frequently experienced by girls.[148] While social media use entails a diverse set of exposures that can have positive effects on adolescent self-esteem,[149] problematic use is more common among girls,[150] though it is inconsistently linked to depressed mood.[151,152] Broader economic trends, such as the affordability of higher education,[153,154] credit card debt,[155] as well as macroeconomic and labor market instability,[156,157] may also be related to increasing depression rates among adolescents and young adults, though it is unclear if or how these effects might be specific to girls.

The findings of this meta-regression should be interpreted in light of several limitations. First, the majority of national samples were cross-sectional design, and were only able to assess prevalent depression status. Incident depression studies following children into adolescence and adulthood would supplement this and other studies of temporal trends across samples, by describing the complex etiology of depression and the depression gap. Second, there was significant heterogeneity in age ranges across studies. Any attempt to cross-classify studies by age would inevitably require a compromise to find cut-offs that were

conceptually meaningful but also yielded adequate sample sizes. In other words, a consequence of making the age group categories comparable across studies (i.e., observed ages) involved truncating the age ranges within each sample (i.e., true ages), potentially introducing measurement error because observed age range was sometimes different from the true age range. This measurement error was likely non-differential as it was not related to the depression gap outcome, so any bias would attenuate age estimates, potentially masking age differences as well as evidence of age by time interaction estimates. In this study, the reported age effects were robust to an alternative set of age groupings (i.e., 10-17, 18-25, 26-35, 36-45, 46-55, 56+), suggesting that the age trends were not artifactual. Third, despite being robust to differences in depression instruments overall, it is possible that differences across diagnostic criteria, which changed from DSM-III to DSM-5 during the study period, may partially explain the observed heterogeneity over time. There are two reasons to believe that these changes did not explain the study findings, however. First, any effect of diagnostic definitions would likely be consistent across all age groups. Second, in a sensitivity analysis, age by time interaction was tested among only studies that used DSM-IV depression (k=71), and the results were consistent with the overall sample. Finally, among included studies of diagnostic depression, there were missing data and evidence of potential publication bias, which may have distorted the summary estimates of the depression gap. However, evidence from multiple imputation models, and trim-and-fill sensitivity analyses suggested that this bias was minimal.

In conclusion, with a sample of 813,189 respondents, representing eight decades of age, and spanning a time period of 35 years, the present study finds evidence of a persistent depression gap, highlighting a major health disparity between women and men that may be increasing in the youngest ages. Future research is needed to understand the causes of these trends in greater detail, in order to inform depression prevention and treatment efforts, and reverse potentially growing depression disparities.

VI. Changes in the depression gender gap from 1992-2014: cohort effects and mediation by gendered social position

INTRODUCTION

Depression, the gender depression gap, and social stress theory

Major depressive disorder is a persistent[3] state of low mood and apathy that can affect the thoughts, behavior, feelings, and sense of well-being of the afflicted.[1] Compared with men, women have a higher incidence and prevalence of major depression[37,50,71] and depressive symptoms[54,158] throughout the world (hereafter referred to as the depression gap), though some evidence suggests that the gap may be changing over time.[38–40]

In studies of the depression gap, depression is often defined by a diagnostic threshold,[14] however there is considerable heterogeneity in the definition and operationalization of the depression construct. In addition to diagnostic criteria, studies of the depression gap also define depression as psychological distress,[52] operationalized using more general measures of somatic or internalizing symptoms.[53] Each of these measures represent related constructs, though differences may potentially influence the magnitude of the gap. Existing studies generally converge on higher mean symptoms among women compared with men.[54] In spite of this potential variation, I will hereafter refer to gender differences in these measures as the depression gap.

In studying the causes of the depression gap, many approaches have tested whether known causes of inter-individual variation in depression are more prevalent among women than men. For example, if childhood adversity increases depression risk generally, and women experience more childhood adversity than men, then childhood adversity may also explain the depression gap. To date, putative causes of inter-

[3] Symptoms must be present for at least two weeks to meet current criteria for major depressive disorder

individual variation in depression risk include genes,[159] hormones,[160–167] and stressful life events,[168–174] however, there is no evidence to support genetic[23–25] or hormonal[15,16] causes of the gender depression gap. Also, the conclusions about whether gender differences in stress explain the depression gap are dependent on what stressors are included in the stress inventories, therefore evidence regarding stress as an explanation for the depression gap is inconclusive.[175–178]

Instead of relying solely on differences in the type and frequency of stressors that men and women experience, social stress theory focuses on ways that group differences in social position may act as an upstream determinant of stress and the depression gap.[58] The theory posits that social categories, such as race, sexual orientation, and gender, confer access to resources and advantages differentially based on member's relative position within each category.[26] In other words, social position itself may be a source of stress and may determine the availability of resources to respond to stress.[172]

Women have historically experienced a disadvantaged social position compared with men.[28] While evidence indicates that women and men have similar levels of exposure to general life stressors, the stressful events to which women are exposed are more likely to cause depression, partially because women have fewer resources to respond to stress,[179] and are socialized to use more depressogenic coping strategies.[143,180,181] Thus, the conceptual model, shown in Figure 2.1, positions gender as a cause of social position, and social position as a cause of greater incidence and persistence of depression through stress-related pathways.[182]

In part, gendered social position is created and reinforced by gender norms,[183,184] which are socially acceptable behaviors defined for individuals in a given social and historical context.[185] Gender norms are influential from an early age,[34,41,42] as individuals learn, shape, and respond to a normative set of beliefs regarding the nature and appropriate behavior of women and men.[186] Part of this socialization includes the roles that are traditionally expected of women and men. In the United States, contemporary notions of "traditional gender roles" were first formalized during the industrial revolution, and dictate that men obtain a formal education then enter the paid workforce, while women provide unpaid domestic labor, primarily in the form of housework and childcare.[185] Traditional gender roles reinforce women's disadvantaged social

position relative to men, as the gendered division of paid and unpaid labor restricts the opportunities of women to attain an education and engage in the labor market.[183] Thus, the divisions or relative composition of these roles among women and men may serve as indicators of gendered social position.

Gendered social position may influence depression risk through both absolute and relative deprivation.[187] At an absolute level, gender roles determine access to personal (i.e., not shared within a partnership) socioeconomic resources[45,60,125–128] (e.g., income) and psychosocial resources, such as self-efficacy[129,130] and self-esteem.[131] These resources reduce exposure to stressors,[32] and also mitigate the effects of stressors[29–31] in ways that influence the risk of depression.[28,35,36,133] At a relative level, gendered social position reflects the maldistribution of power, authority, and opportunity between men and women.[183] This is exemplified by the gendered wage gap, which quantifies the extent to which men and women with equal effort or qualifications are differentially rewarded in the workplace.[188] Distinct from the effects of absolute resource acquisition, these relative inequalities represent an additional source of depression risk, by reinforcing gender differences in the opportunity to attain equal social position in the division of workplace and domestic roles.[132,133]

Changes in women's social position may influence the depression gap over time

Over the past 50 years, education and employment opportunities have become increasingly available to women, coinciding with changes in gender norms. Women surpassed men in earning a college degree in the early 1990's.[43,44] Since 1955, employment rates among working-age women nearly doubled to 70%,[59] and today, women are nearly as likely to be employed as men.[60] The time women spend on housework has also decreased.[61,62] All of these changes have been accompanied and facilitated in part by a suite of changes in women's domestic roles – women also increased birth control use,[64,65] waited longer to marry,[63] had fewer children,[66] and had children at later ages.[46] Childbearing significantly interrupts education and workplace experience trajectories, thus a woman's decision to delay or reduce these domestic obligations likely increases her opportunities to pursue education and career-oriented work.[60] Based on

social stress theory, changes in these indicators of gendered social position (education, employment, housework) should predict decreases in women's risk of depression, thereby narrowing the depression gap over time.

Changes in the depression gap over time can be characterized as a birth cohort effect, capturing the cumulative effects of exposure to specific historical conditions from birth onward.[189] The cumulative nature of these effects means that the experience of gender socialization during childhood and adolescence has long-term implications for depression. Historical changes in these conditions, evidenced by changing gendered social position, may yield different cumulative effects for depression across cohorts.

There is some evidence that the depression gap has changed; however, evidence is limited and inconsistent. Early retrospective studies of DSM-III depression reported inconsistent results across cohorts born from 1905-1965; some showed that the depression gap was narrowing,[190,191] and some showed it was stable.[40] Evidence of narrowing depression symptom levels have been reported in longitudinal[39,122,192] and cross-sectional[38] studies of more recently born adults. One meta-analysis of symptoms reported a decrease in Children's Depression Inventory scores across 1964-1988 birth cohorts, though this study was limited to childhood depression.[67] Another narrative review of studies of children and adolescents reported that internalizing symptoms (measured with the General Health Questionnaire,[193] and the Strengths and Difficulties Questionnaire[194]) were increasing among girls and boys.[147] Neither of these reviews reported whether changes over time differed between boys and girls. A meta-analysis of depressive symptoms among women and men ages 12-70+ reported that gender differences increased from 1991-2014, though the authors did not specifically examine cohort trends.[54]

Overall evidence regarding the degree and nature of changes in the depression gap over time is inconclusive, in part due to four limitations. First, individual studies of cohort effects often rely on retrospective reporting of symptoms, which may introduce recall bias, and/or obscure incident vs. recurrent cases. Second, follow-up periods in longitudinal studies are often too short to identify temporal trends in depression. Third, studies do not always report changes in depression levels separately for male and female respondents. Reporting the gap without stratified estimates may obscure these important trends, and thus

limit the ability to study potential sources of the changes underlying the depression gap. Changing social position may decrease women's depression, while potentially influencing men's depression at the same time, influencing the magnitude and variation in the depression gap. For example, changing gender norms are also redefining men's domestic responsibilities. These changes may be a source of chronic stress, especially when they conflict with men's expected role as the breadwinner in a partnership.[61,195] On the other hand, norm changes may benefit men. They may gain psychosocial resources from increased opportunities to care for their children, and share in household labor,[196] which may decrease their depression risk. Fourth, less attention has been paid to gender differences in levels of depression symptoms, compared with the gender gap in diagnostic depression. Examining cohort effects in gender differences in symptom-based assessments of depression will shed important light on gender patterns in sub-threshold levels of depression, which can cause significant impairment,[69] and also increase the probability of progression to major depression.[197–199]

To address the above limitations, I investigated the temporal trends in the gender differences in depression symptoms in a large prospective sample. These individuals comprise several sequential birth cohorts, followed over time through the period of early- to middle-adulthood as they establish education, employment, and division of domestic labor patterns. I hypothesized that 1) there have been temporal changes in the depression gap across birth cohorts in recent decades, 2) those changes were due to decreasing depression rates among women, and 3) decreasing depression rates were mediated primarily by increasing education and employment rates among women compared with men, as well as increasing equity in the division of housework between men and women.

METHODS

Sample

Data were from the National Longitudinal Surveys (NLS), two ongoing prospective surveys of employment, education, domestic labor, and health of American adolescents and adults. Detailed information for each survey has been previously published.[200,201] The source population, birth year range, interview waves, sample sizes, gender distributions, and attrition rates are summarized in table 2.1. The NLS surveys were combined to create a single analytic sample of respondents interviewed biennially from 1992-2014, the period when depression symptoms were asked of both men and women.

Data cleaning, processing, and imputation

NLS data are processed and cleaned by the Bureau of Labor Statistics, which maintains a publicly available repository of data and documentation.[202]

At each wave, respondent data were either non-missing, missing with a 'reason for non-interview', or missing with no reason. 'Reasons for non-interview' were recorded by NLS interviewers, with the following responses: deceased, unable to be reached for that wave, or refused multiple interview requests. Some respondents who were missing at one wave (and not deceased) were contacted and subsequently interviewed at a later wave whenever possible.

After these exclusions, the final dataset included 13,666 respondents at baseline with 47,646 years of observation over the study period (3.7 observations per respondent). Within this dataset, 15% of outcome responses were missing. In the main analysis, missing data were multiply imputed and combined with corrected standard errors.[203] Ten imputation models were run, using the following variables: gender, age, birth year, and all valid outcome data. Imputed model estimates were compared to unimputed estimates in order to examine the robustness of analytic models to the degree of missing data.

Measures

Independent variables

The primary independent variables were the respondent's gender (male/female) and their birth year, grouped into four a priori defined 10-year periods to create successive birth cohorts from 1955-1994. However, there were no respondents for birth years 1955-56 and 1966-71, creating gaps within these two birth cohorts. Overall, birth cohorts included the following years (0=1957-1964; 1=1965, 1972-1974; 2=1975-1984; 3=1985-1994). This categorization is conventional in demographic research and distinguishes cohorts roughly according to historically meaningful groups, from "Baby Boomers" (1955-1974) through early "Millennials" (1985-1994).[204,205] The person-years of observation for each birth cohort overall and stratified by gender are detailed in table 2.2. Together, gender and birth cohort were used to estimate cohort effects on the depression gap.

Dependent variables

Center for Epidemiologic Studies Depression Scale

The primary dependent variable was the 7-item Center for Epidemiologic Studies Depression (CESD) symptom score.[206] In completing the CESD, individuals rate how often over the past two weeks they experienced symptoms associated with depression, such as restless sleep, poor appetite, and feeling lonely (score range: 0-21, higher scores indicate more symptomatology; all items are listed in Supplementary Appendix 2.1).[207] A CESD score of 8 or higher represents high-CESD symptoms.[208] As an instrument to identify individuals with probable depression, the 7-item CESD has demonstrated high internal consistency,[209] good sensitivity, specificity and construct validity,[210–212] in community samples,[213] for different age groups,[214–217] and genders.[218] The CESD scores of individuals within each cohort were averaged at each wave. The depression gap was operationalized as both differences in CESD scores between women and men, and the risk difference of high-CESD scores for women minus men. In both forms, a positive difference indicates women's higher scores.

Mediators

Three variables were considered as indicators of relative gendered social position: the ratio of women to men with a college degree or more (i.e., *college ratio*), the ratio of women to men who are full-time employed (i.e., *employment ratio*), and the ratio of hours of daily housework reported by women vs. men (i.e., *housework ratio*). Each cohort-level indicator was operationalized in relative terms in order to highlight changes over time as increasing or decreasing gender parity in gendered social position. Further, indicators were defined at the population-level as they represent indicators of the overall context of gendered social position at a given historical timepoint. For example, the ratio of college completion captures to some extent the equality of opportunity for women to attain a higher education, rather than the individual-level rates of college completion within the sample. I hypothesized that this relative effect would be related to changes in the depression gap. Mediator data were incorporated from external nationally-representative data, described in detail below.

The *college ratio* was defined as the proportion of women vs. men from each birth cohort who reported attaining a college degree or more. Data were incorporated from the US Census Current Population Survey (CPS) annual historical data.[219]

The *employment ratio* was defined as the ratio of annual rates of full-time year-round employment of women vs. men. These data were also incorporated from the US Census CPS historical data,[219] as the average annual labor force participation rate among noninstitutionalized civilian adults age 20 or older.

The *housework ratio* was defined as the ratio of the average number of hours spent doing housework in a week reported by women vs. men.[220] Data were incorporated from a series of harmonized studies[4] that measured time-use decennially from 1965-1995 and annually from 2003-2018.[221] Housework

[4] Studies include 1965-66 Americans' Use of Time Study; 1975-76 Time Use in Economic and Social Accounts; 1985 Americans' Use of Time; 1995 Electric Power Research Institute (EPRI) Study

included activities such as cooking, cleaning, yard work, and shopping, but not time spent providing child care, in order to apply to respondents without children in the analysis.

In each cohort, members were assigned the value of each mediator that was extant in the population when they were between 20 years old. For example, education level for the 1955-64 birth cohort corresponded to the ratio of women to men who earned a college degree from 1975-1985. This range was chosen to represent the period when college, employment, and housework statuses were most salient for individuals, while also maximizing the range of overlap in ages across cohorts.

Confounding variables

Conceptually no variables met criteria for confounding of the relationship between cohort and depression or gender and depression (i.e. no variables cause birth cohort or gender). In contrast, I considered confounders of the mediator-outcome relationship. There are likely other causes of changes in education, employment, and housework ratios, apart from gendered social position that are also related to the depression gap, such as exogenous macroeconomic trends (e.g., wage suppression and the increasing necessity of dual-income households[222]). To control for these sources of confounding and estimate a valid indirect effect, the mediation models were adjusted for the population annual unemployment rate (defined as the proportion of the population actively seeking current employment) and marriage rate. The Directed Acyclic Graph depicting the causal structure underlying the analytical model can be found in Supplementary Figure 2.1. Models were also adjusted for respondent age, the mean age within each cohort, to account for inter-cohort age differences,[223] and the gender-specific values of each mediator to control for the absolute effect of each mediator and isolate the relative effect of each mediator. The employment ratio mediation model was also adjusted for the gender-specific average weekly number of hours of paid work.

Analysis

Descriptive analysis

The distributions of the study variables were summarized as means and standard deviations (SD) for continuous variables and percentages of categorical variables. Distributions were calculated for the overall sample, stratified by gender, and further stratified by 10-year birth cohort.

Hierarchical mixed modeling

I used a series of hierarchical mixed models to estimate the magnitude of the gender depression gap over time, and its relationship with birth cohort and education, employment, and housework ratios. These models accounted for the nested structure of the data, and also allowed for the specification of random effects, which were used to account for the proportion of total variation that is due to within-individual CESD scores over time, separately from the variation in between-individual CESD scores.[223,224] Like with standard linear models, hierarchical models assume linearity between dependent and independent variables, and that model errors are statistically independent, normally distributed, and homoscedastic.[225] To assess the functional form of CESD scores over time, I first fit a model with age as an independent variable and tested additional polynomial age variables to determine the best-fitting model using ANOVA. The analysis was then implemented in three sequential steps. First, I examined the overall depression gap (i.e., gender differences in CESD scores). Second, I tested whether the depression gap differed across cohorts (see Equation 1). Third, I tested whether education, employment, and housework ratios mediated the depression gap across cohorts (see methodological appendix, Equations 2 and 3).

[1] $$\mathrm{CESD_{ti}} = \gamma_{00} + \gamma_{01}\mathrm{G_j} + \gamma_{02}\mathrm{C_j} + \gamma_{03}\mathrm{cohA_j} + \gamma_{04}\mathrm{GC} + \mathrm{A_{ti}}(\gamma_{10} + \gamma_{11}\mathrm{G_j} + \gamma_{12}\mathrm{C_j} + \mathrm{w_{1i}}) + \mathrm{e_{ti}} + \mathrm{w_{0i}}$$

Where G is gender, C is birth cohort, A is age, and cohA is the mean age within each cohort, to account for inter-cohort differences in age. In equation 1, γ_{00} is the estimated CESD score with all covariates at their reference levels, γ_{01} is the mean gender difference in CESD score, γ_{02} is the mean cohort difference in CESD scores, γ_{03} is a covariate to adjust for mean age differences across cohorts, γ_{04} is the between-cohort differences in the gender differences in mean CESD scores, γ_{10} is the expected change in CESD score with age, γ_{11} is the between-cohort gender differences in CESD score changes, γ_{12} is the change with age between cohorts. The random coefficients include: e_{ti} which is the within-individual variation in CESD scores, w_{0i} is the between-individual variation in initial CESD score, and w_{1i} is the between-individual variation in changes in CESD scores by age, after controlling for cohort differences. All variance terms are assumed $e_{ti} \sim N(0,\sigma^2)$, $w_{0i} \sim N(0, r_{u0})$, and $w_{1i} \sim N(0, r_{u1})$.

Separate models estimated continuous CESD scores using linear models and binary high-CESD symptoms outcome using log-binomial models. Birth cohort was first considered as a set of indicator variables, to examine the linearity of changes in CESD scores between cohorts. All continuous variables were centered to facilitate interpretation.[226] Model fit was assessed using log likelihood and Bayesian information criterion (BIC) statistics.[224] Random intercepts and random slopes were estimated and included if they explained more than 10% of the total model variance.[227,228] Further model fitting details are detailed in the methodological appendix.

Mediation

In analytic steps 3-4, I conducted a mediation analyses to test whether three indicators of gendered social position mediated the gender gap in CESD scores over time. Separate models were estimated for each mediator.

To implement the mediation analysis, I first tested for additive statistical interaction between cohort and each mediator, to determine the method needed to estimate a valid indirect effect.[229] Next, I specified

a mediator model to estimate the conditional distribution of the mediator given the exposure, and an outcome model to estimate the conditional distribution of the outcome given the exposure, mediator, and observed confounders. Each model's estimated parameters were used to decompose the direct and mediated (indirect) parameter estimates in the total mediation model.[230] This approach estimates standard errors using quasi-Bayesian Monte Carlo methods based on normal approximation, to construct 95% confidence intervals around both the direct and indirect parameter estimates and the proportion of the total effect that was mediated[230] for both linear and binary outcomes.[231,232]

All statistical analyses were conducted in R (version 3.5.1).[233] All hierarchical linear models were implemented using the "lme4" package,[234] multiple imputation was implemented with "merTools"[235] and "amelia"[236] packages, and mediation models were implemented using the "mediation" package[231] using a maximum likelihood estimator robust to non-normality.[224]

RESULTS

Descriptive

Descriptive statistics, overall and stratified by gender, are presented in table 2.3. Compared with men, women reported higher CESD scores (4.57 (SD=4.2) vs. 3.63 (SD=3.7)), and a higher prevalence of high-CESD symptoms (0.20 (SD=0.4) vs. 0.13 (SD=0.3)). Overall the proportions with a college degree or more were similar between women and men (28% (SD=0.08) vs. 27% (SD=0.03)), while women were less likely to be employed (54% (SD=0.04) vs. 72% (SD=0.03)). On average, women did nearly three times more housework than men (20.3 (SD=6.1) vs. 8.1 (SD=1.9) hours per week). There were no gender differences in the annual unemployment rate (6% (SD=0.02)) and marriage rate (58% (SD=0.05)).

Gender differences in study variables comparing women vs. men, further stratified by birth cohort are presented in Supplementary Table 2.2. Gender differences narrowed across cohorts in both CESD scores, (1955: 4.38 vs. 3.24; 1985: 4.44 vs. 3.85) and the prevalence of high-CESD symptoms (1955: 20% vs. 12%; 1985: 17% vs. 13%). The ratio of women to men who earned a college degree changed from

minority women in the 1955-64 cohort (21% vs. 26% among men) to majority women in the 1975-84 cohort (38% vs. 29%). Gender differences in full-time employment narrowed across cohorts, due to increases in employment rates among women (41% to 55%) and decreases among men (80% to 68%). Housework ratios among women vs. men also narrowed, due to declining time spent by women (25.0 to 16.6 hours per week) and increasing time spent by men (6.9 to 9.1 hours per week).

Hierarchical Mixed Models

Main effects

Based on a priori criteria, the best-fitting model contained quadratic and cubic age terms, (χ^2=7.13, p-value=0.008) and included random intercepts, which accounted for 33-36% of the total variance in each model (additional model-fitting details are provided in Supplementary Table 2.3 and in the methodological appendix). Table 2.4 presents the estimated changes in CESD scores across cohort in model 1, and changes in the gender depression gap across cohort in models 2 and 3. In model 1, CESD scores decreased across more recently born cohorts (CESD difference= -0.15; 95% CI= -0.2, -0.09); across all cohorts, the gender depression gap was 0.96 points (95% CI= 0.86, 1.06). In model 2, compared with the 1955-64 cohort, the gender gap was not appreciably different in the 1965-74 cohort, (CESD difference= -0.25; 95% CI= -0.76, 0.26), however, the gap was 0.36 points lower in the 1975-84 cohort (95% CI= -0.63, -0.10), and 0.54 points lower in the 1985-94 cohort (95% CI= -0.79, -0.28). These coefficients approximated a linear decrease, therefore, subsequent analyses assumed linear changes in the gender depression gap across cohorts. In model 3, assuming a linear change, the gender depression gap narrowed by 0.18 points in each birth cohort (95% CI= -0.26, -0.10).

Cohort effects in CESD scores among gender-stratified models are presented in Table 2.5 and visually in Figure 2.4. Among women, CESD scores decreased by 0.41 points across cohorts (95% CI= -0.63, -0.19), while among men, there was a 0.05-point decrease (95% CI= -0.40, -0.04). When compared

to the dataset with no imputation, results were not meaningfully different in magnitude of the model estimates, although standard errors were slightly smaller overall.

Mediation analysis

There was no evidence of additive interaction between the exposure and each mediator, indicating the indirect effects were equal across all levels of the mediator. Table 2.6 details the results of mediation of the gender depression gap across cohorts by three indicators of gendered social position (measured as ratios comparing women to men). In model 1A, an increasing ratio of college degree attainment (i.e., as women became more likely to earn a college degree than men) mediated 39% of the gender depression gap across cohorts (95% CI= 0.18, 0.78). In model 1B, the female to male employment ratio did not mediate the gender depression gap across cohorts, nor did the female to male housework ratio (model 1C).

In the analysis estimating the gender gap in high-CESD symptoms, findings were generally similar to the models estimating CESD score differences. The risk of high-CESD symptoms decreased across cohorts overall (see Supplementary Table 2.4), among both women (RD=-0.32; 95% CI= -0.42, -0.23) and men (RD=-0.17; 95% CI=-0.27, -0.08) (see Supplementary Table 2.5).

DISCUSSION

This study leveraged change over time to test whether gendered social position influences the gender depression gap. There were four central findings: 1) the gender depression gap decreased by an average of 0.18 points between each 10-year birth cohort from 1955 to 1994; 2) the decreasing pattern in the gender depression gap was due primarily to decreasing CESD scores among women; 3) relative to men, women's social position, indicated by college completion, employment rates, and the division of housework improved across cohorts; 4) changes in female-to-male education ratios appeared to explain up to 39% of the observed changes in the gender depression gap.

To my knowledge, this is the first cohort analysis of the adult depression gap in cohorts born after 1975 in the US. Previous studies are based on samples with cohorts born from 1905-1975, which are substantially earlier than the 1955-1994 cohorts represented in the present study. Among them, some reported similar trends of a narrowing gender depression gap among younger cohorts.[39,122] In the Epidemiologic Catchment Area study, researchers found that gender differences in DSM-III depression risk had been decreasing across 1905-1965 birth cohorts.[71,72,123] In contrast, other studies of cohort effects in the depression gap have reported no effects or an increasing gap. For example, the National Comorbidity Study Replication found no evidence of change in the gender gap in DSM-III-R depression across 1936-1975 birth cohorts.[40] Using longitudinal data from the Americans' Changing Lives study, Yang and Lee found evidence that the depression gap increased across cohorts born in 1915 to 1955.[124] In a repeated cross-sectional study, the depression gap appeared to narrow from 2005-2014, among young adults ages 18-25.[38] Compared with much of the existing literature, the findings of this study describe trends in gendered social position among more recently born cohorts through the end of the 20th Century.

Alongside the narrowing gender depression gap, gender ratios in college degree attainment, employment, and the division of housework also narrowed or changed direction in the case of college degree attainment, highlighting three meaningful trends in the improvement of women's social position. The college ratio reversed from majority male in the 1955 cohort to majority female in the 1975 and 1985 cohorts, driven by a near doubling in women's college completion rates, from 21% to 38%. The employment ratio narrowed due both to increases in employment among women and to decreases among men. Despite this, women in the youngest cohort were still less likely to be employed than men. The decreasing housework ratio across cohorts was mostly due to women doing less housework, rather than men doing more. Levels among men increased from the 1955-1965 cohorts, but remained relatively stable thereafter. This suggests that the amount of unpaid work is decreasing overall, replaced by both technological improvements that reduce labor demand,[237] and by paid domestic workers.[61,238] Therefore, while women's share of housework has decreased, the trend toward equity has not come from substantial increases in housework by men.

I found additional evidence to partially support the hypothesis that changes in women's social position relative to men, specifically gender ratios of college completion, mediated changes in the gender depression gap and depression scores among women. Increases in women's college completion, relative to men, mediated 39% of the gender gap in CESD scores across cohorts. Researchers have reported similar associations in cross-national comparisons of the depression gap and gendered social position.[239–242] In one such study, Seedat and colleagues found that the depression gap was lower among countries with greater gender equity, based on an aggregated measure of women's education, employment, marital timing, and use of birth control.[243] The findings in the present study were robust to the adjustment in the base rates for each mediator, and mediation of cohort differences was specific to women's depression scores in stratified models. Results suggest that women may gain additional mental health benefits due to an increasing parity in gendered social position, distinct from the benefits afforded by absolute increases in women's socio-economic status. This highlights the importance of relative measures of social position between men and women.[183,244] The results of the mediation analysis in this study suggest that gender parity in education attainment is beneficial the depression gap. This finding is supported by previous research that has shown that parity in educational attainment between men and women signals a more equal opportunity structure,[245,246] which may act to decrease the depression gap through social stress pathways.[247,248]

In contrast, changes in employment ratios did not mediate the CESD cohort effects. While employment opportunities expanded for women, a gap in the employment ratio remained. This gap highlights one of many persistent barriers to equal participation in the labor market[60,249–252] that may negate any positive effects of expanded opportunities overall. In addition, the effects of trends toward equity in housework at the population level did not mediate the observed decreases in the depression gap. Most studies of the effects of the division of housework on depression have focused on comparing housework levels between men and women within (heterosexual) marriages.[253,254] These studies have found that the perception of equity in housework inhibits depression for women in those households.[131]

Overall, there was inconsistent support of the hypothesis that changing gender social position would partially explain variation in the depression gap. There are several possible conceptual and

methodological explanations for these inconsistencies. First, changing gendered social position was operationalized with only three indirect indicators. While education, employment, and housework trends encompass meaningful changes in gendered social position, additional indicators would likely capture other important aspects of the gender landscape that are related to variation in the depression gap. One potentially important element is the gendered division of childcare. Time spent on childcare among fathers also increased during this period.[255,256] Further, this trend may yield unique mental health benefits of expanded gender role opportunities for men,[131,196,257,258] such as the slight decrease in men's CESD scores observed in this study. Second, education, employment, and housework ratios may have changed due to other economic or social causes not measured in this study. If these causes were also related to the depression gap, they would cause unmeasured confounding of the mediator-outcome relationship. In order for unmeasured confounding to explain why factors such as employment and housework ratios did not mediate the interaction between cohort and gender on depression, such unexplained factors would need to be positively related to employment ratios as well as the depression gap, while also being negatively related to housework ratios and positively related to the depression gap. As an example, trends in multi-generational family structures have increased since the 1970s,[259] which has allowed a greater number of women to obtain full-time employment, increasing the employment ratio, and also relieving women of some of their domestic labor burden, decreasing the housework ratio.[260] At the same time, older cohabitating adults (e.g., the parents of middle-aged working adults) may themselves become sources of additional domestic labor, increasing depression symptoms among the women who care for them[261] and the depression gap in turn. I attempted to limit this bias by controlling for unemployment, weekly hours of paid work, and marriage rates, however, confounding may still be a source of bias. Third, the mediator values were assigned to respondents based on Census-based population averages when respondents were age 20, in order to capture the age when the context of gendered social position was most salient for respondents. It may be that this assumption was satisfied for education, given college attendance and completion typically occurs around age 20, however, employment and housework ratios may be more important for women in their 30s and beyond. Unfortunately, I was not able to sufficiently vary the values to correspond with other ages,

given the age ranges across cohorts (see below for more details), however, future work should interrogate this potential source of measurement error with additional data sources.

This research should be interpreted in light of several limitations. First, respondents' birth years were not evenly distributed within each cohort. Specifically, no respondent was born in 1955 and 1956, as well as from 1966-1971. Therefore, samples sizes for the 1965-74 cohort were much smaller than those of other cohorts, and estimates were interpolated over these missing years. When modeled as indicator variables, the depression gap clearly decreased among the 1985 and 1995 cohorts, which had coverage across all birth years. This supports the main conclusion that the depression gap is decreasing over time. Related, the age ranges of each cohort did not completely overlap, and estimates should be interpreted with the acknowledgement of off-support inference by age. Future work should seek to replicate this study using data with more complete coverage for these particular birth cohorts. Third, depression scores as measured in this study refer to short-term (2-week) prevalence, which may not reflect true depression status over the 2+ year period between interviews. Fifth, depression scores were measured with a symptom scale, which may measure sub-threshold depressive symptoms but not diagnostic criteria. Depression is a complex construct to measure and, even though the CESD has been widely used as a measure of depression for over 40 years,[207] including in studies of the gender depression gap,[54] it is likely a measure of more general psychological distress and demoralization. Finally, the focus of this study is almost entirely on the effects of changing social context on depression. Men face an excess of externalizing mental health and substance use problems,[262] and researchers have argued that the gender depression gap simply reflects different manifestations of the same stress response, rather than greater psychiatric morbidity among women.[263–265] Indeed, similar methods have been applied to understand historical variation in heavy and disordered alcohol use in men and women.[266] While the goal of the present study was to focus specifically on depression, future work to integrate these two bodies of research, potentially from a trans-diagnostic[267] or dimensional[268] perspective could provide a more comprehensive application of social stress theory to the understanding of social context and health.

The scope of this analysis was to explore the variation in the gender depression gap in the US population overall, but future work might acknowledge and examine heterogeneity in the gender depression gap changes across other intersecting social categories. Differences likely exist within and between other important categories that were not comprehensively measured in this study, such as racialized status, gender identity, and sexual orientation. Each group is influenced by distinct norms and has experienced unique social changes during the study period. Future research with detailed attention to these and other groups would likely prove an important contribution to the social stress literature from an intersectional perspective.

This study is strengthened by the prospective measurement of depression symptoms, and the incorporation of population-representative measures of education, employment, and domestic labor status. Prospective study designs may be especially important in cohort analyses, as cohort effects are highly sensitive to differential recall bias.[269,270] Also, studying the depression gap highlights the importance of gendered social position as a social determinant that is not directly observable at the individual-level, and examining population changes over time may allow effects of exposures that are ubiquitous in the population at any single point in time.

In this study, I find evidence to suggest that the opportunities and responsibilities dictated by gender norms are becoming less restrictive and women's social position is improving relative to prior cohorts. While important gaps remain in understanding these changes, this work highlights several mechanisms through which depression risk is reinforced, and identifies opportunities to reduce the depression gap between men and women.

VII. Bringing home the benefits: Do pro-family employee benefits mitigate the risk of depression from competing workplace and domestic labor roles?

INTRODUCTION

During latter half of the 20th Century, women's participation in higher education and the workplace drastically increased in the United States.[271] Today, compared with earlier generations, women are more likely to earn a college degree[245] and be employed full-time.[59] Domestic roles changed as well; women now spend less time doing housework,[61,62] wait longer to marry[63] and have children,[46] and have fewer children overall,[66] due in part to increased use of birth control.[64,65]

However, in spite of these historic changes, significant gender disparities in workplace and domestic labor remain.[272] For example, the birth of a child is more disruptive to the employment trajectories of working mothers, who are more likely to take parental leave or exit the workforce entirely to raise children than working fathers.[251,252] Furthermore, working mothers are often regarded as either unreliable employees or unreliable mothers,[273] while working fathers are rewarded.[274] Among men and women with similar workplace obligations, women still spend more time on daily domestic labor and childcare.[275–277] This imbalance persists among women who out-earn their male spouse,[278] suggesting that these patterns are not solely due to intrahousehold socioeconomic differences.

As a result of these incomplete changes, women are now more likely to hold concurrent workplace and domestic roles, which may affect mental health, including depression.[47] Depression in the US is prevalent, destabilizing, and approximately twice as likely to occur among women than men. Characterized by sadness, loss of interest, as well as psychomotor, sleep, and appetite changes, the negative impacts of depression on women's' lives is widespread. However, there is limited consensus on the impact of concurrent role obligations on depression, including whether the effects would be positive or negative. Some studies have found that multiple roles reduce depression risk by increasing sources of social support, self-complexity, and material resources.[35,279–282] Others have found that the potential for conflict- and overload-related stress increases with the number of roles, thus increasing depression risk.[134–137]

More specifically, holding multiple roles may have negative effects when the time, energy, and attention resources needed to satisfy one role compete with the resources needed to satisfy another role.[283]

In this context, employers have increasingly offered employee benefits as a way to address the burden from competing roles and to retain employees with children.[48] I refer to these as pro-family benefits, as they emerged to improve an employee's ability to manage work-family responsibilities. Pro-family benefits include paid family-leave, childcare subsidies, and flexible working schedules, among others. They are distinct from other employee benefits that are not intended to buffer the effects of competing roles (e.g., retirement pensions, health insurance).[284]

Evidence to date suggests that pro-family benefits may mitigate the burden of competing role demands, particularly among female employees. Women with access to pro-family benefits are less likely to experience work-family conflict,[285] competing responsibilities,[286] and job dissatisfaction[287] and as a result are more likely to remain employed,[288,289] and maintain pre-childbirth work hours.[290]

Pro-family benefits are associated with better physical health[291] and general well-being.[292] Positive mental health effects have also been shown,[293] however, the literature on the effect of pro-family employee benefits on depression is limited in at least three ways. First, studies typically consider exposure to a single policy, such as the availability and length of maternity leave[294,295] or work schedule flexibility,[296] rather than to multiple policies in the same population with attention to the potential cumulative effects of multiple benefits. Second, studies have not tested the specificity of the effect of pro-family benefits, by comparing them to the effects of policies that are not specifically family-related. Workplace benefits, as a whole, may be related to employee mental health through alternative pathways,[297–299] so any effect of pro-family policies may be a non-specific indicator of a generally positive workplace environment. In effect, the broader workplace context may confound any observed association between pro-family benefits and depression. Comparing the effects of pro-family benefits to other benefits will attempt to adjust for this source of confounding and strengthen the previous theoretical claim that pro-family policies mitigate the negative effects of competing roles to reduce depression risk. Third, studies often rely on cross-sectional data[300–302] which are limited due to a lack of temporality. Longitudinal studies are necessary to understand the potential

bias from reverse causation and selection. If depression and exposure to competing roles both increase the risk of selection out of the workforce, then estimates of the main effect of competing roles and depressive symptoms would be spuriously attenuated. Selection bias might also arise if individuals with depression and competing roles are less likely to work in jobs with employee benefits. This would also attenuate interaction estimates.

To address the above limitations, the present study examined if the burden of competing workplace and domestic roles increased women's subsequent risk of depression, and considered the extent to which that risk was buffered by access to multiple pro-family benefits in the workplace. The study addressed the following specific aims: 1) to estimate the relationship between competing roles (working and raising children vs. working and *not* raising children) and depression; 2) to assess whether that relationship varied by the presence of any pro-family benefits as well as the number of available benefits; 3) to assess whether any observed buffering effect by pro-family benefits was similar to the buffering effects due to non-family-related benefits; 4) to estimate the potential impact of two important sources of selection bias due to prior depressive symptoms and competing role status: selection out of employment and selection out of jobs with available employee benefits.

METHODS

Sample

Data were from the National Longitudinal Surveys (NLS), a series of prospective studies of employment, education, domestic experiences, and health of American adults. NLS data are processed and cleaned by the Bureau of Labor Statistics, which maintains a publicly available repository of data and documentation.[202] Detailed information for each study has been previously published.[200,201,303–306] The analytic sample comprised data from the five most recent biennial interviews (2006-2014) of the NLSY79 and NLSY79 Young Adult samples, ranging from ages 17-57, representing the time when childbearing and

full-time employment were most prevalent. The sample included only respondents who were employed at each wave, to ensure that the entire sample had the potential to be exposed to employee benefits, and avoid structural positivity violations.[5] For example, a respondent who reported being employed in 2006, unemployed in 2008 and 2010 and employed in 2012 and 2014 would contribute three years of observation to the study follow-up. The final analytic sample included 12,239 person-years for 3993 women (mean of 3.1 observations per respondent).

Measures

Center for Epidemiologic Studies Depression Scale

The primary dependent variable was the Center for Epidemiologic Studies Depression (CESD) symptom score. Symptoms were assessed using the 7-item CESD scale,[206] in which individuals rate how often over the past week they experienced symptoms associated with depression, such as restless sleep, poor appetite, and feeling lonely (score range: 0-21, higher scores indicate more symptomatology; all items are listed in Supplementary Appendix 2.1).[207] A dichotomous CESD score 8 or higher represents high-CESD symptoms.[208] As an instrument to identify individuals with probable depression, the 7-item CESD scale has demonstrated high internal consistency,[209] good sensitivity, specificity and construct validity,[210–212] in community samples,[213] in various age[214–217] and gender groups.[218]

Competing roles

[5] Respondents who were unemployed, self-employed in an unincorporated business, or enlisted in the military were not asked about employee benefits.

The independent variable was a measure of competing roles, defined as currently employed with at least one child (ages 0-17) living in the home. A person could contribute time as exposed and unexposed to competing roles during the study period, e.g., if they had a child (unexposed → exposed) or their child turned 18 (exposed → unexposed). The comparison group for those with competing roles was defined as those who were currently employed with no children living in the home.

Employee benefits

In the NLS, respondents reported the availability (yes/no) of nine employee benefits in their current job at each interview. Benefits were classified as either pro-family or non-family-related, in line with previous research, based on their hypothesized potential to alleviate some of the burden and stress due to competing workplace and domestic roles.[307] Pro-family benefits included: family leave, flexible scheduling, and employer-provided childcare. Non-family-related benefits included: dental insurance, medical insurance, life insurance, profit sharing, retirement pension programs, and training/educational opportunities. Each group of benefits was considered as binary (e.g., any vs. no benefits) and as a count variable (0 vs. 1, 2+). Benefits were tested for interaction with competing roles on CESD scores.

Confounding and selection bias

Figure 3.1 presents the analytic model including the structure of confounding and selection bias over time. The following variables were considered as confounders of the relationship between competing roles and CESD scores: continuous age, continuous education level (highest completed grade), continuous number of hours of paid work per week, continuous income, marital status (never, currently, formerly married) and race/ethnicity (Hispanic, non-Hispanic Black, other). In addition, several variables were considered as confounders of the relationship between employee benefits and CESD scores, including demographic (age, education), domestic (number of children (1, 2, 3+), number of children under age 5 (0,

1, 2+)), and workplace variables (hours of paid work per week, employer type (government, private sector, non-profit), income, and industry (21 categories, see Supplementary table 3.1 for a list of all industries)). Additionally, pro-family benefits may be related to CESD scores through alternative pathways not related to competing risk. To attempt to isolate the pathway of interest, unconfounded by (unmeasured) general workplace quality, non-family-related benefits were also included as confounders of pro-family benefits.

Because only employed respondents reported the availability of employee benefits, the sample was necessarily selected on employment status, which may have introduced selection bias. A sensitivity analysis attempted to quantify the magnitude of this source of selection bias by estimating the strength of the association between selection out of employment related to CESD symptom scores, high CESD symptoms, and competing role status in prior waves. Additionally, to account for potential bias from differential selection into jobs with pro-family benefits, I examined the probability of holding a job with pro-family benefits based on prior CESD symptom scores or high-CESD symptoms and competing role status.

Imputation

Within the sample of valid respondents, 15% of the outcome data were missing. I used multiple imputation with 10 combined datasets for these missing data. Standard errors were corrected to account for repeated analyses.[203] Imputation models included sex, age, year, and CESD scores from non-missing interviews. Imputed model estimates were compared with non-imputed estimates in order to examine the robustness of analytic models to the degree of missing data.

Analysis

The distributions of the study variables were summarized as means and standard deviations (SD) for continuous variables, or proportions for categorical variables for all available person-time over the study

period. ANOVA and chi-squared statistics were used to test whether the means or proportions of each variable differed by competing role status.

The analytic aims were addressed using a series of generalized estimating equations (GEE), to account for uneven follow-up periods and the clustering of observations within individuals over time.[308] All models were estimated for both continuous CESD scores and binary high-CESD symptoms, using linear and log-binomial models respectively, with cluster-robust standard errors.[309]

The first analytic aim was to estimate the relationship between competing roles (working and raising children vs. working and *not* raising children) and depression, and the relationship between any pro-family benefits and depression. This was accomplished by regressing CESD symptoms on competing roles status, both with and without adjustment for confounding variables. The association between pro-family employee benefits and CESD symptoms was estimated, in order to test the average effect of benefits on CESD symptoms

The second analytic aim was to assess whether that relationship varied by the presence of any pro-family benefits as well as the number of available benefits. This was accomplished with two steps: first, a model tested for interaction between competing roles and pro-family benefits, to examine whether pro-family employee benefits buffered the risk of competing roles on CESD symptoms. Interaction was tested using cross-product methods in linear models, and using the relative excess risk due to interaction (RERI) and 95% CIs[310] in the log-binomial models. Interaction was tested both before and after adjustment for non-pro-family benefits in order to assess the potential confounding due to (unmeasured) general workplace quality. This adjustment provided further evidence of the specificity of any buffering effects from pro-family benefits. I hypothesized the buffering effect of pro-family benefits would be robust to adjustment for non-family-related benefits. Second, in addition to directly testing for interaction, models were also stratified to examine the magnitude of the risk from competing roles in the presence vs. absence (and count) of pro-family benefits.

The third analytic aim was to assess whether any observed buffering effect by pro-family benefits was similar to the buffering effects due to non-family-related benefits. This was accomplished by estimating

the depression risk of competing roles, stratified by exposure to non-family-related employee benefits (both binary and counts), with adjustment for pro-family benefits. This was done to compare the buffering effects of this alternative type of benefits, which may affect the depression risk from competing roles, albeit through alternative pathways.

The fourth analytic aim was to estimate the potential bias due to selection out of employment and selection into jobs with available employee benefits, based on prior depressive symptoms and competing role status. This aim was accomplished with two sets of GEE models. One set of models regressed employment status on lagged CESD symptoms and competing role status, and one set of models regressed pro-family benefit status on lagged CESD symptoms and competing role status.

RESULTS

Study means and proportions for all available person-time over the study period by competing roles status for all study variables are presented in Table 3.1 and Figure 3.2. Overall, the prevalence of competing roles was 61%. Compared to those not in competing roles, women in competing roles were younger (47.1 (7.4) vs. age 48.9 (3.5)), completed fewer grades in school (13.4 (2.6) vs. 14.8 (2.7) years of school), earned less income ($38845 (33691) vs. $50559 (44562)), worked fewer weekly hours for pay (42.5 (7.9) vs. 44.2 (8.3)), were more likely to have pro-family benefits (85.8% vs. 83.2%) and less likely to have non-family-related benefits (85.4% vs. 89.1%), and reported higher CESD scores (3.99 (4.1) vs. 3.26 (3.7)) and a higher prevalence of high-CESD symptoms (17% (0.4)) vs. 11% (0.3)). Marital status, race/ethnicity, and employer type also differed among those with vs. without competing roles.

Aim 1: estimate the relationship between competing roles (working and raising children vs. working and not raising children) and depression, and the relationship between any pro-family benefits and depression

The effects of competing roles and pro-family benefits on CESD scores and high-CESD symptoms are presented in Table 3.2. After adjustment, women in competing roles reported a 0.56-point higher CESD

score (95% CI= 0.15, 0.97) and a 62% greater risk of high-CESD symptoms, compared with women not in competing roles (95% CI= 1.17, 2.25).

Women with any available pro-family benefits reported a 0.83-point lower CESD score (95% CI= -1.36, -0.31) and a 28% lower risk of high-CESD symptoms (RR=0.72; 95% CI= 0.57, 0.91), compared with women without pro-family benefits (see Table 2).

Aim 2: assess whether that relationship varied by the presence of any pro-family benefits as well as the number of available benefits

Table 3.3 presents interaction tests and stratified analysis of the effects of competing roles on CESD scores and high-CESD symptoms in the absence vs. presence of any pro-family benefits and the number of available benefits. In the linear models, the interaction between pro-family benefits and competing roles was associated with CESD scores, both unadjusted (B=-0.51, p=0.017) and adjusted (B=-0.44, p=0.023). In the binary models, there was no evidence of additive interaction, before or after adjustment for non-family-related benefits (RERI= -0.81 (95% CI= -2.18, 0.56) and RERI= -1.05 (95% CI= -2.8, 0.89) respectively) (see footnote a in table 3.3).

In the absence of any available pro-family benefits, those in competing roles reported 6.1 point higher CESD scores (95% CI=1.14, 11.1), compared with those not in competing roles. In contrast, among women *with* access to any pro-family benefits, there was no association between competing roles and CESD scores (difference=0.44; 95% CI=-0.2, 1.0). In examining the number of benefits, no association was found among those reporting exactly one pro-family benefit (CESD score difference= 1.1; 95% CI= -0.08, 2.26) or 2 or more pro-family benefits (CESD difference= -0.01; 95% CI= -0.87, 0.86). The results were similar in models estimating the risk of high-CESD symptoms. The confidence intervals were appreciably wider among those without (vs. with) pro-family benefits because of the low prevalence of no pro-family benefits.

Aim 3: assess whether any observed buffering effect by pro-family benefits was similar to the buffering effects due to non-family-related benefits

Table 3.4 presents the effects of competing roles on CESD scores and high-CESD symptoms stratified by the absence vs. presence of any non-family-related benefits and the number of available benefits. Women in competing roles *without* any non-family-related benefits reported 3.59 point higher CESD scores than those not in competing roles (95% CI=1.24, 5.95) while among women *with* access to these benefits, there was no association between competing roles and CESD symptoms. No association was found among those with exactly one non-family-related benefit (CESD difference= 2.09; 95% CI= -0.26, 4.44) and 2 or more benefits (CESD difference= 0.44; 95% CI= -0.73, 1.62). The results were similar in models estimating the risk of high-CESD symptoms. Compared with the dataset with no imputation, multiply imputed model estimates were not meaningfully different, although standard errors were slightly smaller overall. The confidence intervals were appreciably wider among those without (vs. with) non-family-related benefits because of the low prevalence of no non-family-related benefits.

Aim 4: estimate the potential impact of two important sources of selection bias due to prior depressive symptoms and competing role status: selection out of employment and selection into jobs with available employee benefits

Table 3.5 estimates the magnitude of bias due to selection out of the workforce at each interview, based on the CESD score and competing role status in the prior interview. The risk of becoming unemployed among those exposed to competing roles in the prior interview ranged from 1.051 in 2012 (95% CI=1.034, 1.068) to 0.997 in 2006 (95% CI=0.98, 1.014). The risk of becoming unemployed based on prior CESD scores was effectively null at every wave. The patterns among those with high-CESD symptoms were similar to the models of CESD scores. Selection into jobs with pro-family benefits vs. jobs without pro-family benefits was also considered. The risk of reporting available pro-family benefits among those exposed to competing roles in the prior interview ranged from 1.025 in 2012 (95% CI=0.983, 1.068) to 0.96 in 2006 (95% CI=0.932, 0.989). The risk of reporting available pro-family benefits based on prior CESD scores was effectively null at every wave. The patterns among those with high-CESD symptoms were similar to the models of CESD scores. Overall, selection risks were small due to competing roles and

null due to prior CESD scores, suggesting that the results were not substantially influenced by selection bias.

DISCUSSION

The present study examined the effect of competing gender roles on women's depression and whether that effect was buffered by the availability of pro-family employee benefits. There were four central findings: 1) among women in the workplace, those with competing gender roles reported a higher CESD score and a greater risk of high-CESD symptoms than those without competing roles; 2) the depression risk from competing roles was attenuated in the presence of pro-family benefits, supporting the buffering effects hypothesis; however, 3) the depression risk from competing roles was also attenuated in the presence of non-family-related employee benefits, suggesting that access to employee benefits in general, regardless of whether pro-family or not, is associated with reduced depressive symptoms among women in competing roles; 4) there was no evidence that the effects were attributable to selection out of employment or into jobs with pro-family benefits.

Popularized as a result of the 'Second Shift' faced by women,[271] the mental health effects of holding competing roles is still an active area of research in the current social context. Holding multiple roles can be good for mental health and well-being,[35,279,280,311] however, the present study found that competing roles were associated with increased depressive symptoms, suggesting that the competition for the resources needed to satisfy both workplace and domestic labor roles is adverse for mental health.[47] These findings are concordant with previous studies that have shown a detrimental effect of excessive role demands.[134–137,311,312] This competition has likely been increasing over time, as trends show that Americans work longer hours overall and earn a wage premium for overwork (more than 50 hours of work per week),[249] and women spend more time providing care to their children,[255] even into adulthood.[313]

In this context, workplace benefits intended to help employees balance competing domestic and workplace demands[48,314] have become more widely available.[315] The buffering effects of pro-family

employee benefits in this study are generally concordant with previous research, and extend the evidence base to documenting benefits for depression symptoms. Access to paid family leave increases the use of preventative health care,[316] and decreases fatigue, anxiety, and depressive symptoms among new mothers.[317] The use of flexible working policies improves mother-child bonding during early childhood,[318] decreasing the mother's depression risk in turn.[319] Research on the impact of the availability of employer-provided or subsidized child care on employee mental health is very limited, though there is indirect evidence of positive mental health effects of this specific benefit. Increasing the affordability of childcare increases employee retention, which likely reduces the depression risk associated with job turnover.[320,321]

The presence of non-family-related employee benefits also reduced the depression risk from competing roles, suggesting that the attenuation of stress from competing roles is not limited to benefits designed to specifically address role competition among working women with children. There are several potential explanations for these broader positive effects. First, the availability of personal health insurance improves access to preventive care and treatment for those at risk for depression, including new mothers.[322] Second, access to pensions and other retirement benefits decreases job changes and associated stress,[299] and may provide financial security among parents, who may be particularly concerned about future financial burdens. Third, jobs with a wide array of employee benefits are good jobs; they signal greater occupational prestige,[298] and are associated with higher employee self-esteem and general satisfaction[297] than jobs without benefits. All of these characteristics could potentially reduce stress and subsequent depression risk, related to competing roles. Though I did attempt to account for alternative explanations by adjusting model estimates for work hours, income, industry, and employer type, the non-specific nature of these effects highlights a broader positive impact of workplace policies to reduce depressive symptoms among women with children.

The analytic sample was restricted to employed respondents, which would cause selection bias if employment status was related to competing roles and depression. While I could not avoid this selection, given the study design, I attempted to account for the potential magnitude of the bias in a sensitivity analysis. To the extent that these selection pressures are measured and captured in the short-term (i.e., 2-

year lags), the magnitude of selection bias in this study appeared to be minimal. Also, it is also possible that healthier individuals may be more likely to select into jobs with more benefits, though related evidence to date suggests that selection effects are secondary and are not likely to explain the results of this study.[323] Indeed, the sensitivity analysis of selection into jobs with pro-family benefits found minimal evidence of bias.

This research should be interpreted in light of several limitations. First, employee benefits were measured based on *self-reported availability*. There is evidence that employees' knowledge of their workplace benefits is underestimated,[324] which would introduce measurement error in these variables. If the error is entirely random, this may have attenuated model estimates. On the other hand, it is plausible that employees with children (i.e., those with competing roles) would report the availability of family leave policies more accurately than those without children.[324] If recall was independent of depressive symptoms, then interaction estimates may also be attenuated. Regardless, the buffering effects estimated in this study may be best defined as the effect of the awareness of employee benefits, and future research should utilize a more objective measure of benefits (e.g., Human Resource data) to clarify this potential measurement issue. Second, the NLS does not directly measure the amount of time spent providing childcare among respondents. However, since women still provide the majority of childcare on average,[275–277] the presence of a child in the home would likely be a valid proxy for increased domestic labor responsibilities. Nonetheless, it is important to acknowledge that time spent on childcare has been increasing among fathers.[255,256] The mental health effects among women (and men) who experience more equitable or male-dominated domestic labor arrangements may be positive,[325] or negative,[326] but more research is needed in the face of changing childcare trends. Third, depression scores as measured in this study refer to short-term (2-week) prevalence, which may not reflect true depression status over the 2+ year period between interviews, or capture depression incidence. While the longitudinal design did establish temporality at each interview, a study with earlier age of follow up and ascertainment of incident depression cases would help to further understand the risk of competing roles. Fourth, the CESD is a symptom scale, not diagnostic criteria. Depression is a complex construct to measure and, even though the CESD has been widely used as

a measure of depression for over 40 years,[207] it is likely a measure of more general psychological distress and demoralization.

Despite these limitations, this study is strengthened by the use of a large study of women, covering a wide age range that likely captured the peak period in the lifecourse where employment and raising children were most likely to be in competition. Also, including a wide array of employee benefits allowed for an examination of cumulative buffering effects of multiple benefits on competing roles and CESD scores, and a comparison of the effects with those of non-family-related benefits. From this comparison, I found evidence that both pro-family and non-family-related employee benefits may buffer the depression risk from competing workplace and domestic roles.

In conclusion, patterns of women's participation in the workplace and domestic labor suggest that the gender revolution is incomplete, and competition from dual workplace and domestic labor roles may represent a risk for depression for women. Working women with children inevitably face overlap in the responsibilities that both roles demand, and though workplace benefits may help to attenuate that risk, more fundamental social changes are needed to address the residual gender inequalities in paid and unpaid labor roles in the United States today.

VIII. Conclusion

There is a large literature across disciplines aimed at understanding the causes of the depression gap, defined as differences in depression prevalence between men and women. Women consistently report higher depression risk and more average symptoms than men, which has led many to study genetic or hormonal hypotheses, as reflective of static biological differences between men and women. To date however, no consistent evidence has been presented to support a biological explanation for the depression gap. On the contrary, studies have reported meaningful variation in the magnitude of the depression gap over place and time, patterns which align with hypotheses that the depression gap is a product of the social environment and may covary with the social context of gender. Social stress theory has been used as a framework to potentially understand the social causes of the depression gap.[28,35,36] Social stress theory highlights women's disadvantaged social position relative to men, positioning gender differences in socio-economic opportunities as social stressors,[28] while also acknowledging how gender socialization teaches women to respond to stressors in depressogenic ways from an early age.[29–34]

This dissertation applied social stress theory to better understand the social causes of the depression gap, doing so with three related aims. The first aim was to summarize the evidence for variation or stability in the depression gap in recent decades, through a systematic review and meta-regression of depression gap studies over time and by age. The second aim examined the evidence for a changing depression gap across birth cohorts, and tested the extent to which any changes over time were mediated by changing gender differences in education, employment, and housework rates, three indicators of broader trends in gendered social position through the 21st Century. Finally, acknowledging that broad social changes may be uneven and incomplete, the third aim examined whether women in the workforce with competing domestic labor roles were at increased risk of depression, and whether pro-family workplace benefits buffered the effects of competing roles. The following sections summarize the key findings of the dissertation, synthesize results across aims where applicable, and discuss the strengths and limitations. The final section highlights the overall contributions of this dissertation to understanding the etiology of the depression gap, and

proposes future research directions to mitigate future gender disparities in depression and mental health more broadly.

The systematic review of the depression gap literature identified an overall excess of depression among women compared with men from 1982-2017. This pattern was found in studies that measured the gap both as differences in mean symptom levels and those based on a diagnostic threshold. In this study, the summary depression gap in diagnostic studies was twice as high for women as men, in line with the summary statistic that is often cited when referring to the depression gap in the literature. However, the main hypothesis of this dissertation was that there would be meaningful variation in the magnitude of the depression gap over time, which was not supported by the results of the meta-regression. Overall there was no evidence of change in the depression gap over time, however, patterns varied by age group. Among adults ages 20 and older, the depression gap did not change over the study period. However, the depression gap increased over time among respondents ages 10-19, compared to the reference group of respondents ages 60 and older. This is contrary to the expectation of a decreasing depression gap for girls and women of all ages, given the hypothesis that gendered social position reflects broader changes in norms and the process of gender socialization in the overall population.

The aim 1 analysis also identified variation in the magnitude of the depression gap across depression instruments, particularly among studies measuring differences in depression symptoms. Compared with the Center for Epidemiologic Studies Depression scale (CESD), the depression gap was higher when measured with the 9-item Patient Health Questionnaire (PHQ-9) (SMD=0.16; 95% CI=0.06, 0.25), and 'other' instruments (e.g., the Mental Health Inventory Depression subscale) (SMD=0.14; 95% CI=0.14, 0.22). The diversity of instruments used in depression gap studies underscored the heterogeneity in the definition and operationalization of the depression construct itself. In addition to major depressive disorder, studies estimated the depression gap using measures of major depressive episode,[14] psychological distress,[52] or more specific measures of somatic or internalizing behaviors.[53] Diagnostic depression was measured using DSM-III, DSM-IV, and DSM-5 definitions, each of which represent one snapshot of an evolving conceptualization of the depression construct over time, according to the main authoritative body

on mental illness in the US. The conflation of these tools as interchangeable measures of the depression gap may obscure what makes these related constructs distinct, and may hinder efforts to understand the relationship between gender, stress, and mental health. Testing differences through measurement invariance methods may highlight important differences in the salience, type, or timing of certain aspects of depression for men and women.[327–329] For example, women are more likely to endorse the somatic symptoms of depression than cognitive/affective symptoms,[330–332] therefore scales with more somatic symptoms of depression would likely estimate a different depression gap than scales with more cognitive symptoms. When ignored, these issues represent a potential source of bias across depression gap studies.[333] Accounting for them, however, may offer an opportunity to better understand differences in the etiology and experience of depression for women and men.

Aim 2 reported evidence of variation in the depression gap across four 10-year birth cohorts born from 1955-1994 and followed from 1992-2014. The results suggested that the depression gap had narrowed across birth cohorts, specifically that the gap was smaller among those born from 1985-1994 than those born in prior decades since 1955. Further, evidence of inter-cohort changes in depression symptoms stratified by gender showed that decreases in the gap were primarily due to decreasing depression symptoms among women.

Overall, aims 1 and 2 characterized time trends in different ways, but both offered insight into how the depression gap has changed over time. In aim 1, there was no evidence of change in the depression gap over time, however, patterns varied by age group. The age by time effects suggested that the depression gap is not changing among adults ages 20 and older, but that depression disparities may be increasing among adolescents. The birth cohort effects in aim 2 also corresponded to age (assigned by birth cohort) by time effects. The results of this analysis suggested the depression gap was decreasing among young- and middle-aged adults in subsequent birth cohorts, highlighting broad population trends toward a decreasing depression gap. Taken together, the results of these studies highlight evidence that the depression gap may be expanding and contracting for different age groups. There are at least two potential explanations for these differences. First, age and time differences between the analytical samples may limit the ability to

directly compare results. The age range in the aim 2 sample with the greatest overlap between birth cohorts (i.e., the most complete range to test cohort effects) was approximately ages 20-50. This was more limited than the age range of the aim 1 sample, which included respondents ages 10-99. Therefore, the aim 2 sample could not estimate trends among adolescents or respondents older than age 51. Further, the most recent birth cohort was born from 1985-1994, substantially older than the adolescent age group in the aim 1 sample, which was born from 1998-2007. It is possible that increasing depression gap trends have only emerged very recently, and might eventually be replicated using aim 2 methods, when depression scores are measured in the youngest NLS respondents. Second, the way that age groups were defined to create the aim 1 meta-analytic sample age-related might have introduced non-differential measurement error, biasing estimates of age trends toward the null. Third, the unit of analysis and sample sizes differed between aims. As a meta-regression, the unit of analysis in aim 1 was an independent depression gap estimate, which yielded a sample size of 76 diagnostic-based and 68 symptoms-based estimates. These small sample sizes are common in meta-analytic and meta-regression studies, and may limit the power to identify a statistically meaningful effect, especially when models include additional covariates. In contrast, the unit of analysis in aim 2 was the individual, and included a sample size of 13,666 individuals. The a priori statistical power of this analysis was very high and sufficient to identify the hypothesized trends in the depression gap by birth cohort (see appendix 1). Short of conducting a post-hoc power calculation for aim 1, which can yield biased estimates and misleading interpretations of null findings,[334] it is plausible that the lower statistical power of aim 1 may partially explain the lack of evidence of change over time.

Finally, these divergent findings may simply highlight the limitations of the overall theoretical framework of gendered social position and social stress to explain the depression gap. Other theories may be better suited to reconcile and predict the observed age and time trends. One potential approach might be to focus more explicitly on the accumulation of resources that may be deployed to prevent health problems.[335] This accumulation is often differential across social group identities, and may vary across age and historical time. Applied to the depression gap, adolescent girls may have relatively few resources (e.g., social capital, coping strategies) to prevent depression, while as adults, changes over time have increased

women's resources and their depression risk in turn. Alternatively, the depression gap may reflect social, historical, and biological processes that are too complex to be explained by a single unifying theory. Developing and testing alternative explanations represents an important step to test of the validity of social stress theory in explaining the depression gap and predicting variation by age and time. Additional inquiries to build on the findings of this dissertation are discussed in more detail below.

The increase in the depression gap in adolescents warrants further scrutiny. Available literature has confirmed through multiple data sources that depressive symptoms have been increasing in adolescence, especially among girls.[141,142,336] Hypotheses about why the depression gap emerges in adolescence have focused on social and relational factors, such as psychosocial stressors, trauma, and interpersonal conflict among peers, which become more prevalent in adolescence.[139,140] Girls are more likely to experience harassment and bullying,[337] social isolation as a result of friend conflict.[338] Further, these negative experiences may be exacerbated by depressogenic coping strategies such as rumination and internal versus external attribution of negative experiences, which are more commonly utilized by girls.[143,144] Some scholars have advanced the hypothesis that digital media, including social media, is a new risk factor for mental health problems,[150] particularly among girls.[148,150] The use of digital and social media among adolescents has exponentially increased in the last ten years, thus if it is a cause of mental health problems, particularly among girls, it is a viable hypothesis to explain the change in the depression gap. However, the relationship between social media use and depression is inconsistent.[151,152] A recent review reported positive, negative, and no relationship between time spent using social networking sites and depression.[339] A more general review reported that eight studies found that time using social networks increased the risk of depression, however, twice as many found no effect.[340] It is likely that the relationship between social media use and depression is complicated, and multifaceted. Future research is needed to better understand the type, duration, and nature of social media use among adolescent boys and girls, in order to fully characterize its role in emerging depression gap trends.

In addition to estimating cohort effects in the depression gap, aim 2 tested specific ways through which the depression gap might be changing as mediated through three indicators of social position, which

have been changing over time, differentially by gender. There was evidence that increasing parity in education levels mediated 39% of the overall decreasing trend in the gap, supporting the hypothesis that decreasing relative differences in education levels signal a more equal opportunity structure for women. By contrast, trends in employment and housework differences did not mediate the gap. In these two indicators, the degree of convergence between men and women has been relatively less complete, which may partially explain the lack of an association. Based on census data, aim 2 showed that trends in education have been much more favorable for women than trends in employment. Women have achieved a higher average education level than men for over 20 years, however, women still report lower employment rates than men, and experience numerous additional sources of workplace-related stress and discrimination.[188,252,341] Housework ratios decreased from 4.41 in the 1955-64 cohort to 1.84 in the 1985-94 cohort; this residual disparity in domestic labor may act to reinforce the depression gap. These trends emphasize the importance of power and opportunity differentials by gender as fundamental causes of health,[335] and highlight potentially positive and negative (or null) consequences for the depression gap through social stress pathways.[27]

While aim 2 identified decreasing trends in the depression gap, due in part to increasing parity in social position, it was important to acknowledge the potential negative consequences of shifting workplace and domestic gender roles that may increase women's depression and widen the depression gap.[342,343] Aim 3 examined this hypotheses, framing the incomplete nature of changes in gendered social position, specifically the lack of parity in domestic labor roles, as a depression risk for women. The results showed that women in competing workplace and domestic roles reported an average of 0.56 more depressive symptoms compared with women who were not in both roles (95% CI: 0.15, 0.97). Access to employee benefits buffered the risk of competing roles, such that the effect of competing roles was not associated with greater depression symptoms among women with access to these benefits. However, similar patterns were identified among women with access to pro-family and non-family-related employee benefits, suggesting that the buffering effect on competing roles was not limited to benefits designed to specifically address role competition among working women with children. While these results do suggest that greater

access to these policies would decrease depression among women, a more fundamental and equitable intervention would seek to achieve gender parity in domestic labor roles, and reduce the source of role competition for women. Fortunately, there is evidence showing how this might be achieved, by making pro-family employee benefits less gendered. An intervention in Sweden showed that targeted policies to incentivize new fathers to take family leave appear to benefit maternal postpartum depression risk.[344] However, even when they are available, men's utilization of these types of policies is low,[324,345] emphasizing that significant barriers remain in changing social norms about domestic labor roles for working men and women.

Overall, this dissertation provided evidence of variation to support the hypothesis that the depression gap is not a result of fixed biological mechanisms that differ by gender. Rather it is meaningfully related to the normative social environment, through which gender roles, responsibilities, and opportunities available to women and men are defined and reinforced. Aim 1 found evidence that the current social context might be increasing the depression gap for adolescents, and aim 2 showed that the depression gap has been decreasing in subsequent birth cohorts from 1955 to 1994, driven mostly by women's decreasing depression levels relative to men. These trends were partially explained by increases over time in the average education level of women vs. men. Aim 3 found that, among women in the workplace, those with competing gender roles reported more CESD symptoms than those without competing roles, and that the depression risk from competing roles was buffered by access to both pro-family and non-family-related employee benefits, suggesting that access to employee benefits in general reduced depressive symptoms among women in competing roles.

The conclusions of this dissertation should be considered in light of several limitations. First, depression in Aims 2 and 3 was measured using the CESD scale, which may yield an incomplete picture of depression. Even though the CESD has been widely used as a measure of depression for over 40 years,[207] and is correlated with diagnostic depression measures, it is more likely a measure of general psychological distress and demoralization.[52] Further, even assuming it is an indirect measure of depressive symptoms, CESD items were measured for the past 2-weeks, and did not query incidence, duration, or severity of

symptoms. Given the limited assessment of depressive symptoms, CESD scores may be an imprecise measure of an individual's most valid depression status, as symptoms likely fluctuate greatly over the 2+ year period between study interviews. On the other hand, the CESD exhibits a strong trait component,[346] suggesting that within-individual variation between interviews may be minimal. Also, though the analytical aims did use the longitudinal nature of the data to establish temporality during the study periods, prevalent CESD symptoms were measured and thus studies could not account for prior depression levels. Incidence studies with diagnostic measures would add to the findings of this dissertation with a more complete picture of depression.

Second, cohort analyses are always limited by incomplete data capture on the oldest and youngest age groups across cohorts, and were additionally limited in NLS given the age ranges of participants. An ideal cohort analysis utilizes groups of individuals born in different years but followed for the same period of their lives (e.g., from age 10-50), such that there is complete coverage for individuals at all ages. In aim 2, the age range of interest was defined as early- to middle-adulthood (approximately ages 18-50), chosen to represent the period in life when education, employment, and housework were most salient for individuals and relative gender differences in them were most influential for depression. However, the analysis was limited by incomplete overlap in ages across birth cohorts. Therefore, model estimates were partially based on the interpolation of data for both the youngest ages (15-20) and the oldest ages (40-50). This issue is rarely discussed in cohort analysis studies, but should be considered as a limitation and was examined with sensitivity analyses. Specifically, in aim 2, cohort trends were examined in a subset of the sample with complete overlap in ages, and findings were similar to those found overall. This suggested that the overall findings were robust to these off-support data issues.

One consistent source of tension in completing this dissertation was whether and how to account for intersectionality in the depression gap. The central questions of this dissertation likely have different answers and implications for women based on other intersecting social group identities, such as race/ethnicity, sexual orientation, and socio-economic status. It would be easy enough to acknowledge these differences and include additional stratified analyses to test and present these additional sources of

heterogeneity. However, doing so in a responsible way would have required extensive theoretical framing to inform thoughtful and transparent hypotheses about expected variation across these groups. Not doing so would increase the likelihood that results would be misinterpreted and could potentially reinforce well-worn stereotypes about social causes of health disparities. This extensive framing was beyond the scope of this dissertation and thus no stratified analyses were presented. Future work that thoughtfully integrates intersectionality theory may reveal important differences across diverse social groups, and may yield insight into hypotheses that are supplemental to those examined in this dissertation, with implications for the mental health and social stress literature.

Overall this dissertation leveraged variation over time to examine the etiology of the depression gap. It attempted to do so while avoiding too narrow a scope that is often found in etiological studies based on stress pathways. These studies are often oriented through a socio-medical lens, which frames depression as a consequence of abnormal conditions or exposures.[177] Socio-medical approaches to the depression gap often focus on the effect of stress, arguing that an excess of stress might explain an excess of depression. This approach is often limited by shifting the focus to stress (e.g., stressful life experiences), and away from gender as a fundamental determinant of depression. This dissertation sought to avoid these limitations by maintaining the focus on gender as the exposure, linking the depression gap with the historic subordination of women into positions of less power and opportunity than men.[263] This dissertation showed that the depression gap was partially mediated by the relative social position of men and women, which supports the notion that differences in depression reflect social disparities that would attenuate as socio-economic conditions evolve to become more egalitarian.[58]

This dissertation did employ a socio-medical approach by focusing on solely on depression as an outcome. Studying the effect of stress on a single outcome is often too narrow a scope to understand the relationship between stress and mental health.[27] That is, without considering externalizing disorders that are more common in men than women, such as antisocial behavior and substance abuse, it is not clear whether the depression gap reflects a true mental health disparity by gender, or is simply a difference in the way that stress is expressed.[263] From this perspective, this dissertation represents the initial foundation of a

much broader future inquiry into the consequences of changing gendered social position for the mental health of both women and men. As an alternative to traditional nosology, future work might be informed by transdiagnostic[267] and hierarchical[347] frameworks of psychopathology, which may offer a more comprehensive model to understand the relationship between gender, social roles, and mental health.

The implications of this dissertation for science and public health depend on the lens used to understand the social causes of the depression gap. The gap may be understood through a socio-medical lens, as described above, or through a sociological lens, as one of many consequences of a normative system of social organization. The latter positions the depression gap as a consequence of a social system that functions as intended, conferring advantages to some groups at the cost of disadvantage to others.[177] This dissertation has strengthened my belief that a sociological approach best describes the conditions relevant to the social determinants of health and provides a more constructive lens to understand and predict future trends in the depression gap. For example, a sociological lens may be useful to understand the age and cohort effects identified in this dissertation as the natural process where decreasing depression gap trends, due to increased equity in gendered social position, are displaced in adolescence by newly emerging norms that will potentially reinforce long-established gender disparities in depression in the future.[335] From this perspective, a future research agenda would seek to understand the causes of emerging trends in the adolescent depression gap, to not only reduce the present and future burden of the depression gap, but to understand the fundamental processes through which gendered social position may be reproduced to perpetuate disparities in health and well-being in adolescence and beyond. For example, this approach may inform a related public health trend of increasing gender differences in suicidal behavior among adolescents. Compared with boys, adolescent girls report increasingly higher rates of self-injury,[348] as well as suicidal planning and attempts.[349] They are also more likely to report suicidal ideation as a result of bullying.[350] These co-occurring trends underscore the urgent need to understand the social conditions of adolescence in order to prevent further self-destruction among youth.

IX. Tables and figures

Figure 1.1. Literature search and study selection flowchart

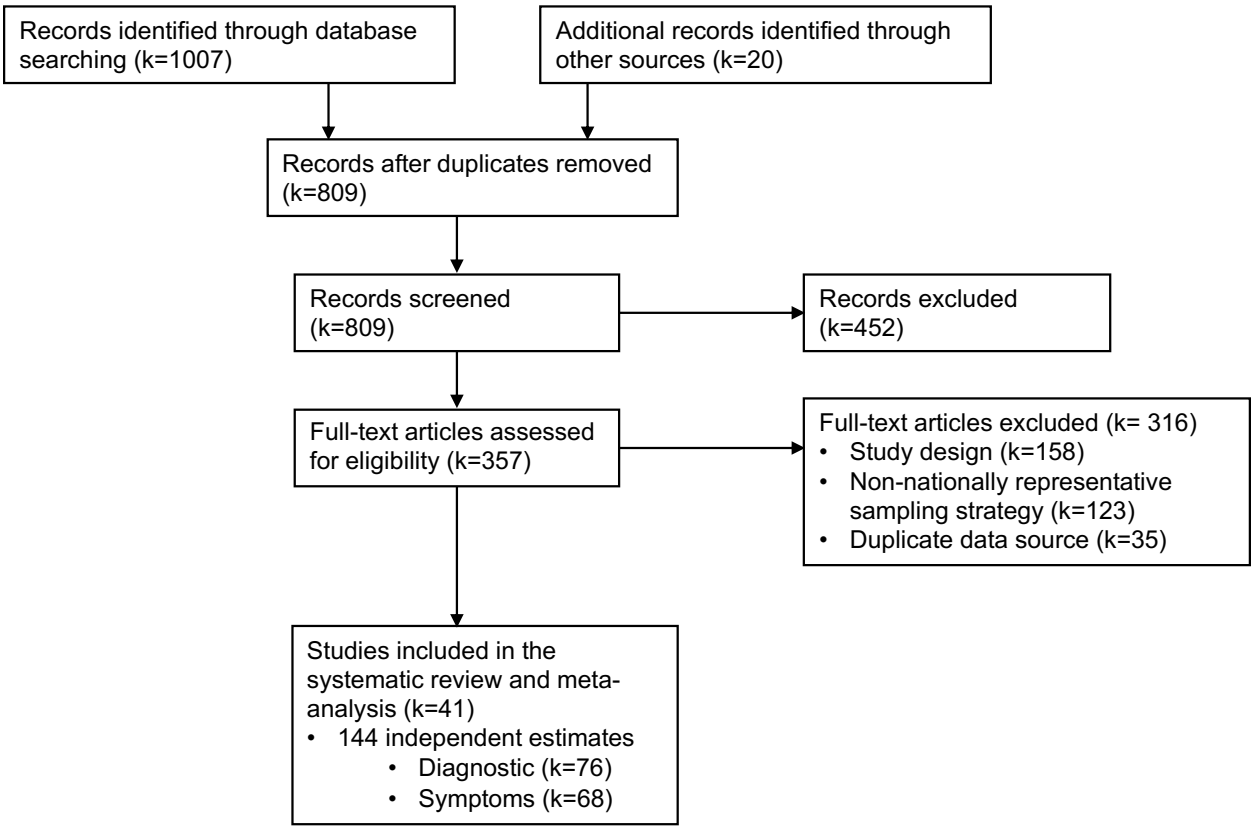


Table 1.1. Studies of gender differences in diagnostic depression, measured as prevalence ratios

Author	BL Year	PR	SE	Age Min	Age Max*	N Men	N Women	Interview & instrument	Study design	Data source	Recall period
Kessler et al, 1993[351]	1990	3.15	1.455	15	24	1010	990	CIDI DSM-III-R	C	National Comorbidity Survey (NCS)	2
	1990	1.50	0.249	25	34	1231	1207	CIDI DSM-III-R	C	NCS	2
	1990	1.89	0.284	35	44	1108	1086	CIDI DSM-III-R	C	NCS	2
	1990	1.52	0.299	45	54	740	726	CIDI DSM-III-R	C	NCS	2
Alaimo et al, 2002[102]	1991	2.28	0.968	15	16	365	389	DIS	C	National Health & Nutrition Epidemiologic Survey (NHANES) III	3
Dawson & Grant, 1997[352]	1992	1.60	0.065	18	99	17819	25043	AUDADIS DSM-IV	C	National Longitudinal Alcohol Epidemiologic Survey (NLAES)	2
Kessler et al, 2010[14]	2001	1.58	0.196	18	34	1375	1658	CIDI DSM-IV	C	National Comorbidity Survey Replication (NCS-R)	2
	2001	2.73	0.483	35	49	1342	1522	CIDI DSM-IV	C	NCS-R	2
	2001	1.48	0.247	50	64	854	1068	CIDI DSM-IV	C	NCS-R	2
	2001	1.71	0.297	65	99	564	894	CIDI DSM-IV	C	NCS-R	2
Center for Behavioral Health Statistics and Quality (CBHSQ), 2004[87]	2004	2.62	0.131	12	17	11363	10938	CIDI DSM-IV	C	National Survey of Drug Use and Health (NSDUH)	2
CBHSQ, 2005[88]	2005	2.96	0.154	12	17	11378	11156	CIDI DSM-IV	C	NSDUH	2
	2005	2.09	0.104	18	25	10697	10444	CIDI DSM-IV	C	NSDUH	2
	2005	1.69	0.096	26	49	7823	9132	CIDI DSM-IV	C	NSDUH	2
	2005	2.00	0.260	50	99	3142	3420	CIDI DSM-IV	C	NSDUH	2
CBHSQ, 2006[89]	2006	2.81	0.151	12	17	11718	11153	CIDI DSM-IV	C	NSDUH	2
	2006	1.81	0.096	18	25	9158	11526	CIDI DSM-IV	C	NSDUH	2
	2006	1.73	0.100	26	49	7431	8606	CIDI DSM-IV	C	NSDUH	2
	2006	1.67	0.217	50	99	2888	3804	CIDI DSM-IV	C	NSDUH	2
CBHSQ, 2007[90]	2007	2.59	0.135	12	17	11524	10909	CIDI DSM-IV	C	NSDUH	2
	2007	1.97	0.100	18	25	10645	11542	CIDI DSM-IV	C	NSDUH	2
	2007	1.72	0.097	26	49	7770	9114	CIDI DSM-IV	C	NSDUH	2
	2007	1.67	0.214	50	99	2857	3509	CIDI DSM-IV	C	NSDUH	2
CBHSQ, 2008[91]	2008	2.91	0.155	12	17	11517	11029	CIDI DSM-IV	C	NSDUH	2
	2008	2.11	0.103	18	25	11166	12039	CIDI DSM-IV	C	NSDUH	2
	2008	1.49	0.091	26	49	7440	8936	CIDI DSM-IV	C	NSDUH	2
	2008	2.14	0.306	50	99	2996	3613	CIDI DSM-IV	C	NSDUH	2
CBHSQ, 2009[92]	2009	2.49	0.129	12	17	11520	11106	CIDI DSM-IV	C	NSDUH	2
	2009	1.93	0.096	18	25	11104	11900	CIDI DSM-IV	C	NSDUH	2
	2009	1.71	0.104	26	49	7591	8729	CIDI DSM-IV	C	NSDUH	2
	2009	1.67	0.213	50	99	3060	3690	CIDI DSM-IV	C	NSDUH	2
CBHSQ, 2010[93]	2010	2.71	0.146	12	17	11140	10820	CIDI DSM-IV	C	NSDUH	2

	2010	2.27	0.091	18	25	17283	16788	CIDI DSM-IV	C	NSDUH	2
CBHSQ, 2011[94]	2011	2.69	0.138	12	17	12028	11482	CIDI DSM-IV	C	NSDUH	2
	2011	1.95	0.077	18	25	17178	17123	CIDI DSM-IV	C	NSDUH	2
Verplaetse et al, 2016[353]	2012	2.01	0.218	18	99	15715	20,386	DSM-V	C	National Epidemiologic Survey of Alcoholism and Related Conditions (NESARC) 3	2
CBHSQ, 2013[95]	2013	2.07	0.405	12	12	1824	1713	CIDI DSM-IV	C	NSDUH	2
	2013	3.42	0.523	13	13	1963	1849	CIDI DSM-IV	C	NSDUH	2
	2013	4.23	0.550	14	14	2026	1865	CIDI DSM-IV	C	NSDUH	2
	2013	3.34	0.366	15	15	1882	1868	CIDI DSM-IV	C	NSDUH	2
	2013	2.54	0.260	16	16	1940	1890	CIDI DSM-IV	C	NSDUH	2
	2013	2.70	0.283	17	17	1914	1760	CIDI DSM-IV	C	NSDUH	2
	2013	1.96	0.095	18	25	10671	11543	CIDI DSM-IV	C	NSDUH	2
	2013	1.71	0.251	26	29	1376	1603	CIDI DSM-IV	C	NSDUH	2
	2013	1.26	0.171	30	34	1529	1802	CIDI DSM-IV	C	NSDUH	2
	2013	1.64	0.269	35	39	1317	1562	CIDI DSM-IV	C	NSDUH	2
	2013	1.57	0.237	40	44	1437	1671	CIDI DSM-IV	C	NSDUH	2
	2013	1.31	0.194	45	49	1440	1613	CIDI DSM-IV	C	NSDUH	2
	2013	1.43	0.282	50	54	837	951	CIDI DSM-IV	C	NSDUH	2
	2013	1.54	0.374	55	59	711	909	CIDI DSM-IV	C	NSDUH	2
	2013	1.41	0.376	60	64	674	719	CIDI DSM-IV	C	NSDUH	2
	2013	3.53	1.399	65	99	1302	1659	CIDI DSM-IV	C	NSDUH	2
CBHSQ, 2017[96]	2017	2.59	0.610	12	12	1329	1269	CIDI DSM-IV	C	NSDUH	2
	2017	4.04	0.696	13	13	1507	1423	CIDI DSM-IV	C	NSDUH	2
	2017	3.63	0.507	14	14	1492	1385	CIDI DSM-IV	C	NSDUH	2
	2017	3.68	0.416	15	15	1460	1427	CIDI DSM-IV	C	NSDUH	2
	2017	2.46	0.246	16	16	1508	1389	CIDI DSM-IV	C	NSDUH	2
	2017	2.20	0.206	17	17	1419	1418	CIDI DSM-IV	C	NSDUH	2
	2017	2.10	0.296	18	18	1070	1036	CIDI DSM-IV	C	NSDUH	2
	2017	1.90	0.263	19	19	976	1002	CIDI DSM-IV	C	NSDUH	2
	2017	1.51	0.180	20	20	973	954	CIDI DSM-IV	C	NSDUH	2
	2017	1.59	0.208	21	21	922	984	CIDI DSM-IV	C	NSDUH	2
	2017	2.21	0.319	22	22	1000	1033	CIDI DSM-IV	C	NSDUH	2
	2017	1.85	0.253	23	23	1006	1155	CIDI DSM-IV	C	NSDUH	2
	2017	1.55	0.205	24	24	975	1139	CIDI DSM-IV	C	NSDUH	2
	2017	1.48	0.206	25	25	1061	1183	CIDI DSM-IV	C	NSDUH	2
	2017	1.83	0.196	26	29	2117	2580	CIDI DSM-IV	C	NSDUH	2
	2017	1.36	0.139	30	34	2631	3088	CIDI DSM-IV	C	NSDUH	2
	2017	2.06	0.267	35	39	2231	2551	CIDI DSM-IV	C	NSDUH	2
	2017	1.91	0.249	40	44	1945	2387	CIDI DSM-IV	C	NSDUH	2
	2017	1.67	0.214	45	49	2075	2450	CIDI DSM-IV	C	NSDUH	2
	2017	1.30	0.274	50	54	901	1093	CIDI DSM-IV	C	NSDUH	2

2017	1.64	0.342	55	59	931	1138	CIDI DSM-IV	C	NSDUH	2
2017	1.87	0.448	60	64	948	1013	CIDI DSM-IV	C	NSDUH	2
2017	1.57	0.343	65	99	2077	2381	CIDI DSM-IV	C	NSDUH	2

Note: BL=baseline; PR=prevalence ratio; SE=standard error; *When the sample age range was described as all ages (e.g., 18 and up)
Study design: C=cross-sectional, BL=Baseline interview of a longitudinal study, L=other wave of longitudinal study (BL year)
Recall period: 2=PY; 3=Lifetime
Race/ Ethnicity: 1=all; 2=NHW; 3=NHB; 4=Hispanic; 5=other

Table 1.2. Studies of gender differences in symptom-based depression, measured as standardized mean differences

Author	BL Year	SMD	SE	Age Min	Age Max*	N Men	N Women	Instrument	Study design	Data source	Retention^	Recall period
Ferketich et al, 2000[100]	1982	0.26	0.02	30	99	2886	5007	CESD	C	National Health & Nutrition Epidemiologic Survey (NHANES) I		1
Everson-Rose et al, 2004[354]	1986	0.31	0.02	24	34	333	407	CESD	BL	American's Changing Lives Survey (ACLS)		1
	1986	0.23	0.05	35	44	228	363	CESD	BL	ACLS		1
	1986	-0.01	0.02	45	54	168	222	CESD	BL	ACLS		1
	1986	0.08	0.05	55	64	251	434	CESD	BL	ACLS		1
	1986	0.12	0.05	65	74	239	526	CESD	BL	ACLS		1
	1986	0.23	0.05	75	99	139	307	CESD	BL	ACLS		1
Inaba et al, 2005[355]	1994	0.27	0.05	28	39	1372	1413	CESD	BL	Natl Survey of Families and Households 2 (NSFH-2)		1
	1994	0.22	0.05	40	49	1013	987	CESD	BL	NSFH-2		1
	1994	0.38	0.05	50	59	594	716	CESD	BL	NSFH-2		1
	1994	0.29	0.04	60	78	856	1220	CESD	BL	NSFH-2		1
Marmorstein et al, 2009[109]	1995	0.12	0.02	12	12	262	329	CESD	BL	Add Health		1
	1995	0.22	0.02	13	13	1039	1218	CESD	BL	Add Health		1
	1995	0.3	0.02	14	14	1319	1472	CESD	BL	Add Health		1
	1995	0.34	0.02	15	15	1778	1883	CESD	BL	Add Health		1
	1995	0.31	0.02	16	16	2061	1991	CESD	BL	Add Health		1
	1995	0.19	0.02	17	17	1981	1940	CESD	BL	Add Health		1
	1995	0.21	0.05	18	18	1512	1427	CESD	BL	Add Health		1
	1995	0.34	0.05	19	19	237	159	CESD	BL	Add Health		1
Neumark-Sztainer et al, 2000[110]	1997	0.26	0.04	10	10	239	267	CDI	C	Commonwealth Fund Survey of Adolescent Girls & Boys		1
	1997	0.02	0.02	11	11	254	305	CDI	C	Commonwealth Fund Survey		1
	1997	0.04	0.05	12	12	386	461	CDI	C	Commonwealth Fund Survey		1
	1997	0.29	0.05	13	13	420	484	CDI	C	Commonwealth Fund Survey		1
	1997	0.22	0.05	14	14	370	462	CDI	C	Commonwealth Fund Survey		1
	1997	0.31	0.02	15	15	361	503	CDI	C	Commonwealth Fund Survey		1
	1997	0.32	0.04	16	16	399	497	CDI	C	Commonwealth Fund Survey		1
	1997	0.25	0.02	17	17	314	372	CDI	C	Commonwealth Fund Survey		1

Mumford et al, 2013[105]	2000	0.48	0.02	15	15	815	765	MHI-D	L (BL=1997)	National Longitudinal Survey of Youth 1997 (NLSY97)	88	1
	2000	0.33	0.02	16	16	819	774	MHI-D	L (BL=1997)	NLSY97	88	1
	2000	0.29	0.05	17	17	811	773	MHI-D	L (BL=1997)	NLSY97	88	1
	2000	0.22	0.02	18	18	766	767	MHI-D	L (BL=1997)	NLSY97	88	1
	2000	0.23	0.05	19	19	657	681	MHI-D	L (BL=1997)	NLSY97	88	1
Song et al, 2011[106]	2005	0.3	0.04	21	64	167	188	CESD	C	---		1
	2005	0.21	0.02	21	64	187	225	CESD	C	---		1
	2005	0.11	0.04	21	64	939	1124	CESD	C	---		1
Shiovitz-Ezra et al, 2009[356]	2005	0.23	0.05	57	64	521	484	CESD	C	National Social Life, Health, and Aging Project (NSHAP)		1
	2005	0.16	0.04	65	74	543	537	CESD	C	NSHAP		1
	2005	0.09	0.05	75	85	373	499	CESD	C	NSHAP		1
Haroz et al, 2014[117]	2006	0.16	0.14	11	12	95	99	CESD-10R	C	Growing up with Media		1
	2006	0.49	0.10	13	14	201	191	CESD-10R	C	Growing up with Media		1
	2006	0.11	0.11	15	17	192	172	CESD-10R	C	Growing up with Media		1
	2009	0.47	0.06	13	14	585	785	CESD-10R	C	Growing up with Media		1
	2009	0.27	0.05	15	17	856	1096	CESD-10R	C	Teen Health and Technology		1
	2009	0.22	0.04	18	18	954	1404	CESD-10R	C	Teen Health and Tech		1
	2006	0.25	0.14	18	18	94	106	CESD-10R	C	Teen Health and Tech		1
Wang et al, 2010[357]	2006	0.29	0.05	11	11	1164	1186	Depressive feelings and behaviors (DFB)	C	Health Behavior in school-aged children (HBSC)		1
	2006	0.44	0.02	12	12	892	951	DFB	C	HBSC		1
	2006	0.49	0.02	13	13	789	997	DFB	C	HBSC		1
	2006	0.51	0.02	14	14	721	742	DFB	C	HBSC		1
	2006	0.59	0.02	15	15	793	804	DFB	C	HBSC		1
Oksuzyan et al, 2010[118]	2006	0.13	0.02	50	54	640	1013	CESD	L (BL=1992)	Health and Retirement Study (HRS)	85*	1
	2006	0.05	0.05	55	59	1051	1472	CESD	L (BL=1992)	HRS	85*	1
	2006	0.1	0.05	60	64	936	1463	CESD	L (BL=1992)	HRS	85*	1
	2006	0.11	0.05	65	69	1537	1879	CESD	L (BL=1992)	HRS	85*	1

	2006	0.12	0.05	70	74	1267	1560	CESD	L (BL=1992)	HRS	85*	1
	2006	0.16	0.02	75	79	906	1128	CESD	L (BL=1992)	HRS	85*	1
	2006	0.11	0.02	80	84	647	917	CESD	L (BL=1992)	HRS	85*	1
	2006	0.04	0.04	85	89	344	649	CESD	L (BL=1992)	HRS	85*	1
	2006	0.04	0.02	90	99	142	379	CESD	L (BL=1992)	HRS	85*	1
Thibodeau et al, 2014[98]	2008	0.29	0.02	18	29	550	500	PHQ-9	C	NHANES 2008		1
	2008	0.34	0.05	30	39	431	447	PHQ-9	C	NHANES 2008		1
	2008	0.3	0.05	40	49	391	452	PHQ-9	C	NHANES 2008		1
	2008	0.23	0.02	50	59	418	400	PHQ-9	C	NHANES 2008		1
	2008	0.29	0.05	60	69	434	459	PHQ-9	C	NHANES 2008		1
	2008	0.25	0.05	70	99	483	482	PHQ-9	C	NHANES 2008		1
Bushman et al, 2012[107]	2011	0.14	0.05	18	90	251	549	CESD	C	-		1
Gettler et al, 2016[99]	2011	0.14	0.04	20	60	1505	933	PHQ-9	C	NHANES 2011-2012		1
Margraf et al, 2016[113]	2013	-0.12	0.14	18	99	1252	1786	DASS-D	BL	Bochum Optimism and mental health		1

Note: BL=baseline; SMD=Standardized mean difference; SE=standard error; *When the sample age range was described as all ages (e.g., 18 and up), upper bound was coded as 99.
^ if sample was not cross-sectional or BL longitudinal; Study design: C=cross-sectional, BL=Baseline interview of a longitudinal study, L=other wave of longitudinal study (BL year)
Race/ Ethnicity: 1=all; 2=NHW; 3=NHB; 4=Hispanic; 5=other; HHANES estimates were stratified by three Hispanic groups: a=among Mexican Americans; b=Puerto Ricans; c=Cuban Americans.

Table 1.3. Distributions of all variables used in meta-regression models

	Diagnostic studies (n=76)	Symptom studies (n=68)
Year Mean (SD)	2010 (6.9)	2001 (7.8)
Year range	1990, 2017	1982, 2013
Age groups	*n (%)*	
All*	3 (3.9)	6 (8.8)
10-19	27 (35.5)	33 (48.5)
20-39	24 (31.6)	4 (5.9)
40-59	11 (14.5)	9 (13.2)
60+	11 (14.5)	16 (23.5)
Symptom period		
Prior-year	74 (97.4)	
Lifetime	2 (2.6)	
Instrument		
DSM-III/III-R	4 (5.3)	
DSM-IV/IV-R	71 (93.4)	
DSM-5	1 (1.3)	
CESD		42 (61.7)
CDI		8 (11.8)
PHQ-9		7 (10.3)
Other		11 (16.2)

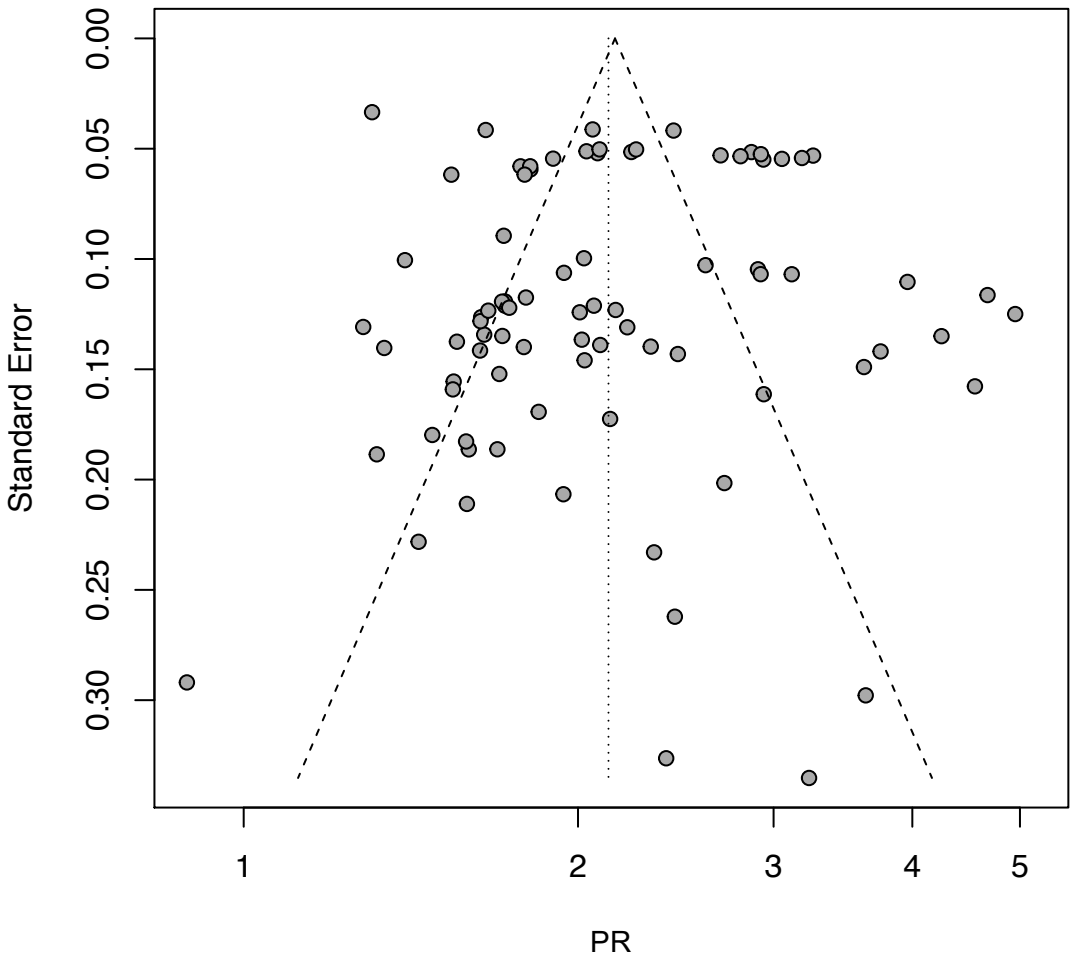
Note: SD=Standard Deviation; DSM=Diagnostic and Statistical Manual; CESD=Center for Epidemiologic Studies Depression scale; CDI=Children's Depression Inventory; PHQ-9=Patient Health Questionnaire; * studies were not included to estimate age effects

Table 1.4. Meta-regression model estimates

	Diagnostic depression gap PR (95% CI)		Symptom-based depression gap SMD (95% CI)	
	Model 1a	Model 2a**	Model 1b	Model 2b*
Intercept	2.35 (1.51, 3.68)	2.22 (1.78, 2.66)	0.27 (0.17, 0.38)	0.26 (0.09, 0.44)
Study Year	0.97 (0.89, 1.06)	0.84 (0.68, 1.04)	-0.03 (-0.05, 0)	-0.03 (-0.06, 0.01)
Age (ref=60+)				
10-19	1.44 (1.19, 1.74)	0.42 (0.13, 1.42)	0.08 (0, 0.15)	0 (-0.26, 0.26)
20-39	0.92 (0.76, 1.11)	0.49 (0.14, 1.71)	-0.01 (-0.1, 0.09)	0.12 (-0.12, 0.36)
40-59	0.85 (0.66, 1.08)	0.66 (0.18, 2.46)	-0.04 (-0.14, 0.05)	-0.07 (-0.3, 0.15)
Age 10-19 x study year		1.27 (1.0, 1.61)		0.05 (0.01, 0.09)
Age 20-39 x study year		1.14 (0.89, 1.45)		-0.03 (-0.09, 0.03)
Age 40-59 x study year		1.06 (0.82, 1.36)		0.02 (-0.04, 0.08)
Diagnostic-depression instrument (ref=DSM-IV)				
DSM-III/IIIR	0.88 (0.54, 1.44)	0.71 (0.4, 1.25)		
Other instrument	0.74 (0.31, 1.76)	1.01 (0.41, 2.44)		
Symptom-based depression instrument (ref=CESD)				
CDI			-0.07 (-0.17, 0.02)	-0.06 (-0.17, 0.04)
PHQ-9			0.13 (0.02, 0.23)	0.15 (0.04, 0.27)
Other instrument			0.15 (0.06, 0.23)	0.13 (0.04, 0.22)

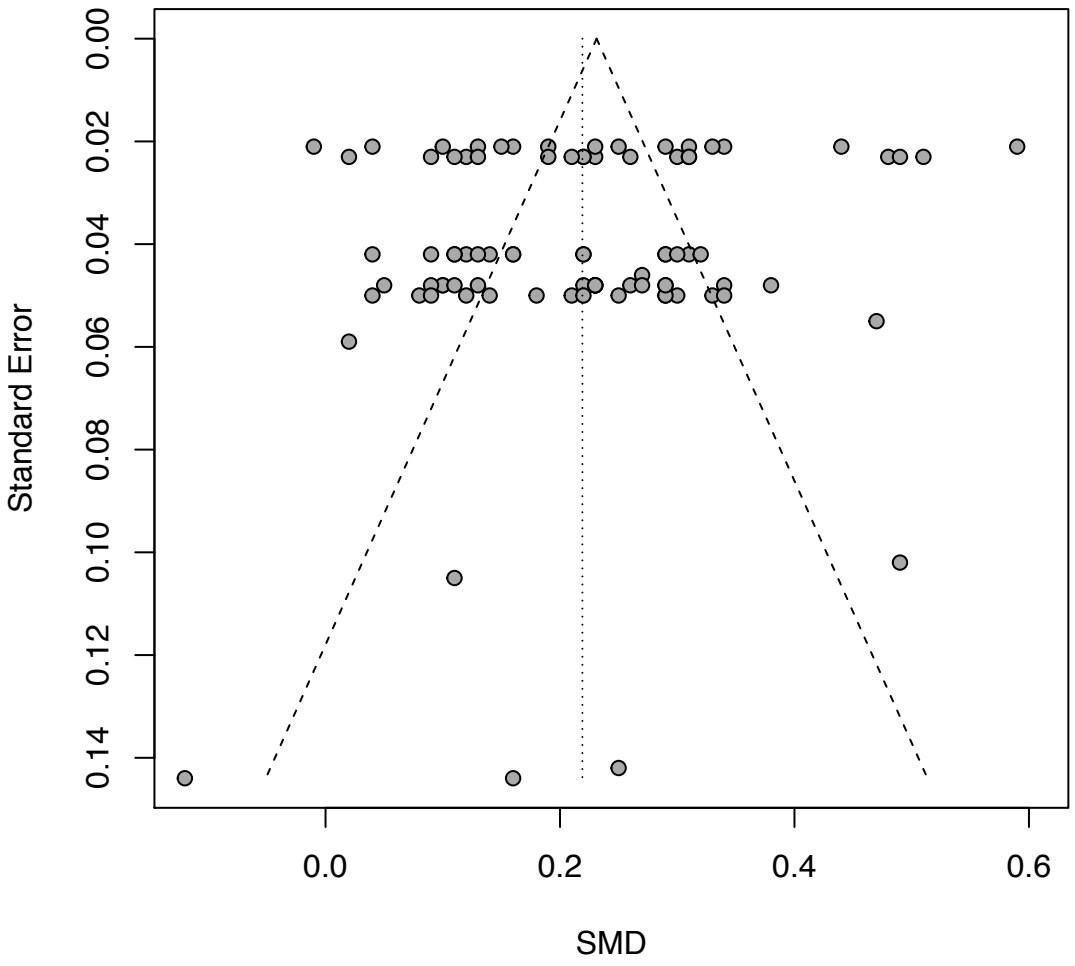
Note: * adjusted for all model 1 variables; PR=prevalence ratio; SMD=standardized mean difference; CI=confidence interval; DSM=Diagnostic and Statistical Manual; CESD=Center for Epidemiologic Studies Depression scale; CDI=Children's Depression Inventory; PHQ-9=Patient Health Questionnaire

Figure 1.2. Funnel plot of studies measuring the diagnostic depression gap



Note: PR = Prevalence ratio. Egger's Test intercept=-0.266 (95% CI=-1.78, 1.24). The trim-and-fill procedure imputed no studies to balance the funnel plot

Figure 1.3. Funnel plot of studies measuring the symptom-based depression gap



Note: SMD=standardized mean difference. Egger's Test intercept=-1.19 (95% CI=-3.5, 1.1) The trim-and-fill procedure imputed 23 studies to balance the funnel plot. Balanced random effects SMD=0.27 (95% CI=0.24; 0.30). Prediction interval=-0.026, 0.57; I^2 = 0.960 (95% CI=0.955, 0.964)

Figure 2.1. Conceptual diagram of the relationship between gender and depression predicted by social stress theory

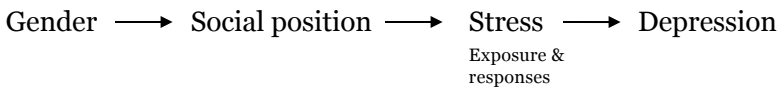


Table 2.1. The source population, birth year range, interview waves, sample sizes, gender distributions, and mean attrition rates per wave for the National Longitudinal Survey samples that comprised the study sample.

Sample name Source population	Birthyear range	Year range of study interviews	N at first interview (% women)	N at last interview	Attrition per wave
NLSY79					
US rep. adults in 1979	1957-1965	1992-2014	12686 (49.5%)	7231	4.8%
NLS YA					
Children of NLSY79 mothers age 15+	1972-1999	1994-2014	980 (52%)	5735	*

Note: rep=representative, BL=baseline; *The sample increases over time because participants are recruited as they turn age 14. See supplemental table 2.1 for sample sizes of the NLS YA by age at each interview wave.

Table 2.2. Person-years of observation for each birth cohort overall and stratified by gender in the study sample.

N (%)	**Total**	**Men**	**Women**
Total	47,646 (100)	24,221 (50.84)	23,425 (49.16)
Birth Cohort			
1955-1964	22,575 (47.4)	11,561 (51.21)	11,014 (48.79)
1965-1974	1444 (3)	754 (52.22)	690 (47.78)
1975-1984	12,935 (27.2)	6456 (49.91)	6479 (50.09)
1985-1994	10,692 (22.4)	5450 (50.97)	5242 (49.03)

Note: no respondents were born in 1955-56 and 1966-71

Table 2.3. Descriptive Statistics of analytic variables averaged across cohorts, overall and stratified by gender

	Total		**Men**	**Women**
	Mean (SD)	*Range*	*Mean (SD)*	*Mean (SD)*
Birth year	1972 (12.2)	[1957, 94]	1972 (12.19)	1973 (12.17)
CESD score	4.09 (4.0)	[0, 21]	3.63 (3.69)	4.57 (4.24)
High CESD score	0.17 (0.37)	[0, 1]	0.13 (0.34)	0.2 (0.4)
% College degree or more †	0.28 (0.05)	[0.08, 0.42]	0.27 (0.03)	0.28 (0.08)
College ratio*	0.98 (0.21)	[1.31, 0.76]		
% employed†	0.63 (0.04)	[0.32, 0.86]	0.72 (0.03)	0.54 (0.04)
Employed ratio*	0.93 (0.4)	[1.77, 0.57]		
Housework hrs/wk ‡	14.1 (7.56)	[4.4, 34.5]	8.1 (1.92)	20.3 (6.11)
Housework ratio*	2.94 (2.62)	[1.8, 3.6]		
Paid work hrs/week**	31.2 (10.3)	[17.1, 47.8]	40.8 (4.07)	21.4 (2.62)
Unemployment rate	0.06 (0.02)	[0.04, 0.10]	0.06 (0.17)	0.05 (0.14)
Marriage rate	0.58 (0.05)	[0.52, 0.68]	0.58 (0.04)	0.55 (0.04)

*Ratio comparing women to men; **Among employed persons; † source: Current Population Survey historical data; ‡ source: 1965-66 Americans' Use of Time Study; 1975-76 Time Use in Economic and Social Accounts; 1985 Americans' Use of Time; 1995 Electric Power Research Institute (EPRI) Study; 2003-2008 American Time Use Study

Table 2.4. Hierarchical mixed model results estimating CESD score differences for birth cohort, gender, and their interaction

	Model 1*	Model 2*	Model 3*
		CESD score difference (95% CI)	
Fixed Effects			
Intercept	3.76 (3.36, 4.17)	3.48 (3.06, 3.91)	3.54 (3.13, 3.96)
Cohort (ref=1955)	-0.15 (-0.2, -0.09)	-0.09 (-0.14, -0.05)	-0.10 (-0.14, -0.05)
Gender (ref=men)	0.96 (0.86, 1.06)	1.10 (0.98, 1.24)	1.14 (1.0, 1.29)
Cohort x gender (ref=1955)†			
1965-74		-0.25 (-0.76, 0.26)	
1975-84		-0.36 (-0.63, -0.1)	
1985-94		-0.54 (-0.79, -0.28)	
Cohort x gender (ref=1955)			-0.18 (-0.26, -0.1)
Random effects			
Intercept	5.48	5.43	5.46
Residual	10.3	10.3	10.3
Fit statistics			
BIC	236575	236598	236568
Log Likelihood	-118239	-118230	-118231

Note: all continuous variables are mean centered; *adjusted for age polynomials and cohort mean age; †modeled as indicator variables

Table 2.5. Hierarchical mixed model results estimating CESD score differences for birth cohort, stratified by gender

	Women*	Men*
	CESD score difference (95% CI)	
Fixed effects		
Intercept	5.03 (4.74, 5.32)	3.91 (3.67, 4.15)
Cohort (ref=1955)	-0.41 (-0.63, -0.19)	-0.05 (-0.08, -0.02)
Random-effects variance		
Individual intercept	6.68	4.3
Residual	11.42	9.22
Fit statistics		
BIC	130756	129079
Log likelihood	-65343	-64504

Note: all continuous variables are mean centered; *adjusted for age polynomials and cohort mean age

Figure 2.2. Predicted CESD scores by age, stratified by gender and cohort

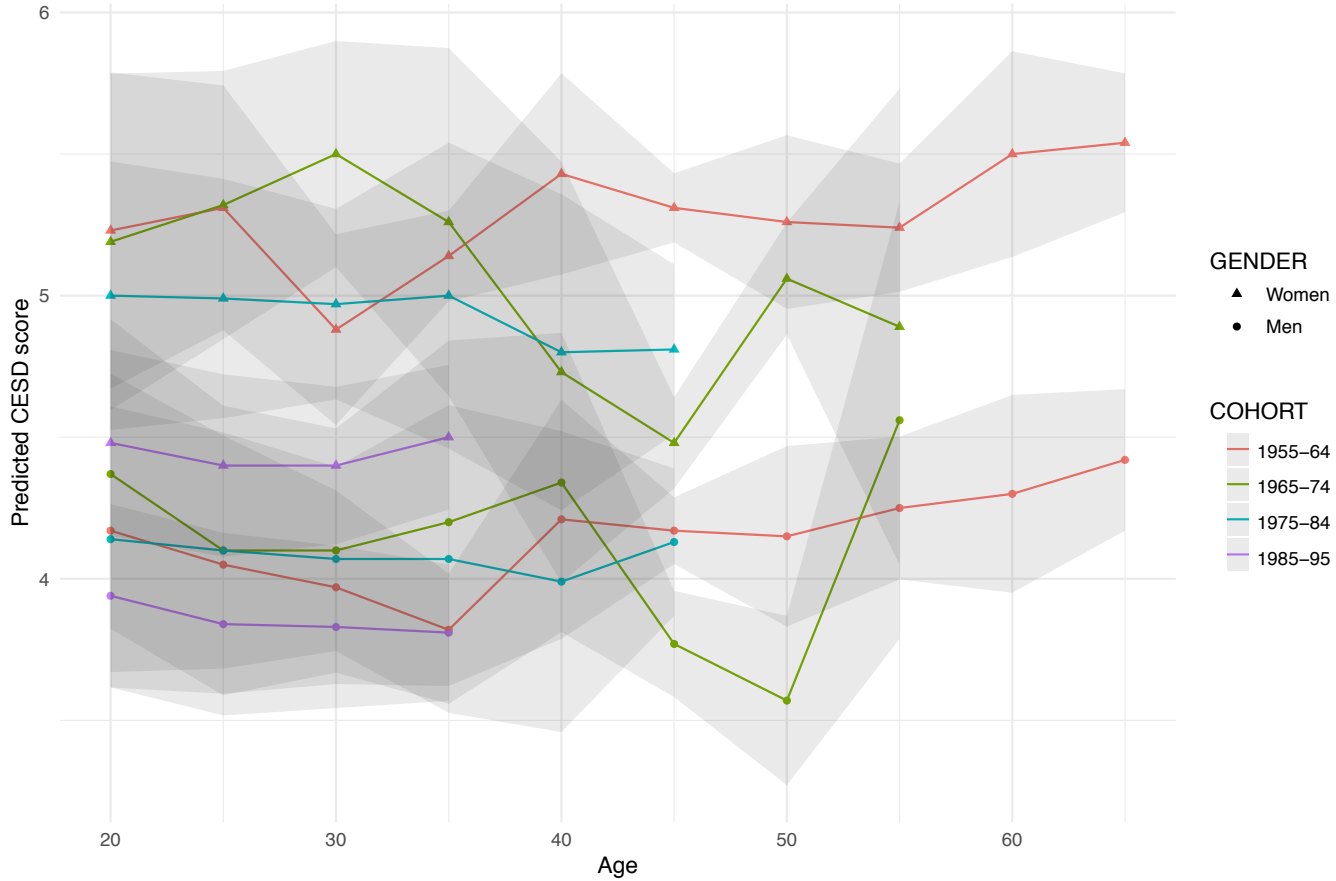
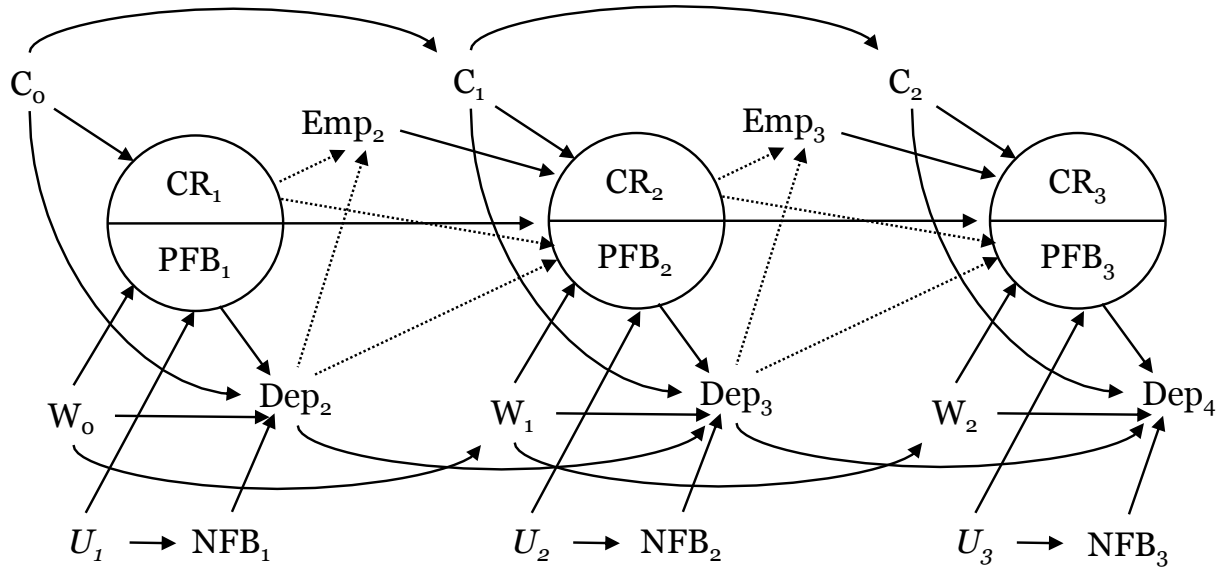


Table 2.6. Hierarchical mixed model results estimating the mediation of the gender gap in CESD scores across cohorts by indicators of gendered social position

	Model 1A	Model 2A	Model 3A
Fixed Effects parameters	*CESD score difference (95% CI)*		
Intercept	3.63 (3.26, 3.99)	5.93 (3.25, 8.61)	6.92 (4.94, 8.9)
Cohort (ref=1955)	-0.49 (-0.66, -0.32)	-0.47 (-0.75, -0.19)	-0.47 (-0.77, -0.17)
Gender	1.11 (0.98, 1.24)	1.11 (0.97, 1.24)	1.02 (0.82, 1.23)
Gender x cohort	-0.11 (-0.18, -0.04)	-0.26 (-0.34, -0.18)	-0.20 (-0.31, -0.09)
College ratio*	-0.13 (-0.21, -0.05)		
Employed ratio*†		0.42 (-0.60, 1.44)	
Housework ratio*			-0.03 (-0.15, 0.10)
Mediation analysis			
Avg. mediated effect	-0.07 (-0.1, -0.04)	0.08 (-0.04, 0.22)	0.02 (-0.13, 0.17)
% mediated	0.39 (0.18, 0.78)	-0.45 (-1.54, 0.26)	-0.108 (-0.26, 0.04)
Random-effects variance			
Individual intercept	4.95	5.37	5.39
Residual	9.82	10.18	10.22
Fit statistics			
BIC	138036	234429	232961
Log likelihood	-68952	-117145	-116406

Note: all models adjusted for age polynomials and cohort mean age; *The ratio of base rates comparing women to men; †Adjusted for number of hours worked per week

Figure 3.1. Directed acyclic graph representing the hypothesized structure of confounding and selection bias



Note: CR = competing roles; PFB = pro-family benefits; NFB = non-family-related benefits; Emp = employment status; Dep = CESD scores; C = age; education; income; work hours; marital status; race/ethnicity; W = age, work hours; # of children; children<5; income; education; employer type; industry; U = unmeasured variable representing an alternative pathway between PFB and CESD scores; solid lines represent sources of potential confounding; dashed lines represent sources of potential selection bias

Table 3.1. Study variable means and percentages for all available person-time from 2006-2014, overall and stratified by competing roles status (n=12,239)

Mean (SD)	**Total**	**Competing roles**	**No competing roles**	**P-value**
Number of children living in household	1.88 (1.31)	---	---	---
Age	47.4 (6.95)	47.1 (7.38)	48.9 (3.51)	<.0001
Highest grade completed	13.62 (2.66)	13.4 (2.6)	14.83 (2.67)	<.0001
Income	40715 (35904)	38845 (33691)	50559 (44562)	<.0001
Hours of paid work per week	42.8 (7.95)	42.5 (7.85)	44.2 (8.33)	<.0001
High CESD symptoms	0.16 (0.37)	0.17 (0.38)	0.11 (0.31)	0.0015
CESD score	3.90 (4.1)	3.99 (4.14)	3.26 (3.74)	0.001
Percentage				
Competing roles	61.0	---	---	---
Any pro-family benefits	83.6	85.8	83.2	0.0047
Any non-family-related benefits	86.0	85.4	89.1	<.0001
Both benefit types available	80.9	80.4	83.8	0.0004
Marital Status				
Never married	17.2	13.4	37.3	<.0001
Currently married	50.4	53.0	36.5	
Formerly married	32.5	33.6	26.2	
Race/ethnicity				
Hispanic	19.5	20.5	14.3	<.0001
NH Black	33.1	33.6	30.6	
Other	47.4	46.0	55.1	
Employer type				
Government	27.0	27.1	27.2	<.0001
Private sector	57.8	60.4	60.9	
Non-profit	15.3	12.5	11.9	

Note: CESD=Center for Epidemiologic Studies Depression scale

Figure 3.2. Prevalence of competing roles (employed with children living in the home) from 2006-2014

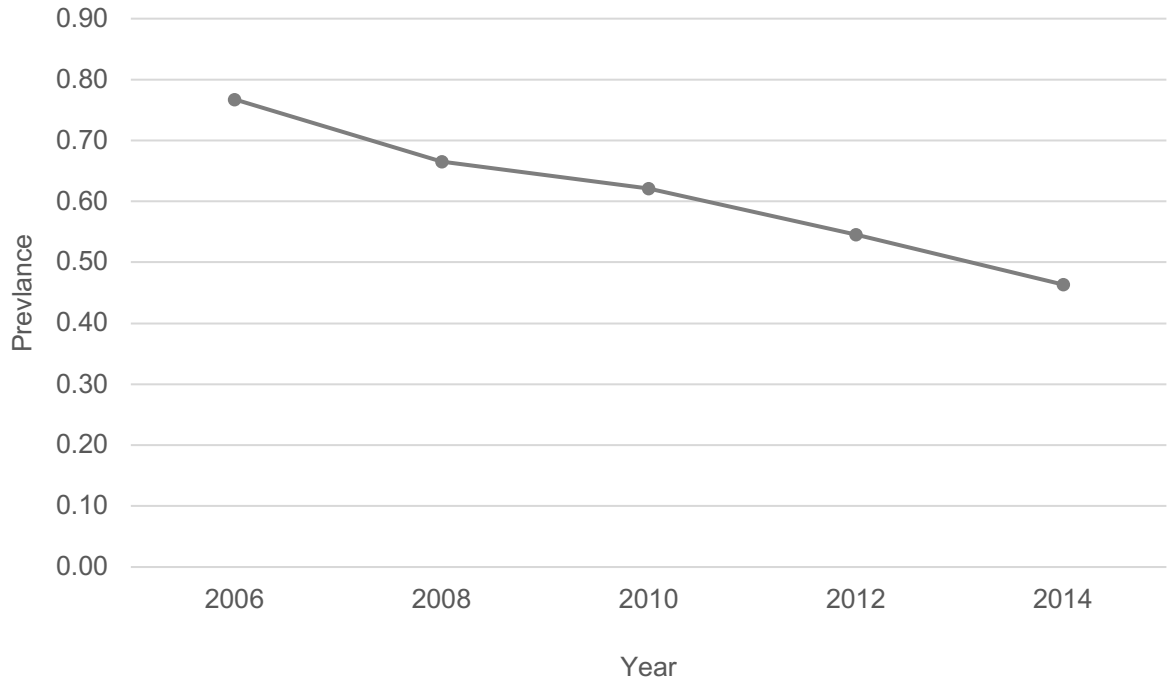


Table 3.2. CESD symptom score differences and the risk of high-CESD symptoms among time spent in competing roles vs. not in competing roles and women with any vs. no pro-family employee benefits, between 2006-2014

	CESD score difference (95% CI)		**RR of high-CESD symptoms (95% CI)**	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Competing roles	0.72 (0.32, 1.12)	0.56 (0.15, 0.97)[a]	1.59 (1.18, 2.14)	1.62 (1.17, 2.25)[a]
Any pro-family employee benefits (ref=none)	-0.70 (-1.07, -0.33)	-0.83 (-1.36, -0.31)[b]	0.69 (0.59, 0.81)	0.72 (0.57, 0.91)[b]

Note: Competing roles are defined as working with children living in the respondent's household (ref=working with no children living in the household); CESD=Center for Epidemiologic Studies Depression scale; CI=Confidence Interval
[a] adjusted for year, age, race/ethnicity, hours of paid work per week, employer type, industry, education
[b] adjusted for year, industry, employer type, and hours of paid work per week

Table 3.3. CESD symptom score differences among women in competing roles vs. not in competing roles, stratified by the availability (any vs. none) and a count of pro-family employee benefits, between 2006-2014

Availability of pro-family benefits	**CESD score difference (95% CI)*[a]**	**RR of high-CESD symptoms (95% CI)*[b]**
No benefits	6.1 (1.14, 11.1)	2.62 (1.15, 5.97)
Any benefits (ref=none)	0.44 (-0.21, 1.0)	1.38 (0.87, 2.19)
One benefit (ref=none)	1.1 (-0.08, 2.26)	1.67 (0.88, 3.17)
Two or more benefits (ref=none)	-0.01 (-0.87, 0.86)	1.04 (0.6, 1.81)

*adjusted for age, race/ethnicity, hours of paid work per week, employer type, industry, education, number of children, children under 5, non-family-related benefits

Note: Competing roles are defined as working with children living in the respondent's household (ref=working with no children living in the household); CESD=Center for Epidemiologic Studies Depression scale; CI=Confidence Interval. Interaction test: H0: competing roles x benefits=0;

[a] Interaction B=-0.51, p=0.017 without adjustment for non-family-related benefits; B=-0.44, p=0.023 with adjustment for non-family-related benefits

[b] RERI= -0.81 (95% CI= -2.18, 0.56) without adjustment for non-family-related benefits; RERI= -1.05 (-2.8, 0.89) with adjustment for non-family-related benefits

Table 3.4. The risk of high-CESD symptoms among women in competing roles vs. not in competing roles, stratified by the availability of any (any vs. none) and a count of non-family-related employee benefits, between 2006-2014

Availability of non-family-related benefits	**CESD score difference (95% CI)***	**RR of high-CESD symptoms (95% CI)***
No benefits	3.59 (1.24, 5.95)	3.15 (1.45, 6.83)
Any benefits (ref=none)	0.57 (-0.61, 1.74)	1.38 (0.87, 2.19)
One benefit (ref=none)	2.09 (-0.26, 4.44)	3.11 (0.87, 11.11)
Two or more benefits (ref=none)	0.44 (-0.73, 1.62)	1.17 (0.81, 1.69)

*adjusted for age, race/ethnicity, hours of paid work per week, employer type, industry, education, number of children, children under 5, pro-family benefits

Note: Competing roles are defined as working with children living in the respondent's household (ref=working with no children living in the household); CESD=Center for Epidemiologic Studies Depression scale; CI=Confidence Interval.

Table 3.5. The risk of being unemployed and reporting available pro-family benefits, based on prior CESD symptom score, high CESD symptoms, and competing role status

RR (95% CI)	Unemployed (ref= employed)	Any pro-family benefits (ref= no pro-family benefits)
Competing roles		
2004	1.038 (1.028, 1.047)	0.96 (0.932, 0.989)
2006	0.997 (0.98, 1.014)	0.958 (0.927, 0.989)
2008	1.049 (1.033, 1.065)	0.997 (0.96, 1.035)
2010	1.021 (1.008, 1.033)	0.997 (0.961, 1.034)
2012	1.051 (1.034, 1.068)	1.025 (0.983, 1.068)
CESD score		
2004	0.991 (0.988, 0.993)	0.996 (0.993, 1.0)
2006	1.007 (1.005, 1.009)	0.997 (0.993, 1.001)
2008	0.997 (0.994, 1.0)	0.998 (0.995, 1.002)
2010	0.998 (0.996, 1.001)	0.996 (0.992, 1.0)
2012	0.998 (0.995, 1.001)	0.997 (0.992, 1.002)
High CESD-symptoms		
2004	0.938 (0.916, 0.96)	0.996 (0.993, 1)
2006	1.058 (1.034, 1.081)	0.997 (0.993, 1.001)
2008	1.011 (0.983, 1.038)	0.998 (0.995, 1.002)
2010	1.009 (0.984, 1.034)	0.996 (0.992, 1)
2012	1.021 (0.991, 1.051)	0.997 (0.992, 1.002)

Note: CESD=Center for Epidemiologic Studies Depression scale RR=risk ratio, CI=Confidence Interval

X. References

1. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. (American Psychiatric Pub, 2013).
2. Weissman, M. M. & Klerman, G. L. Sex differences and the epidemiology of depression. *Archives of general psychiatry* **34**, 98–111 (1977).
3. Young, M. A., Fogg, L. F., Scheftner, W. A., Keller, M. B. & Fawcett, J. A. Sex differences in the lifetime prevalence of depression: does varying the diagnostic criteria reduce the female/male ratio? *Journal of Affective Disorders* **18**, 187–192 (1990).
4. Silverstein, B. Gender difference in the prevalence of clinical depression: the role played by depression associated with somatic symptoms. *American Journal of Psychiatry* **156**, 480–482 (1999).
5. Fennig, S., Schwartz, J. E. & Bromet, E. J. Are diagnostic criteria, time of episode and occupational impairment important determinants of the female: male ratio for major depression? *Journal of affective disorders* **30**, 147–154 (1994).
6. Tousignant, M., Brosseau, R. & Tremblay, L. Sex biases in mental health scales: do women tend to report less serious symptoms and confide more than men? *Psychological Medicine* **17**, 203–215 (1987).
7. Coryell, W., Endicott, J. & Keller, M. Major depression in a nonclinical sample: demographic and clinical risk factors for first onset. *Archives of general psychiatry* **49**, 117–125 (1992).
8. Griffiths, K. M., Christensen, H. & Jorm, A. F. Predictors of depression stigma. *BMC psychiatry* **8**, 1 (2008).
9. Kendler, K. S., Davis, C. G. & Kessler, R. C. The familial aggregation of common psychiatric and substance use disorders in the National Comorbidity Survey: a family history study. *The British Journal of Psychiatry* **170**, 541–548 (1997).
10. Gove, W. R. & Geerken, M. R. Response bias in surveys of mental health: An empirical investigation. *American Journal of Sociology* 1289–1317 (1977).

11. Stommel, M. *et al.* Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). *Psychiatry research* **49**, 239–250 (1993).

12. Steer, R. A., Beck, A. T. & Brown, G. Sex differences on the revised Beck Depression Inventory for outpatients with affective disorders. *Journal of Personality Assessment* **53**, 693–702 (1989).

13. Kessler, R. C. Epidemiology of women and depression. *Journal of affective disorders* **74**, 5–13 (2003).

14. Kessler, R. C. *et al.* Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: results from the WHO World Mental Health Survey Initiative. *Depression and anxiety* **27**, 351–364 (2010).

15. Angold, A. & Costello, E. J. Puberty and depression. *Child and Adolescent Psychiatric Clinics* **15**, 919–937 (2006).

16. Brooks-Gunn, J., Graber, J. A. & Paikoff, R. L. Studying links between hormones and negative affect: Models and measures. *Journal of Research on Adolescence* **4**, 469–486 (1994).

17. Piccinelli, M. & Wilkinson, G. Gender differences in depression. *The British Journal of Psychiatry* **177**, 486–492 (2000).

18. Cai, N. *et al.* Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–588 (2015).

19. Kohli, M. A. *et al.* The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron* **70**, 252–265 (2011).

20. Thomson, P. *et al.* Sex-specific association between bipolar affective disorder in women and GPR50, an X-linked orphan G protein-coupled receptor. *Molecular psychiatry* **10**, 470 (2005).

21. Schulze, T. G. *et al.* Association between a functional polymorphism in the monoamine oxidase A gene promoter and major depressive disorder. *American journal of medical genetics* **96**, 801–803 (2000).

22. Stine, O. C. *et al.* Initial genome screen for bipolar disorder in the NIMH genetics initiative pedigrees: chromosomes 2, 11, 13, 14, and X. *American journal of medical genetics* **74**, 263–269 (1997).

23. Krishnan, V. & Nestler, E. J. Linking molecules to mood: new insight into the biology of depression. *American Journal of Psychiatry* **167**, 1305–1320 (2010).

24. Faraone, S. V., Lyons, M. J., Tsuang, M. T. & Rao, D. Sex differences in affective disorder: genetic transmission. *Genetic epidemiology* **4**, 331–343 (1987).

25. Kendler, K. S. & Prescott, C. A. A population-based twin study of lifetime major depression in men and women. *Archives of general psychiatry* **56**, 39–44 (1999).

26. Lynch, J. Socioeconomic position. in *Social Epidemiology* (2000).

27. Aneshensel, C. S. Social stress: Theory and research. *Annual review of sociology* **18**, 15–38 (1992).

28. Aneshensel, C. S., Phelan, J. C. & Bierman, A. The sociology of mental health: Surveying the field. in *Handbook of the sociology of mental health* 1–19 (Springer, 2013).

29. Kunz-Ebrecht, S. R., Kirschbaum, C. & Steptoe, A. Work stress, socioeconomic status and neuroendocrine activation over the working day. *Social science & medicine* **58**, 1523–1530 (2004).

30. Brydon, L., Edwards, S., Mohamed-Ali, V. & Steptoe, A. Socioeconomic status and stress-induced increases in interleukin-6. *Brain, Behavior, and Immunity* **18**, 281–290 (2004).

31. Hackman, D. A. & Farah, M. J. Socioeconomic status and the developing brain. *Trends in cognitive sciences* **13**, 65–73 (2009).

32. Persson, R. *et al.* Physiological and psychological reactions to work in men and women with identical job tasks. *European journal of applied physiology* **105**, 595 (2009).

33. Hernandez, D. C. & Pressler, E. Gender disparities among the association between cumulative family-level stress & adolescent weight status. *Preventive medicine* **73**, 60–66 (2015).

34. Risman, B. J. & Davis, G. From sex roles to gender structure. *Current Sociology* **61**, 733–755 (2013).

35. Barnett, R. C. & Hyde, J. S. Women, men, work, and family: An expansionist theory. *American psychologist* **56**, 781 (2001).

36. Aneshensel, C. S. Marital and employment role-strain, social support, and depression among adult women. *Stress, social support, and women* 99–114 (1986).

37. Karg, R. S. *et al.* Past year mental disorders among adults in the United States: Results from the 2008–2012 Mental Health Surveillance Study. *CBHSQ Data Review* (2014).

38. Mojtabai, R., Olfson, M. & Han, B. National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics* e20161878 (2016).

39. Kessler, R. C. & McRae, J. A. Trends in the Relationship Between Sex and Psychological Distress: 1957- 1976. *American Sociological Review* **46**, 443–452 (1981).

40. Kessler, R. C. *et al.* Sex and depression in the National Comorbidity Survey. II: Cohort effects. *Journal of affective disorders* **30**, 15–26 (1994).

41. Connell, R. W. *Gender and power: Society, the person and sexual politics*. (John Wiley & Sons, 2014).

42. Risman, B. J. Gender as a social structure theory wrestling with activism. *Gender & society* **18**, 429–450 (2004).

43. England, P. & Li, S. Desegregation stalled: The changing gender composition of college majors, 1971-2002. *Gender & Society* **20**, 657–677 (2006).

44. England, P. *et al.* Why are some academic fields tipping toward female? The sex composition of US fields of doctoral degree receipt, 1971–2002. *Sociology of education* **80**, 23–42 (2007).

45. Cotter, D. A., Hermsen, J. M. & Vanneman, R. Gender inequality at work. (2004).

46. Martin, J. A., Hamilton, B. E., Osterman, M. J., Curtin, S. C. & Mathews, T. Births: final data for 2012. (2013).

47. Schmidt, S., Roesler, U., Kusserow, T. & Rau, R. Uncertainty in the workplace: Examining role ambiguity and role conflict, and their link to depression—a meta-analysis. *European Journal of Work and Organizational Psychology* **23**, 91–106 (2014).

48. Stich, J. *Employee Benefits Survey*. www.ifebp.org/benefitsurvey2016 (2018).

49. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. *Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health*. https://www.samhsa.gov/data/ (2017).

50. Kessler, R. C., McGonagle, K. A., Swartz, M., Blazer, D. G. & Nelson, C. B. Sex and depression in the National Comorbidity Survey I: Lifetime prevalence, chronicity and recurrence. *Journal of affective disorders* **29**, 85–96 (1993).

51. Young, M. A., Fogg, L. F., Scheftner, W. A., Keller, M. B. & Fawcett, J. A. Sex differences in the lifetime prevalence of depression: does varying the diagnostic criteria reduce the female/male ratio? *Journal of Affective Disorders* **18**, 187–192 (1990).

52. Ridner, S. H. Psychological distress: concept analysis. *Journal of advanced nursing* **45**, 536–545 (2004).

53. Angst, J. & Dobler-Mikola, A. Do the diagnostic criteria determine the sex ratio in depression? *Journal of Affective Disorders* **7**, 189–198 (1984).

54. Salk, R. H., Hyde, J. S. & Abramson, L. Y. Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. *Psychol Bull* **143**, 783–822 (2017).

55. Ripke, S. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry* **18**, 497–511 (2013).

56. Angold, A. & Worthman, C. W. Puberty onset of gender differences in rates of depression: a developmental, epidemiologic and neuroendocrine perspective. *J Affect Disord* **29**, 145–158 (1993).

57. Nottelmann, E. D., Inoff-Germain, G., Susman, E. J. & Chrousos, G. P. Hormones and behavior at puberty. (1990).

58. Horwitz, A. V. The sociological study of mental illness: A critique and synthesis of four perspectives. in *Handbook of the sociology of mental health* 95–112 (Springer, 2013).

59. Lee, M. & Mather, M. *US labor force trends*. vol. 63 (Population Reference Bureau, 2008).

60. England, P. *Households, employment, and gender: A social, economic, and demographic view*. (Routledge, 2017).

61. Bianchi, S. M., Milkie, M. A., Sayer, L. C. & Robinson, J. P. Is anyone doing the housework? Trends in the gender division of household labor. *Social forces* **79**, 191–228 (2000).

62. Blau, F. D. & Kahn, L. M. *Female labor supply: Why is the US falling behind?* (2013).

63. Vespa, J. *The Changing Economics and Demographics of Young Adulthood: 1975–2016*. https://www.census.gov/content/dam/Census/library/publications/2017/demo/p20-579.pdf (2017).

64. Rosenzweig, M. R. & Seiver, D. A. Education and contraceptive choice: A conditional demand framework. *International Economic Review* 171–198 (1982).

65. Goldin, C. & Katz, L. F. *The power of the pill: oral contraceptives and women's career and marriage decisions*. (2000).

66. Rosenzweig, M. R. & Schultz, T. P. The demand for and supply of births: Fertility and its life cycle consequences. *The American Economic Review* **75**, 992–1015 (1985).

67. Twenge, J. M. & Nolen-Hoeksema, S. Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: A meta-analysis. *Journal of abnormal psychology* **111**, 578 (2002).

68. Jane Costello, E., Erkanli, A. & Angold, A. Is there an epidemic of child or adolescent depression? *Journal of child psychology and psychiatry* **47**, 1263–1271 (2006).

69. Lewinsohn, P. M., Solomon, A., Seeley, J. R. & Zeiss, A. Clinical implications of" subthreshold" depressive symptoms. *Journal of abnormal psychology* **109**, 345 (2000).

70. O'Connor, E., Rossom, R. & Henninger, M. *Screening for Depression in Adults: An Updated Systematic Evidence Review for the U.S. Preventive Services Task Force: Depression Symptom Rating Scales.* (Agency for Healthcare Research and Quality (US), 2016).

71. Robins, L. N. Psychiatric disorders in America. in *The Epidemiologic Catchment Area Study* (1991).

72. Weissman, M. M., Bruce, M. L., Leaf, P. J., Florio, L. P. & Holzer, C. Affective disorders. *Psychiatric disorders in America* 53–80 (1991).

73. Schildkraut, J. & Klein, F. The classification and treatment of depressive states. *Manual of Psychiatric Therapeutics. Boston, Mass: Little, Brown & Co* **47**, (1975).

74. de Figueiredo, J. M. Some issues in research on the epidemiology of demoralization. *Comprehensive Psychiatry* **24**, 154–157 (1983).

75. Shader, R. I. Demoralization revisited. *Journal of Clinical Psychopharmacology* (2005).

76. Innovation, V. H. *Covidence systematic review software*. (Veritas Health Innovation Melbourne, Victoria Australia, 2017).

77. Fleiss, J. L., Levin, B. & Paik, M. C. *Statistical methods for rates and proportions*. (John Wiley & Sons, 2013).

78. Lipsey, M. W. & Wilson, D. B. *Practical meta-analysis.* (SAGE publications, Inc, 2001).

79. Cohen, J. Statistical power analysis for the social sciences. (1988).

80. Sutton, A. J. Publication bias. *The handbook of research synthesis and meta-analysis* **2**, 435–452 (2009).

81. Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Bmj* **315**, 629–634 (1997).

82. Duval, S. & Tweedie, R. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**, 455–463 (2000).

83. IntHout, J., Ioannidis, J. P., Rovers, M. M. & Goeman, J. J. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ open* **6**, e010247 (2016).

84. Schwarzer, G. meta: An R package for meta-analysis. *R news* **7**, 40–45 (2007).

85. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* **36**, 1–48 (2010).

86. Pigott, T. D. Handling missing data. *The handbook of research synthesis and meta-analysis* **2**, 399–416 (2009).

87. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2004).

88. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2005).

89. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2006).

90. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2007).

91. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2008).

92. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2009).

93. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2010).

94. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2011).

95. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2013).

96. Center for Behavioral Health Statistics and Quality. *National Survey on Drug Use and Health: Mental Health Detailed Tables*. (2017).

97. Hasin, D. S., Goodwin, R. D., Stinson, F. S. & Grant, B. F. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Archives of general psychiatry* **62**, 1097–1106 (2005).

98. Thibodeau, M. A. & Asmundson, G. J. G. The PHQ-9 assesses depression similarly in men and women from the general population. *Personality and Individual Differences* **56**, 149–153 (2014).

99. Gettler, L. T. & Oka, R. C. Are testosterone levels and depression risk linked based on partnering and parenting? Evidence from a large population-representative study of US men and women. *Social Science & Medicine* **163**, 157–167 (2016).

100. Ferketich, A. K., Schwartzbaum, J. A., Frid, D. J. & Moeschberger, M. L. Depression as an antecedent to heart disease among women and men in the NHANES I study. *Archives of internal medicine* **160**, 1261–1268 (2000).

101. Shah, A. J., Veledar, E., Hong, Y., Bremner, J. D. & Vaccarino, V. Depression and history of attempted suicide as risk factors for heart disease mortality in young individuals. *Archives of general psychiatry* **68**, 1135–1142 (2011).

102. Alaimo, K., Olson, C. M. & Frongillo, E. A. Family food insufficiency, but not low family income, is positively associated with dysthymia and suicide symptoms in adolescents. *The Journal of nutrition* **132**, 719–725 (2002).

103. Merikangas, A. K., Mendola, P., Pastor, P. N., Reuben, C. A. & Cleary, S. D. The association between major depressive disorder and obesity in US adolescents: results from the 2001–2004 National Health and Nutrition Examination Survey. *Journal of behavioral medicine* **35**, 149–154 (2012).

104. Hardie, J. H. The consequences of unrealized occupational goals in the transition to adulthood. *Social science research* **48**, 196–211 (2014).

105. Mumford, E. A., Liu, W., Hair, E. C. & Yu, T.-C. Concurrent trajectories of BMI and mental health patterns in emerging adulthood. *Soc Sci Med* **98**, 1–7 (2013).

106. Song, L. Social capital and psychological distress. *J Health Soc Behav* **52**, 478–492 (2011).

107. Bushman, B. J., Moeller, S. J., Konrath, S. & Crocker, J. Investigating the Link Between Liking Versus Wanting Self-Esteem and Depression in a Nationally Representative Sample of A merican Adults. *Journal of Personality* **80**, 1453–1469 (2012).

108. Toussaint, L. L., Williams, D. R., Musick, M. A. & Everson-Rose, S. A. The association of forgiveness and 12-month prevalence of major depressive episode: Gender differences in a probability sample of US adults. *Mental Health, Religion and Culture* **11**, 485–500 (2008).

109. Marmorstein, N. R. Longitudinal associations between alcohol problems and depressive symptoms: early adolescence through early adulthood. *Alcohol Clin Exp Res* **33**, 49–59 (2009).

110. Neumark-Sztainer, D. & Hannan, P. J. Weight-related behaviors among adolescent girls and boys: results from a national survey. *Arch Pediatr Adolesc Med* **154**, 569–577 (2000).

111. Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M. & Merikangas, K. R. Major depression in the National Comorbidity Survey–Adolescent Supplement: prevalence, correlates, and treatment. *Journal of the American Academy of Child & Adolescent Psychiatry* **54**, 37-44. e2 (2015).

112. Henderson, C. *et al.* Neighbourhood Characteristics, Individual Level Socioeconomic Factors, and Depressive Symptoms in Young Adults: The CARDIA Study. *Journal of Epidemiology and Community Health (1979-)* **59**, 322–328 (2005).

113. Margraf, J., Lavallee, K., Zhang, X. & Schneider, S. Social Rhythm and Mental Health: A Cross-Cultural Comparison. *PLoS One* **11**, e0150312 (2016).

114. Zemore, S. E., Karriker-Jaffe, K. J. & Mulia, N. Temporal trends and changing racial/ethnic disparities in alcohol problems: results from the 2000 to 2010 National Alcohol Surveys. *Journal of addiction research & therapy* **4**, (2013).

115. Toussaint, L. L., Williams, D. R., Musick, M. A. & Everson-Rose, S. A. The association of forgiveness and 12-month prevalence of major depressive episode: Gender differences in a probability sample of US adults. *Mental Health, Religion and Culture* **11**, 485–500 (2008).

116. Ojard, C., Donnelly, J. P., Safford, M. M. & Griffin, R. Psychosocial stress as a risk factor for sepsis: a population-based cohort study. *Psychosomatic medicine* **77**, 93 (2015).

117. Haroz, E. E., Ybarra, M. L. & Eaton, W. W. Psychometric evaluation of a self-report scale to measure adolescent depression: the CESDR-10 in two national adolescent samples in the United States. *J Affect Disord* **158**, 154–160 (2014).

118. Oksuzyan, A. *et al.* Cross-national comparison of sex differences in health and mortality in Denmark, Japan and the US. *Eur J Epidemiol* **25**, 471–480 (2010).

119. Dohrenwend, B. P. & Dohrenwend, B. S. Perspectives on the past and future of psychiatric epidemiology. The 1981 Rema Lapouse Lecture. *American Journal of Public Health* **72**, 1271–1279 (1982).

120. Frank, J. D. & Frank, J. B. *Persuasion and healing: A comparative study of psychotherapy*. (JHU Press, 1993).

121. Roberts, R. E. Epidemiological issues in measuring preventive effects. (1987).

122. Srole L & Fischer A. The midtown manhattan longitudinal study vs 'the mental paradise lost' doctrine: A controversy joined. *Archives of General Psychiatry* **37**, 209–221 (1980).

123. Wickramaratne, P. J., Weissman, M. M., Leaf, P. J. & Holford, T. R. Age, period and cohort effects on the risk of major depression: results from five United States communities. *Journal of clinical epidemiology* **42**, 333–343 (1989).

124. Yang, Y. & Lee, L. C. Sex and race disparities in health: Cohort variations in life course patterns. *Social Forces* **87**, 2093–2124 (2009).

125. Ross, C. E. & Huber, J. Hardship and depression. *Journal of Health and Social Behavior* 312–327 (1985).

126. Lorant, V. *et al.* Socioeconomic inequalities in depression: a meta-analysis. *American journal of epidemiology* **157**, 98–112 (2003).

127. Belle, D. & Doucet, J. Poverty, inequality, and discrimination as sources of depression among US women. *Psychology of Women Quarterly* **27**, 101–113 (2003).

128. Cotter, D., Hermsen, J. M. & Vanneman, R. The end of the gender revolution? Gender role attitudes from 1977 to 2008. *American Journal of Sociology* **117**, 259–89 (2011).

129. Linville, P. W. Self-complexity and affective extremity: Don't put all of your eggs in one cognitive basket. *Social cognition* **3**, 94 (1985).

130. Rafaeli-Mor, E. & Steinberg, J. Self-complexity and well-being: A review and research synthesis. *Personality and Social Psychology Review* **6**, 31–58 (2002).

131. Glass, J. & Fujimoto, T. Housework, paid work, and depression among husbands and wives. *Journal of Health and Social Behavior* 179–191 (1994).

132. Gove, W. R. The relationship between sex roles, marital status, and mental illness. *Social forces* **51**, 34–44 (1972).

133. Gove, W. R. & Tudor, J. F. Adult sex roles and mental illness. *American journal of Sociology* **78**, 812–835 (1973).

134. Hecht, L. M. Role conflict and role overload: Different concepts, different consequences. *Sociological Inquiry* **71**, 111–121 (2001).

135. Goode, W. J. A theory of role strain. *American sociological review* 483–496 (1960).

136. Hock, E. & DeMeis, D. K. Depression in mothers of infants: The role of maternal employment. *Developmental Psychology* **26**, 285 (1990).

137. Pearlin, L. I. Sex roles and depression. in *Life-span developmental psychology* 191–207 (Elsevier, 1975).

138. Blakemore, S.-J., Burnett, S. & Dahl, R. E. The role of puberty in the developing adolescent brain. *Human brain mapping* **31**, 926–33 (2010).

139. Hankin, B. L., Mermelstein, R. & Roesch, L. Sex differences in adolescent depression: Stress exposure and reactivity models. *Child development* **78**, 279–295 (2007).

140. Vicary, J. R., Klingaman, L. R. & Harkness, W. L. Risk factors associated with date rape and sexual assault of adolescent girls. *Journal of Adolescence* **18**, 289 (1995).

141. Copeland, W. *et al.* Outcomes of Early Pubertal Timing in Young Women: A Prospective Population-Based Study. *American Journal of Psychiatry* **167**, 1218–1225 (2010).

142. Ge, X., Conger, R. D. & Elder, G. H. Coming of age too early: Pubertal influences on girls' vulnerability to psychological distress. *Child development* **67**, 3386–3400 (1996).

143. Nolen-Hoeksema, S. Sex differences in unipolar depression: evidence and theory. *Psychological bulletin* **101**, 259 (1987).

144. Johnson, D. P. & Whisman, M. A. Gender differences in rumination: A meta-analysis. *Personality and Individual Differences* **55**, 367–374 (2013).

145. Collishaw, S. Annual research review: secular trends in child and adolescent mental health. *Journal of Child Psychology and Psychiatry* **56**, 370–393 (2015).

146. Kessler, R. C. *et al.* Sex and depression in the National Comorbidity Survey. II: Cohort effects. *Journal of affective disorders* **30**, 15–26 (1994).

147. Bor, W., Dean, A. J., Najman, J. & Hayatbakhsh, R. Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Aust N Z J Psychiatry* **48**, 606–616 (2014).

148. Kessel Schneider, S., O'Donnell, L. & Smith, E. Trends in cyberbullying and school bullying victimization in a regional census of high school students, 2006-2012. *Journal of school health* **85**, 611–620 (2015).

149. Steinfield, C., Ellison, N. B. & Lampe, C. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology* **29**, 434–445 (2008).

150. Lenhart, A., Smith, A., Anderson, M., Duggan, M. & Perrin, A. Teens, technology and friendships. (2015).

151. Augner, C. & Hacker, G. W. Associations between problematic mobile phone use and psychological parameters in young adults. *International journal of public health* **57**, 437–441 (2012).

152. Orben, A. & Przybylski, A. K. The association between adolescent well-being and digital technology use. *Nature Human Behaviour* **3**, 173 (2019).

153. Archuleta, K. L., Dale, A. & Spann, S. M. College Students and Financial Distress: Exploring Debt, Financial Satisfaction, and Financial Anxiety. *Journal of Financial Counseling and Planning* **24**, 50–62 (2013).

154. Roberts, R. *et al.* Mental and physical health in students: the role of economic circumstances. *British Journal of Health Psychology* **5**, 289–297 (2000).

155. Fitch, C., Hamilton, S., Bassett, P. & Davey, R. The relationship between personal debt and mental health: a systematic review. *Mental Health Review Journal* **16**, 153–166 (2011).

156. Larson, R. W. Globalization, societal change, and new technologies: What they mean for the future of adolescence. *Journal of Research on Adolescence* **12**, 1–30 (2002).

157. Mortimer, J. T. & Larson, R. W. Macrostructural trends and the reshaping of adolescence. *The changing adolescent experience: Societal trends and the transition to adulthood* 1–17 (2002).

158. Nolen-Hoeksema, S., Larson, J. & Grayson, C. Explaining the gender difference in depressive symptoms. *J Pers Soc Psychol* **77**, 1061–1072 (1999).

159. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry* **157**, 1552–1562 (2000).

160. O'hara, M. W. & Swain, A. M. Rates and risk of postpartum depression—a meta-analysis. *International review of psychiatry* **8**, 37–54 (1996).

161. Gavin, N. I. *et al.* Perinatal depression: a systematic review of prevalence and incidence. *Obstetrics & Gynecology* **106**, 1071–1083 (2005).

162. Paulson, J. F. & Bazemore, S. D. Prenatal and postpartum depression in fathers and its association with maternal depression: a meta-analysis. *Jama* **303**, 1961–1969 (2010).

163. Dennerstein, L., Lehert, P., Burger, H. & Dudley, E. Mood and the menopausal transition. *The Journal of nervous and mental disease* **187**, 685–691 (1999).

164. Rubinow, D. R., Johnson, S. L., Schmidt, P. J., Girdler, S. & Gaynes, B. Efficacy of estradiol in perimenopausal depression: so much promise and so few answers. *Depression and anxiety* **32**, 539–549 (2015).

165. Freeman, E. W., Sammel, M. D., Lin, H. & Nelson, D. B. Associations of hormones and menopausal status with depressed mood in women with no history of depression. *Archives of general psychiatry* **63**, 375–382 (2006).

166. Slap, G. B. Oral contraceptives and depression: impact, prevalence and cause. *Journal of Adolescent Health Care* **2**, 53–64 (1981).

167. Oinonen, K. A. & Mazmanian, D. To what extent do oral contraceptives influence mood and affect? *Journal of affective disorders* **70**, 229–240 (2002).

168. Dohrenwend, B. S. & Dohrenwend, B. P. *Stressful life events: Their nature and effects*. (John Wiley & Sons, 1974).

169. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry* (1999).

170. Repetti, R. L., Taylor, S. E. & Seeman, T. E. Risky families: family social environments and the mental and physical health of offspring. *Psychological bulletin* **128**, 330 (2002).

171. Taylor, S. E., Kemeny, M. E., Reed, G. M., Bower, J. E. & Gruenewald, T. L. Psychological resources, positive illusions, and health. *American psychologist* **55**, 99 (2000).

172. Pearlin, L. I., Menaghan, E. G., Lieberman, M. A. & Mullan, J. T. The stress process. *Journal of Health and Social behavior* 337–356 (1981).

173. Pearlin, L. I., Schieman, S., Fazio, E. M. & Meersman, S. C. Stress, health, and the life course: Some conceptual perspectives. *Journal of Health and Social Behavior* **46**, 205–219 (2005).

174. McEwen, B. S. Protective and damaging effects of stress mediators. *New England journal of medicine* **338**, 171–179 (1998).

175. Tolin, D. F. & Foa, E. B. Sex differences in trauma and posttraumatic stress disorder: a quantitative review of 25 years of research. *Psychological bulletin* **132**, 959 (2006).

176. Hatch, S. L. & Dohrenwend, B. P. Distribution of traumatic and other stressful life events by race/ethnicity, gender, SES and age: a review of the research. *American journal of community psychology* **40**, 313–332 (2007).

177. Aneshensel, C. S., Rutter, C. M. & Lachenbruch, P. A. Social structure, stress, and mental health: Competing conceptual and analytic models. *American sociological review* 166–178 (1991).

178. Young, E. & Korszun, A. Sex, trauma, stress hormones and depression. *Molecular psychiatry* **15**, 23–28 (2010).

179. Wheaton, B. The Sociogenesis of Psychological Disorder: An Attributional Theory. *Journal of Health and Social Behavior* **21**, 100–124 (1980).

180. Nolen-Hoeksema, S. Responses to depression and their effects on the duration of depressive episodes. *Journal of abnormal psychology* **100**, 569 (1991).

181. Wheaton, B. Stress, Personal Coping Resources, and Psychiatric Symptoms: An Investigation of Interactive Models. *Journal of Health and Social Behavior* **24**, 208–229 (1983).

182. Dohrenwend, B. J. Socioeconomic status (SES) and psychiatric disorders. Are the issues still compelling? *Social psychiatry and psychiatric epidemiology* **25**, 41–47 (1990).

183. Connell, R. *Gender and power*. (Polity Press, 2003).

184. Acker, J. Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations. *Gender & Society* **4**, 139–158 (1990).

185. Rose, S. O. *What is gender history*. (Polity, 2010).

186. Lorber, J. *Paradoxes of gender*. (Yale University Press, 1994).

187. Marmot, M. G. Understanding Social Inequalities in Health. *Perspectives in Biology and Medicine* **46**, S9–S23 (2003).

188. Platt, J., Prins, S., Bates, L. & Keyes, K. Unequal depression for equal work? How the wage gap explains gendered disparities in mood disorders. *Social Science & Medicine* **149**, 1–8 (2016).

189. Ryder, N. B. The cohort as a concept in the study of social change. in *Cohort analysis in social research* 9–44 (Springer, 1985).

190. Weissman, M. M. *et al.* Sex differences in rates of depression: cross-national perspectives. *Journal of affective disorders* **29**, 77–84 (1993).

191. Medalia, C. Historical Changes in the Gender Gap in Depression in the United States, 1971 to 2008.

192. Murphy, J. M. Trends in depression and anxiety: men and women. *Acta Psychiatrica Scandinavica* **73**, 113–127 (1986).

193. Goldberg, D. GHQ-12. *London: NFER-Nelson* (1978).

194. Goodman, A. & Goodman, R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry* **48**, 400–403 (2009).

195. Springer, K. W., Lee, C. & Carr, D. Spousal Breadwinning Across 30 Years of Marriage and Husbands' Health: A Gendered Life Course Stress Approach. *Journal of aging and health* **31**, 37–66 (2019).

196. Bird, C. E. Gender, household labor, and psychological distress: The impact of the amount and division of housework. *Journal of Health and Social Behavior* 32–45 (1999).

197. Cuijpers, P., de Graaf, R. & van Dorsselaer, S. Minor depression: risk profiles, functional disability, health care use and risk of developing major depression. *Journal of affective disorders* **79**, 71–79 (2004).

198. Fergusson, D. M., Horwood, L. J., Ridder, E. M. & Beautrais, A. L. Subthreshold depression in adolescence and mental health outcomes in adulthood. *Archives of general psychiatry* **62**, 66–72 (2005).

199. Shankman, S. A. *et al.* Subthreshold conditions as precursors for full syndrome disorders: a 15-year longitudinal study of multiple diagnostic classes. *Journal of Child Psychology and Psychiatry* **50**, 1485–1494 (2009).

200. Center for Human Resource Research. National Longitudinal Survey of Youth, 1979: Archival Version. (2007) doi:10.3886/ICPSR04683.

201. Center for Human Resource Research. Children of the NLSY79, 1979-2014. (2017) doi:10.3886/ICPSR03959.

202. Bureau of Labor Statistics. National Longitudinal Study main website. https://www.nlsinfo.org/ (2018).

203. Rubin, D. B. *Multiple imputation for nonresponse in surveys*. vol. 81 (John Wiley & Sons, 2004).

204. Hughes, M. E. & Angela, M. *The lives and times of the baby boomers*. (Russell Sage Foundation New York, 2004).

205. Yang, Y. Is old age depressing? Growth trajectories and cohort variations in late-life depression. *Journal of health and social behavior* **48**, 16–32 (2007).

206. Cole, J. C., Rabin, A. S., Smith, T. L. & Kaufman, A. S. Development and validation of a Rasch-derived CES-D short form. *Psychological assessment* **16**, 360 (2004).

207. Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement* **1**, 385–401 (1977).

208. Vilagut, G., Forero, C. G., Barbaglia, G. & Alonso, J. Screening for Depression in the General Population with the Center for Epidemiologic Studies Depression (CES-D): A Systematic Review with Meta-Analysis. *PLOS ONE* **11**, e0155431 (2016).

209. Levine, S. Z. Evaluating the seven-item Center for Epidemiologic Studies Depression Scale short-form: a longitudinal US community study. *Social psychiatry and psychiatric epidemiology* **48**, 1519–1526 (2013).

210. Harder, V. S., Morral, A. R. & Arkes, J. Marijuana use and depression among adults: Testing for causal associations. *Addiction* **101**, 1463–1472 (2006).

211. Cseh, A. The effects of depressive symptoms on earnings. *Southern Economic Journal* 383–409 (2008).

212. Prause, J., Dooley, D. & Huh, J. Income volatility and psychological depression. *American Journal of Community Psychology* **43**, 57–70 (2009).

213. Carleton, R. N. *et al.* The center for epidemiologic studies depression scale: a review with a theoretical and empirical examination of item content and factor structure. *PloS one* **8**, e58067 (2013).

214. Hertzog, C., Van Alstine, J., Usala, P. D., Hultsch, D. F. & Dixon, R. Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* **2**, 64 (1990).

215. Lewinsohn, P. M., Seeley, J. R., Roberts, R. E. & Allen, N. B. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and aging* **12**, 277 (1997).

216. Roberts, R. E., Andrews, J. A., Lewinsohn, P. M. & Hops, H. Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* **2**, 122 (1990).

217. Radloff, L. S. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. *Journal of youth and adolescence* **20**, 149–166 (1991).

218. Cole, S. R., Kawachi, I., Maller, S. J. & Berkman, L. F. Test of item-response bias in the CES-D scale: Experience from the New Haven EPESE study. *Journal of clinical epidemiology* **53**, 285–289 (2000).

219. Census, U. B. *Current Population Survey Historical data series*. https://www.census.gov/data/tables.html.

220. Hamermesh, D. S., Frazis, H. & Stewart, J. Data watch: The American time use survey. *Journal of Economic Perspectives* **19**, 221–232 (2005).

221. Fisher, K. *et al.* Introducing the American Heritage Time Use Study Data Extract Builder (AHTUS-X). *Electron Int J Time Use Res* **12**, 169–173 (2015).

222. Waite, L. J. & Nielsen, M. The rise of the dual-earner family. *Working families: The transformation of the American home* 23–41 (2001).

223. Yang, Y. & Land, K. C. Mixed Effects Models: Hierarchical APC-Growth Curve Analysis of Prospective Cohort Data. in *Age-period-cohort analysis: New models, methods, and empirical applications* 285–312 (Chapman and Hall/CRC, 2016).

224. Jager, J., Schulenberg, J. E., O'Malley, P. M. & Bachman, J. G. Historical variation in drug use trajectories across the transition to adulthood: the trend toward lower intercepts and steeper, ascending slopes. *Development and psychopathology* **25**, 527–543 (2013).

225. Keyes, K. M., Utz, R. L., Robinson, W. & Li, G. What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971–2006. *Social science & medicine* **70**, 1100–1108 (2010).

226. Miyazaki, Y. & Raudenbush, S. W. Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological methods* **5**, 44 (2000).

227. Tom, A., Bosker, T. A. S. R. J. & Bosker, R. J. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. (Sage, 1999).

228. Seltzer, M. & Choi, K. Sensitivity analysis for multilevel models. *Multilevel modeling: Methodological advances, issues, and applications. Mahwah, NJ: Lawrence Earlbaum* (2002).

229. VanderWeele, T. *Explanation in causal inference: methods for mediation and interaction*. (Oxford University Press, 2015).

230. Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychological methods* **15**, 309 (2010).

231. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. (2014).

232. Imai, K., Keele, L., Tingley, D. & Yamamoto, T. Causal mediation analysis using R. in *Advances in social science research using R* 129–154 (Springer, 2010).

233. R Core Team. R: A Language and Environment for Statistical Computing. (2013).

234. Bates, D., Sarkar, D., Bates, M. D. & Matrix, L. The lme4 package. *R package version* **2**, 74 (2007).

235. Knowles, J. E. & Frederick, C. merTools: tools for analyzing mixed effect regression models. *R package version 0.3. 0* (2016).

236. Honaker, J., King, G. & Blackwell, M. Amelia II: A program for missing data. *Journal of statistical software* **45**, 1–47 (2011).

237. Sayer, L. C. Trends in housework. *Dividing the domestic: Men, women, and household work in cross-national perspective* 19–38 (2010).

238. Bianchi, S. M., Robinson, J. P. & Milke, M. A. *The changing rhythms of American family life*. (Russell Sage Foundation, 2006).

239. Van de Velde, S., Bracke, P. & Levecque, K. Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Social Science & Medicine* **71**, 305–313 (2010).

240. Joyce, P. R., Oakley-Browne, M. A., Wells, J. E., Bushnell, J. A. & Hornblow, A. R. Birth cohort trends in major depression: increasing rates and earlier onset in New Zealand. *Journal of Affective Disorders* **18**, 83–89 (1990).

241. Wauterickx, N. & Bracke, P. Unipolar depression in the Belgian population. *Social Psychiatry and Psychiatric Epidemiology* **40**, 691–699 (2005).

242. Hosseinpoor, A. R. *et al.* Social determinants of self-reported health in women and men: understanding the role of gender in population health. *PloS one* **7**, e34799 (2012).

243. Seedat, S. *et al.* Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. *Arch. Gen. Psychiatry* **66**, 785–795 (2009).

244. Pearse, R. & Connell, R. Gender Norms and the Economy: Insights from Social Research. *Feminist Economics* **22**, 30–53 (2016).

245. Goldin, C., Katz, L. F. & Kuziemko, I. The homecoming of American college women: The reversal of the college gender gap. *The Journal of Economic Perspectives* **20**, 133–133 (2006).

246. Diprete, T. A. & Buchmann, C. Gender-specific trends in the value of education and the emerging gender gap in college completion. *Demography* **43**, 1–24 (2006).

247. Ross, C. E. & Mirowsky, J. Sex differences in the effect of education on depression: resource multiplication or resource substitution? *Soc Sci Med* **63**, 1400–1413 (2006).

248. John Mirowsky & Catherine E. Ross. Education, Personal Control, Lifestyle and Health: A Human Capital Hypothesis. *Res Aging* **20**, 415–449 (1998).

249. Cha, Y. & Weeden, K. A. Overwork and the slow convergence in the gender gap in wages. *American Sociological Review* 0003122414528936 (2014).

250. Boyle, P., Cooke, T. J., Halfacree, K. & Smith, D. A cross-national comparison of the impact of family migration on women's employment status. *Demography* **38**, 201–213 (2001).

251. Kleven, H., Landais, C. & Søgaard, J. E. *Children and gender inequality: Evidence from Denmark*. (2018).

252. Goldin, C., Kerr, S. P., Olivetti, C. & Barth, E. The expanding gender earnings gap: evidence from the LEHD-2000 Census. *American Economic Review* **107**, 110–14 (2017).

253. Shelton, B. A. & John, D. The division of household labor. *Annual review of sociology* **22**, 299–322 (1996).

254. Kiecolt-Glaser, J. K. & Newton, T. L. Marriage and health: his and hers. *Psychological bulletin* **127**, 472 (2001).

255. Sayer, L. C., Bianchi, S. M. & Robinson, J. P. Are Parents Investing Less in Children? Trends in Mothers' and Fathers' Time with Children. *American Journal of Sociology* **110**, 1–43 (2004).

256. Raley, S., Bianchi, S. M. & Wang, W. When do fathers care? Mothers' economic contribution and fathers' involvement in child care. *American Journal of Sociology* **117**, 1422–59 (2012).

257. Ozer, E. M. The impact of childcare responsibility and self-efficacy on the psychological health of professional working mothers. *Psychology of Women Quarterly* **19**, 315–335 (1995).

258. Goldberg, A. E. & Perry-Jenkins, M. Division of labor and working-class women's well-being across the transition to parenthood. *Journal of Family Psychology* **18**, 225 (2004).

259. Cohn, D. & Passel, J. A record 64 million Americans live in multigenerational households. https://www.pewresearch.org/fact-tank/2018/04/05/a-record-64-million-americans-live-in-multigenerational-households/ (2018).

260. Muennig, P., Jiao, B. & Singer, E. Living with parents or grandparents increases social capital and survival: 2014 General Social Survey-National Death Index. *SSM Popul Health* **4**, 71–75 (2017).

261. Chisholm, J. F. The sandwich generation. *Journal of Social Distress and the Homeless* **8**, 177–191 (1999).

262. Kessler, R. C. *et al.* Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of general psychiatry* **51**, 8–19 (1994).

263. Horwitz, A. V. *Creating mental illness*. (University of Chicago Press, 2002).

264. Rogers, A. & Pilgrim, D. *A sociology of mental health and illness*. (McGraw-Hill Education (UK), 2014).

265. Mayes, R. & Horwitz, A. V. DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences* **41**, 249–267 (2005).

266. Jager, J., Keyes, K. M. & Schulenberg, J. E. Historical variation in young adult binge drinking trajectories and its link to historical variation in social roles and minimum legal drinking age. *Developmental psychology* **51**, 962 (2015).

267. Caspi, A. *et al.* The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science* **2**, 119–137 (2014).

268. Eaton, N. R. *et al.* An invariant dimensional liability model of gender differences in mental disorder prevalence: evidence from a national sample. *Journal of abnormal psychology* **121**, 282 (2012).

269. Ernst, C. & Angst, J. The Zurich Study. XII. Sex differences in depression. Evidence from longitudinal epidemiological data. *Eur Arch Psychiatry Clin Neurosci* **241**, 222–230 (1992).

270. Lewinsohn, P. M., Rohde, P., Seeley, J. R. & Fischer, S. A. Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *J Abnorm Psychol* **102**, 110–120 (1993).

271. Hochschild, A. R. *The Second Shift: Working Parents And The Revolution*. (Viking, 1989).

272. England, P. The gender revolution: Uneven and stalled. *Gender & society* **24**, 149–166 (2010).

273. Anderson, D. J., Binder, M. & Krause, K. The motherhood wage penalty revisited: Experience, heterogeneity, work effort, and work-schedule flexibility. *ILR Review* **56**, 273–294 (2003).

274. Hodges, M. J. & Budig, M. J. Who gets the daddy bonus? Organizational hegemonic masculinity and the impact of fatherhood on earnings. *Gender & Society* **24**, 717–745 (2010).

275. Lachance-Grzela, M. & Bouchard, G. Why do women do the lion's share of housework? A decade of research. *Sex roles* **63**, 767–780 (2010).

276. Coltrane, S. Research on household labor: Modeling and measuring the social embeddedness of routine family work. *Journal of Marriage and family* **62**, 1208–1233 (2000).

277. Hochschild, A. & Machung, A. *The second shift: Working families and the revolution at home*. (Penguin, 2012).

278. Bittman, M., England, P., Sayer, L., Folbre, N. & Matheson, G. When does gender trump money? Bargaining and time in household work. *American Journal of sociology* **109**, 186–214 (2003).

279. Marshall, N. L. & Barnett, R. C. Work-family strains and gains among two-earner couples. *Journal of Community Psychology* **21**, 64–78 (1993).

280. Crosby, F. J. & Jaskar, K. L. *Women and men at home and at work: Realities and illusions.* (Sage Publications, Inc, 1993).

281. Sieber, S. D. Toward a theory of role accumulation. *American sociological review* 567–578 (1974).

282. Biddle, B. J. *Role theory: Expectations, identities, and behaviors*. (Academic Press, 2013).

283. Greenhaus, J. H. & Powell, G. N. When work and family collide: Deciding between competing role demands. *Organizational Behavior and Human Decision Processes* **90**, 291–303 (2003).

284. Berkman, L. F. & O'Donnell, E. M. The Pro-family Workplace: Social and Economic Policies and Practices and Their Impacts on Child and Family Health. in *Families and Child Health* (eds. Landale, N. S., McHale, S. M. & Booth, A.) 157–179 (Springer New York, 2013). doi:10.1007/978-1-4614-6194-4_12.

285. Kelly, E. L., Moen, P. & Tranby, E. Changing workplaces to reduce work-family conflict: Schedule control in a white-collar organization. *American Sociological Review* **76**, 265–290 (2011).

286. Pavalko, E. K. & Henderson, K. A. Combining care work and paid work: Do workplace policies make a difference? *Research on Aging* **28**, 359–374 (2006).

287. Baltes, B. B., Briggs, T. E., Huff, J. W., Wright, J. A. & Neuman, G. A. Flexible and compressed workweek schedules: A meta-analysis of their effects on work-related criteria. *Journal of Applied Psychology* **84**, 496 (1999).

288. Allen, T. D. Family-supportive work environments: The role of organizational perceptions. *Journal of vocational behavior* **58**, 414–435 (2001).

289. Glass, J. L. & Finley, A. Coverage and effectiveness of family-responsive workplace policies. *Human resource management review* **12**, 313–337 (2002).

290. Mishel, L., Bivens, J., Gould, E. & Shierholz, H. *The state of working America*. (Cornell University Press, 2012).

291. Greenhaus, J. H., Allen, T. D. & Spector, P. E. Health consequences of work–family conflict: The dark side of the work–family interface. in *Employee health, coping and methodologies* 61–98 (Emerald Group Publishing Limited, 2006).

292. Moen, P. & Forest, K. B. Working Parents, Workplace Supports, and Well-Being: The Swedish Experience. *Social Psychology Quarterly* **53**, 117–131 (1990).

293. Perry-Jenkins, M., Smith, J. Z., Wadsworth, L. P. & Halpern, H. P. Workplace policies and mental health among working-class, new parents. *Community, work & family* **20**, 226–249 (2017).

294. Avendano, M., Berkman, L. F., Brugiavini, A. & Pasini, G. The long-run effect of maternity leave benefits on mental health: Evidence from European countries. *Social Science & Medicine* **132**, 45–53 (2015).

295. Aitken, Z. *et al.* The maternal health outcomes of paid maternity leave: A systematic review. *Social Science & Medicine* **130**, 32–41 (2015).

296. Moen, P., Fan, W. & Kelly, E. L. Team-level flexibility, work–home spillover, and health behavior. *Social Science & Medicine* **84**, 69–79 (2013).

297. Dolan, S. *Stress, self-esteem, health and work.* (Springer, 2006).

298. Fujishiro, K., Xu, J. & Gong, F. What does "occupation" represent as an indicator of socioeconomic status?: Exploring occupational prestige and health. *Social science & medicine* **71**, 2100–2107 (2010).

299. Batt, R. & Valcour, P. M. Human resources practices as predictors of work-family outcomes and employee turnover. *Industrial Relations: A Journal of Economy and Society* **42**, 189–220 (2003).

300. Hill, R., Tranby, E., Kelly, E. & Moen, P. Relieving the Time Squeeze? Effects of a White-Collar Workplace Change on Parents. *Journal of Marriage and Family* **75**, 1014–1029 (2013).

301. Kiehl, E. M. & White, M. A. Maternal adaptation during childbearing in Norway, Sweden and the United States. *Scandinavian Journal of Caring Sciences* **17**, 96–103 (2003).

302. Chatterji, P. & Markowitz, S. Family leave after childbirth and the mental health of new mothers. *Journal of Mental Health Policy and Economics* **15**, 61 (2012).

303. Center for Human Resource Research. National Longitudinal Survey of Youth, 1997: Archival Version. (2004) doi:10.3886/ICPSR03959.

304. Center for Human Resource Research. National Longitudinal Survey of Older Men, 1966-1990: Archival Version. (2007) doi:10.3886/ICPSR04675.

305. Center for Human Resource Research (CHRR). National Longitudinal Survey of Mature Women, 1967-2003: Archival Version. (2007) doi:10.3886/ICPSR04681.

306. Center for Human Resource Research (CHRR). National Longitudinal Survey of Young Women, 1968-2003: Archival Version. (2007) doi:10.3886/ICPSR04680.

307. Glass, J. & Fujimoto, T. Employer characteristics and the provision of family responsive policies. *Work and Occupations* **22**, 380–411 (1995).

308. Hardin, J. W. & Hilbe, J. M. *Generalized estimating equations*. (Chapman and Hall/CRC, 2002).

309. Hanley, J. A., Negassa, A. & Forrester, J. E. Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology* **157**, 364–375 (2003).

310. VanderWeele, T. J. & Knol, M. J. A tutorial on interaction. *Epidemiologic Methods* **3**, 33–72 (2014).

311. Thoits, P. A. Identity structures and psychological well-being: Gender and marital status comparisons. *Social Psychology Quarterly* 236–256 (1992).

312. Thoits, P. A. Multiple identities: Examining gender and marital status differences in distress. *American Sociological Review* 259–272 (1986).

313. Fingerman, K. L. Millennials and Their Parents: Implications of the New Young Adulthood for Midlife Adults. *Innov Aging* **1**, (2017).

314. Aumann, K., Galinsky, E. & Matos, K. The new male mystique. *Families and Work Institute national study of the changing workforce* (2011).

315. Blair-Loy, M., Hochschild, A., Pugh, A. J., Williams, J. C. & Hartmann, H. Stability and transformation in gender, work, and family: insights from the second shift for the next quarter century. *Community, Work & Family* **18**, 435–454 (2015).

316. Berger, L. M., Hill, J. & Waldfogel, J. Maternity leave, early maternal employment and child health and development in the US. *The Economic Journal* **115**, F29–F47 (2005).

317. Cooklin, A. R. & FISHER, J. R. Employee entitlements during pregnancy and maternal psychological well-being. *Australian and New Zealand Journal of Obstetrics and Gynaecology* **47**, 483–490 (2007).

318. Estes, S. B. Work–Family Arrangements and Parenting: Are "Family-Friendly" Arrangements Related to Mothers' Involvement in Children's Lives? *Sociological Perspectives* **48**, 293–317 (2005).

319. Cooklin, A. R., Canterford, L., Strazdins, L. & Nicholson, J. M. Employment conditions and maternal postpartum mental health: results from the Longitudinal Study of Australian Children. *Archives of women's mental health* **14**, 217–225 (2011).

320. Baughman, R., DiNardi, D. & Holtz-Eakin, D. Productivity and wage effects of "family-friendly" fringe benefits. *International Journal of Manpower* **24**, 247–259 (2003).

321. Lee, S. Women's work supports, job retention, and job mobility: Child care and employer-provided health insurance help women stay on jobs. (2004).

322. Frank, R. G., McGuire, T. G., Normand, S.-L. T. & Goldman, H. H. The Value Of Mental Health Care At The System Level: The Case Of Treating Depression: Contrary to popular belief, mental health benefits in private health insurance represent solid value for the money spent. *Health Affairs* **18**, 71–88 (1999).

323. Kim, T. & von Dem Knesebeck, O. Perceived job insecurity, unemployment and depressive symptoms: a systematic review and meta-analysis of prospective observational studies. *International archives of occupational and environmental health* **89**, 561–573 (2016).

324. Blair-Loy, M. & Wharton, A. S. Employees' use of work-family policies and the workplace social context. *Social Forces* **80**, 813–845 (2002).

325. Kalmijn, M. & Monden, C. W. S. The division of labor and depressive symptoms at the couple level: Effects of equity or specialization? *Journal of Social and Personal Relationships* **29**, 358–374 (2012).

326. Kramer, K. Z. & Pak, S. Relative Earnings and Depressive Symptoms among Working Parents: Gender Differences in the Effect of Relative Income on Depressive Symptoms. *Sex Roles* **78**, 744–759 (2018).

327. Byrne, B. M., Baron, P. & Campbell, T. L. Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender. *Journal of Research on Adolescence* **3**, 127–143 (1993).

328. Van de Velde, S., Bracke, P., Levecque, K. & Meuleman, B. Gender differences in depression in 25 European countries after eliminating measurement bias in the CES-D 8. *Social Science Research* **39**, 396–404 (2010).

329. Carle, A. C., Millsap, R. E. & Cole, D. A. Measurement bias across gender on the Children's Depression Inventory: Evidence for invariance from two latent variable models. *Educational and Psychological Measurement* **68**, 281–303 (2008).

330. Dekker, J., Koelen, J. A., Peen, J., Schoevers, R. A. & Gijsbers-Van Wijk, C. Gender differences in clinical features of depressed outpatients: preliminary evidence for subtyping of depression? *Women & health* **46**, 19–38 (2008).

331. Silverstein, B. Gender difference in the prevalence of clinical depression: the role played by depression associated with somatic symptoms. *Am J Psychiatry* **156**, 480–482 (1999).

332. Silverstein, B. Gender differences in the prevalence of somatic versus pure depression: A replication. *The American Journal of Psychiatry* **159**, 1051–2 (2002).

333. Delisle, V. C., Beck, A. T., Dobson, K. S., Dozois, D. J. A. & Thombs, B. D. Revisiting gender differences in somatic symptoms of depression: much ado about nothing? *PLoS ONE* **7**, e32490 (2012).

334. Hoenig, J. M. & Heisey, D. M. The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician* **55**, 19–24 (2001).

335. Link, B. G. & Phelan, J. Social conditions as fundamental causes of disease. *Journal of health and social behavior* 80–94 (1995).

336. Keyes, K. M., Gary, D., O'Malley, P. M., Hamilton, A. & Schulenberg, J. E. Recent increases in depressive symptoms among US adolescents: trends from 1991-2018. *Social Psychiatry and Psychiatric Epidemiology* (2019).

337. Kowalski, R. M., Giumetti, G. W., Schroeder, A. N. & Lattanner, M. R. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. *Psychol Bull* **140**, 1073–1137 (2014).

338. Ging, D. & Norman, J. O. Cyberbullying, conflict management or just messing? Teenage girls' understandings and experiences of gender, friendship, and conflict on Facebook in an Irish second-level school. *Feminist Media Studies* **16**, 805–821 (2016).

339. Keles, B., McCrae, N. & Grealish, A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth* 1–15 (2019) doi:10.1080/02673843.2019.1590851.

340. Seabrook, E. M., Kern, M. L. & Rickard, N. S. Social Networking Sites, Depression, and Anxiety: A Systematic Review. *JMIR Mental Health* **3**, e50 (2016).

341. Blau, F. D. & Kahn, L. M. The gender pay gap have women gone as far as they can? *The Academy of Management Perspectives* **21**, 7–23 (2007).

342. McDonald, P. Workplace Sexual Harassment 30 Years on: A Review of the Literature. *International Journal of Management Reviews* **14**, 1–17 (2012).

343. Babcock, J. C., Waltz, J., Jacobson, N. S. & Gottman, J. M. Power and violence: the relation between communication patterns, power discrepancies, and domestic violence. *Journal of consulting and clinical psychology* **61**, 40 (1993).

344. Persson, P. & Rossin-Slater, M. *When Dad Can Stay Home: Fathers' Workplace Flexibility and Maternal Health*. (2019).

345. Sandberg, J. C. The effects of family obligations and workplace resources on men's and women's use of family leaves. *RESEARCH IN SOCIOLOGY OF WORK* 261–281 (1999).

346. Orme, J. G., Reis, J. & Herz, E. J. Factorial and discriminant validity of the Center for Epidemiological Studies Depression (CES-D) scale. *J Clin Psychol* **42**, 28–33 (1986).

347. Kotov, R. The Hierarchical Taxonomy of Psychopathology (HiTOP): A Dimensional Alternative to Traditional Nosologies. *Journal of abnormal psychology (1965)* (2017) doi:10.1037/abn0000258.

348. Muehlenkamp, J. J., Williams, K. L., Gutierrez, P. M. & Claes, L. Rates of non-suicidal self-injury in high school students across five years. *Arch Suicide Res* **13**, 317–329 (2009).

349. Lowry, R., Crosby, A. E., Brener, N. D. & Kann, L. Suicidal thoughts and attempts among US high school students: trends and associated health-risk behaviors, 1991–2011. *Journal of Adolescent Health* **54**, 100–108 (2014).

350. Pontes, N. M., Ayres, C. G. & Pontes, M. C. Additive Interactions Between Gender and Bullying Victimization on Depressive Symptoms and Suicidality: Youth Risk Behavior Survey 2011–2015. *Nursing research* **67**, 430–438 (2018).

351. Kessler, R. C., McGonagle, K. A., Swartz, M., Blazer, D. G. & Nelson, C. B. Sex and depression in the National Comorbidity Survey. I: Lifetime prevalence, chronicity and recurrence. *J Affect Disord* **29**, 85–96 (1993).

352. Dawson D.A. & Grant B.F. Family history of alcoholism and gender: Their combined effects on DSM- IV alcohol dependence and major depression. *J. Stud. Alcohol* **59**, 97–106 (1997).

353. Verplaetse, T. L., Smith, P. H., Pittman, B. P., Mazure, C. M. & McKee, S. A. Associations of Gender, Smoking, and Stress with Transitions in Major Depression Diagnoses. *Yale J Biol Med* **89**, 123–129 (2016).

354. Everson-Rose, S. A., House, J. S. & Mero, R. P. Depressive symptoms and mortality risk in a national sample: confounding effects of health status. *Psychosomatic medicine* **66**, 823–830 (2004).

355. Inaba, A. *et al.* Depression in the United States and Japan: gender, marital status, and SES patterns. *Social science & medicine* **61**, 2280–2292 (2005).

356. Shiovitz-Ezra, S., Leitsch, S., Graber, J. & Karraker, A. Quality of life and psychological health indicators in the national social life, health, and aging project. *J Gerontol B Psychol Sci Soc Sci* **64 Suppl 1**, i30-37 (2009).

357. Wang Z., Goswami D., Fitzgibbon H., Stockmeier C. & Austin M. Gender-specific gene expression of a novel zinc finger transcription factor, KLF7, in the prefrontal cortex of subjects with major depression. *Biol. Psychiatry* **67**, 135S (2010).

358. Assari, S. & Caldwell, C. H. Gender and Ethnic Differences in the Association Between Obesity and Depression Among Black Adolescents. *J Racial Ethn Health Disparities* **2**, 481–493 (2015).

359. Coyne, J. C. & Marcus, S. C. Health Disparities in Care for Depression Possibly Obscured by the Clinical Significance Criterion. *AJP* **163**, 1577–1579 (2006).

360. Danielson, C. K., de Arellano, M. A., Kilpatrick, D. G., Saunders, B. E. & Resnick, H. S. Child maltreatment in depressed adolescents: Differences in symptomatology based on history of abuse. *Child Maltreatment* **10**, 37–48 (2005).

361. Gavin, A. R., Rue, T. & Takeuchi, D. Racial/Ethnic Differences in the Association Between Obesity and Major Depressive Disorder: Findings from the Comprehensive Psychiatric Epidemiology Surveys. *Public Health Reports (1974-)* **125**, 698–708 (2010).

362. Goodwin, R. D. & Gotlib, I. H. Gender differences in depression: The role of personality factors. *Psychiatry Research* **126**, 135–142 (2004).

363. Grant, B. F. Comorbidity between DSM-IV drug use disorders and major depression: Results of a national survey of adults. *Journal of Substance Abuse* **7**, 481–497 (1995).

364. Kessler, R. C. *et al.* Sex and depression in the National Comorbidity Survey. II: Cohort effects. *J Affect Disord* **30**, 15–26 (1994).

365. Oquendo, M. A. *et al.* Ethnic and sex differences in suicide rates relative to major depression in the United States. *Am J Psychiatry* **158**, 1652–1658 (2001).

366. Zinzow, H. M. *et al.* Prevalence and mental health correlates of witnessed parental and community violence in a national sample of adolescents. *Journal of child Psychology and psychiatry* **50**, 441–450 (2009).

367. Guarnaccia, P. J., Guarnaccia, P. J. & Worobey, J. L. The impact of marital status and employment status on depressive affect for hispanic Americans. *Journal of community psychology* **19**, 136–149 (1991).

368. Seaton, E. K., Caldwell, C. H., Sellers, R. M. & Jackson, J. S. The prevalence of perceived discrimination among African American and Caribbean Black youth. *Developmental psychology* **44**, 1288 (2008).

XI. Supplementary material

Supplementary Table 1.1. Analytic dataset for diagnostic studies

Author	Study BL year	Age group	N Men	N Women	Instrument	recall	Women D+	Men D+
Alaimo et al, 2002[102]	1991	1	365	389	0	3	34	14
Assari et al, 2015[358]	2003	1	563	605	4	2	24	25
Avenevoli et al, 2015[111]	2002	1	843	809	4	2		
	2002	1	883	1004	4	2		
	2002	1	966	1044	4	2		
	2002	1	1088	1130	4	2		
	2002	1	1173	1183	4	2		
CBHSQ, 2004[87]	2004	1	11363	10938	4	2	1433	568
CBHSQ, 2005[88]	2005	4	3142	3420	4	2	226	104
	2005	2	7823	9132	4	2	959	485
	2005	1	10697	10444	4	2	1243	610
	2005	1	11378	11156	4	2	1484	512
CBHSQ, 2006[89]	2006	4	2888	3804	4	2	236	107
	2006	2	7431	8606	4	2	921	461
	2006	1	9158	11526	4	2	1187	522
	2006	1	11718	11153	4	2	1316	492
CBHSQ, 2007[90]	2007	4	2857	3509	4	2	218	106
	2007	2	7770	9114	4	2	975	482
	2007	2	10645	11542	4	2	1247	585
	2007	1	11524	10909	4	2	1298	530
CBHSQ, 2008[91]	2008	4	2996	3613	4	2	217	84
	2008	2	7440	8936	4	2	786	439
	2008	2	11166	12039	4	2	1397	614
	2008	1	11517	11029	4	2	1379	495

CBHSQ, 2009[92]	2009	4	3060	3690	4	2	221	110
	2009	2	7591	8729	4	2	838	425
	2009	2	11104	11900	4	2	1261	611
	2009	1	11520	11106	4	2	1299	541
CBHSQ, 2010[93]	2010	1	11140	10820	4	2	1288	490
	2010	2	17283	16788	4	2	1981	899
CBHSQ, 2011[94]	2011	1	12028	11482	4	2	1389	541
	2011	2	17178	17123	4	2	1866	962
CBHSQ, 2013[95]	2013	4	674	719	4	2	51	34
	2013	3	711	909	4	2	71	36
	2013	3	837	951	4	2	88	54
	2013	4	1302	1659	4	2	63	14
	2013	2	1317	1562	4	2	136	70
	2013	2	1376	1603	4	2	165	83
	2013	3	1437	1671	4	2	152	83
	2013	3	1440	1613	4	2	135	92
	2013	2	1529	1802	4	2	157	106
	2013	1	1824	1713	4	2	103	53
	2013	1	1882	1868	4	2	394	119
	2013	1	1914	1760	4	2	343	138
	2013	1	1940	1890	4	2	357	144
	2013	1	1963	1849	4	2	222	69
	2013	1	2026	1865	4	2	339	87
	2013	1	10671	11543	4	2	1339	630
CBHSQ, 2017[96]	2017	3	901	1093	4	2	77	49
	2017	2	922	984	4	2	171	101
	2017	3	931	1138	4	2	92	46
	2017	4	948	1013	4	2	74	37
	2017	2	973	954	4	2	184	124

	2017	2	975	1139	4	2	177	98
	2017	1	976	1002	4	2	174	89
	2017	2	1000	1033	4	2	180	79
	2017	2	1006	1155	4	2	189	89
	2017	2	1061	1183	4	2	157	95
	2017	1	1070	1036	4	2	175	86
	2017	1	1329	1269	4	2	89	36
	2017	1	1419	1418	4	2	362	165
	2017	1	1460	1427	4	2	388	108
	2017	1	1492	1385	4	2	263	78
	2017	1	1507	1423	4	2	206	54
	2017	1	1508	1389	4	2	335	148
	2017	3	1945	2387	4	2	232	99
	2017	3	2075	2450	4	2	213	108
	2017	4	2077	2381	4	2	83	46
	2017	2	2117	2580	4	2	312	140
	2017	2	2231	2551	4	2	235	100
	2017	2	2631	3088	4	2	278	174
Coyne et al, 2006[359]	1999		1590	2696	4	2		
	1999		11612	14903	4	2		
Danielson et al, 2005[360]	1995	1	2020	2003	0	2		
Dawson & Grant, 1997[352]	1992		17819	25043	4	2	2003	891
Gavin et al, 2010[361]	2001		1447	1821	4	2	197	100
	2001		2038	3396	4	2	299	104
	2001		2609	3082	4	2	404	211
Goodwin et al, 2004[362]	1995		1492	1540	4	2		
Grant, 1995[363]	1992		17819	25043	4	3	2752	1540
Hasin et al, 2005[97]	2001	4	927	1255	4	2		
	2001	4	1025	1343	4	2		

	2001	3	1236	1611	4	2		
	2001	2	1407	2060	4	2		
	2001	3	1603	2004	4	2		
	2001	2	1834	2458	4	2		
	2001	3	1876	2142	4	2		
	2001	2	1989	2661	4	2		
	2001	3	2034	2406	4	2		
	2001	4	2177	3846	4	2		
	2001	2	2410	2789	4	2		
Kessler et al, 1993[351]	1990	1	1010	990	3	2	37	12
Kessler et al, 1994[364]	1990	2	1231	1207	3	2	110	75
	1990	3	1108	1086	3	2	148	80
	1990	3	740	726	3	2	82	55
Kessler et al, 2010[14]	2001	4	564	894	4	2	146	54
	2001	4	854	1068	4	2	124	67
	2001	3	1342	1522	4	2	167	54
	2001	2	1375	1658	4	2	211	111
Merikangas et al, 2012[103]	2001	1	2147	2003	0	2		
Oquendo et al, 2001[365]	1982		582	582	0	2	51	22
	1982		1832	1832	0	2	62	27
	1982		4936	4936	0	2	207	59
Shah et al, 2011[101]	1991	2	838	927	0	3		
	1991	2	1065	1178	0	3		
	1991	2	1246	1378	0	3		
Toussaint et al, 2008[108]	1998		563	709	4	2		
Verplaetse et al, 2016[353]	2012		15715	20386	4	1	367	141
Zinzow et al, 2009[366]	2005	1	923	885	0	2		
	2005	1	928	868	0	2		

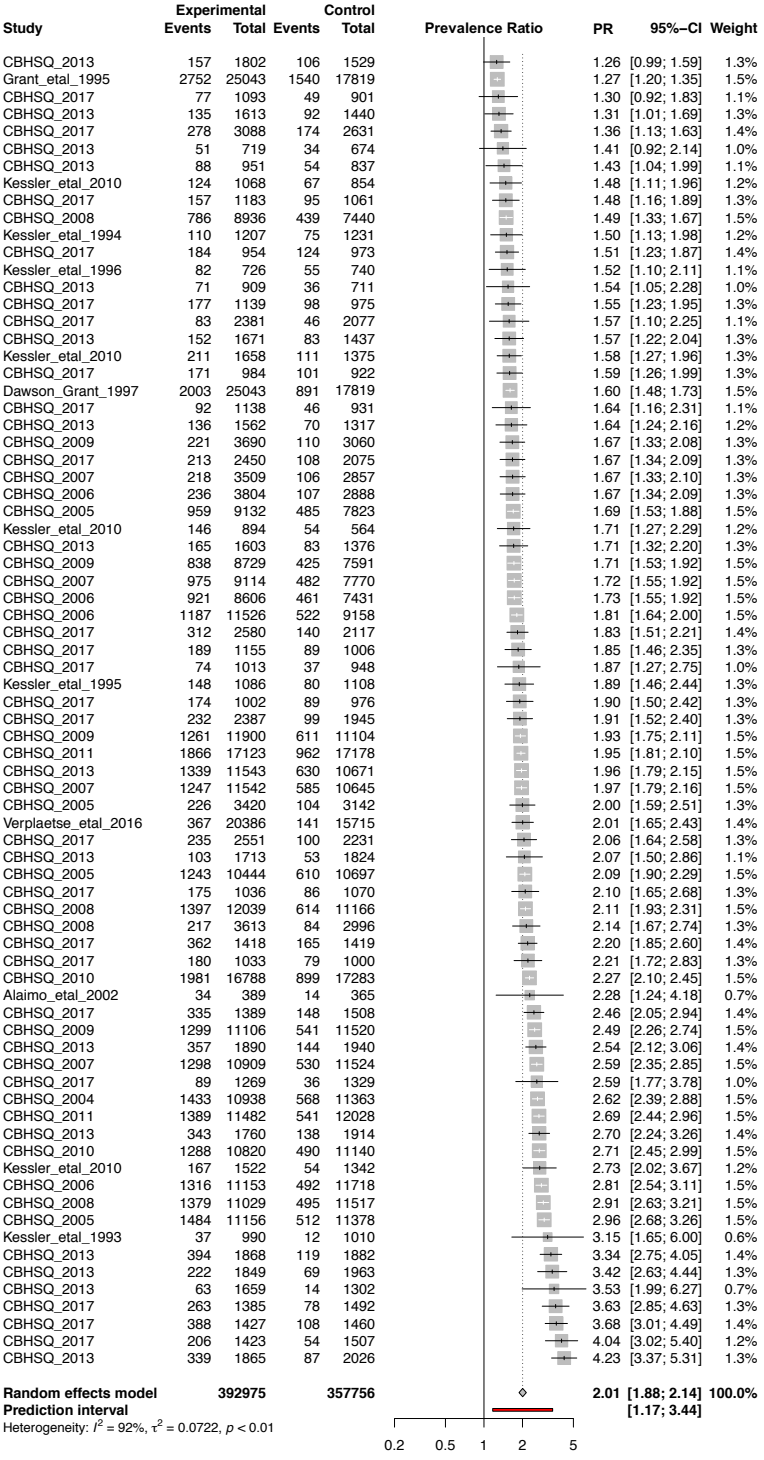
Supplementary Table 1.2. Analytic dataset for symptom-based studies

Author	Study BL year	SMD	SE	Age group	N Men	N Women	Instrument	Symptom period
Bushman et al, 2012[107]	2011	0.08	0.050		251	549	1	1
Everson-Rose et al, 2004[354]	1986	0.23	0.048	4	139	307	1	1
	1986	-0.01	0.021	3	168	222	1	1
	1986	0.23	0.048	3	228	363	1	1
	1986	0.26	0.048	4	239	526	1	1
	1986	0.14	0.050	3	251	434	1	1
	1986	0.31	0.021	2	333	407	1	1
Ferketich et al, 2000[100]	1982	0.26	0.023		2886	5007	1	1
Gettler et al, 2016[99]	2011	0.14	0.042		1505	933	3	1
Guarnaccia et al, 1991[367]	1983	0.23	0.023	2	247	312	1	1
	1983	0.31	0.042	2	416	659	1	1
	1983	0.33	0.050	2	1369	1583	1	1
Hardie et al, 2014[104]	1992	0.23	0.048	2	725	689	1	1
	1992	0.13	0.048	2	1116	1046	1	1
	1992	0.22	0.042	2	1841	1703	1	1
Haroz et al, 2014[117]	2006	0.25	0.142	1	94	106	1	1
	2006	0.16	0.144	1	95	99	1	1
	2006	0.11	0.105	1	192	172	1	1
	2006	0.49	0.102	1	201	191	1	1
	2009	0.47	0.055	1	585	785	1	1
	2009	0.27	0.046	1	856	1096	1	1
	2009	0.22	0.042	1	954	1404	1	1
Henderson et al, 2005[112]	1995	0.18	0.050	2	692	984	1	1
	1995	0.1	0.048	2	830	931	1	1
Inaba et al, 2005[355]	1994	0.38	0.048	3	594	716	1	1
	1994	0.29	0.042	4	856	1220	1	1
	1994	0.22	0.048	3	1013	987	1	1
	1994	0.27	0.048	2	1372	1413	1	1
Margraf et al, 2016[113]	2013	-0.12	0.144		1252	1786	0	1

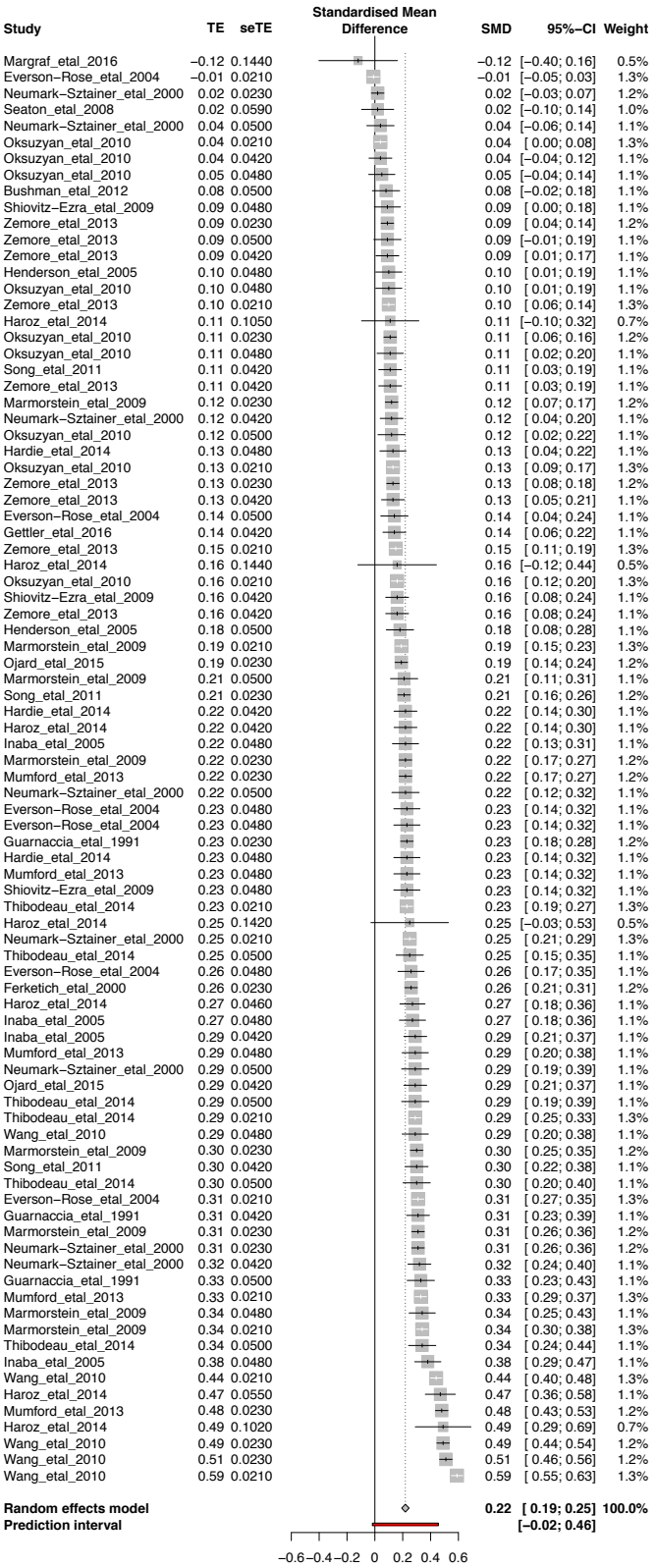
Marmorstein et al, 2009[109]	1995	0.34	0.048	1	237	159	1	1
	1995	0.12	0.023	1	262	329	1	1
	1995	0.22	0.023	1	1039	1218	1	1
	1995	0.3	0.023	1	1319	1472	1	1
	1995	0.21	0.050	1	1512	1427	1	1
	1995	0.34	0.021	1	1778	1883	1	1
	1995	0.19	0.021	1	1981	1940	1	1
	1995	0.31	0.023	1	2061	1991	1	1
Mumford et al, 2013[105]	2000	0.23	0.048	1	657	681	0	1
	2000	0.22	0.023	1	766	767	0	1
	2000	0.29	0.048	1	811	773	0	1
	2000	0.48	0.023	1	815	765	0	1
	2000	0.33	0.021	1	819	774	0	1
Neumark-Sztainer et al, 2000[110]	1997	0.12	0.042	1	239	267	2	1
	1997	0.02	0.023	1	254	305	2	1
	1997	0.25	0.021	1	314	372	2	1
	1997	0.31	0.023	1	361	503	2	1
	1997	0.22	0.050	1	370	462	2	1
	1997	0.04	0.050	1	386	461	2	1
	1997	0.32	0.042	1	399	497	2	1
	1997	0.29	0.050	1	420	484	2	1
Ojard et al, 2015[116]	2005	0.19	0.023		4705	7710	1	1
	2005	0.29	0.042		8751	8802	1	1
Oksuzyan et al, 2010[118]	2006	0.04	0.021	4	142	379	1	1
	2006	0.04	0.042	4	344	649	1	1
	2006	0.13	0.021	3	640	1013	1	1
	2006	0.11	0.023	4	647	917	1	1
	2006	0.16	0.021	4	906	1128	1	1
	2006	0.1	0.048	4	936	1463	1	1
	2006	0.05	0.048	3	1051	1472	1	1
	2006	0.12	0.050	4	1267	1560	1	1
	2006	0.11	0.048	4	1537	1879	1	1

Seaton et al, 2008[368]	2002	0.02	0.059	1	563	605	1	1
Shiovitz-Ezra et al, 2009[356]	2005	0.09	0.048	4	373	499	1	2
	2005	0.23	0.048	4	521	484	1	2
	2005	0.16	0.042	4	543	537	1	2
Song et al, 2011[106]	2005	0.3	0.042		167	188	1	2
	2005	0.21	0.023		187	225	1	2
	2005	0.11	0.042	4	939	1124	1	2
Thibodeau et al, 2014[98]	2008	0.3	0.050	3	391	452	3	1
	2008	0.23	0.021	3	418	400	3	1
	2008	0.34	0.050	2	431	447	3	1
	2008	0.29	0.050	4	434	459	3	1
	2008	0.25	0.050	4	483	482	3	1
	2008	0.29	0.021	2	550	500	3	1
Wang et al, 2010[357]	2006	0.51	0.023	1	721	742	0	1
	2006	0.49	0.023	1	789	997	0	1
	2006	0.59	0.021	1	793	804	0	1
	2006	0.44	0.021	1	892	951	0	1
	2006	0.29	0.048	1	1164	1186	0	1
Zemore et al, 2013[114]	2005	0.09	0.023		383	671	1	1
	2000	0.11	0.042		464	530	1	1
	2000	0.13	0.023		514	847	1	1
	2010	0.1	0.021		517	1078	1	1
	2010	0.09	0.050		517	936	1	1
	2005	0.09	0.042		784	826	1	1
	2005	0.16	0.042		1903	2064	1	1
	2010	0.15	0.021		1904	2695	1	1
	2000	0.13	0.042		2306	2599	1	1

Supplementary Figure 1.1. Forest plot of studies measuring the depression gap as prevalence ratios



Supplementary Figure 1.2. Forest plot of studies measuring the depression gap as standardized mean differences



Supplementary Appendix 2.1. Description of items in the Center for Epidemiologic Studies Depression 7-item Scale

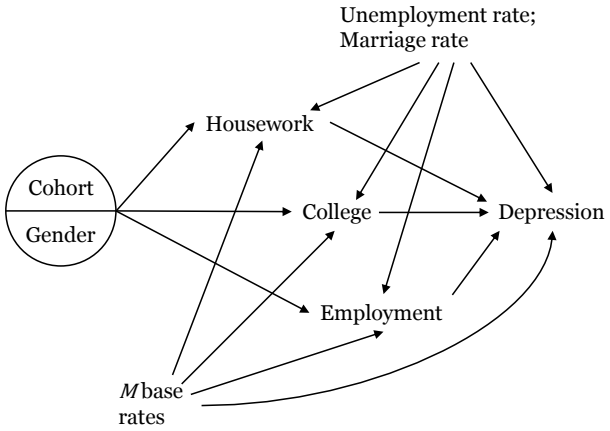
In the past week:

1. I did not feel like eating; my appetite was poor
2. I had trouble keeping my mind on what I was doing
3. I felt depressed
4. I felt that everything I did was an effort
5. My sleep was restless
6. I felt sad
7. I could not get going

Supplementary Table 2.1. Total Number of Interviews in the National Longitudinal Survey of Young Adults by Age (as of December 31st, 2014)

	Number of Young Adult Interviews								
	1	2	3	4	5	6	7	8-9	9-11
Age group									
15-16	214								
17-18	49	308							
19-20	21	56	371						
21-22	26	23	63	521					
23-24	23	31	48	118	550				
25-26	20	31	31	63	111	642			
27-28	15	19	38	49	82	135	587		
29-30	26	27	38	42	55	91	225	488	
31-32	24	18	40	44	77	92	213	388	81
33-34	10	28	30	36	40	78	108	197	355
35-36	6	4	8	13	17	18	31	69	177
>36	24	39	27	28	47	71	91	169	327
Total	458	584	694	914	979	1127	1255	1311	940

Supplementary Figure 2.1. Directed Acyclic Graph depicting the analytic strategy of this study.



Note: M base rates are the gender-specific levels of each mediator, from which the ratios were derived.

Supplementary Table 2.2. Means and Standard Deviations of study variables by gender and cohort

Birth Cohort	1955-1964		1965-1974		1975-1984		1985-1994	
Gender	Men	Women	Men	Women	Men	Women	Men	Women
					Mean (SD)			
Age	36.2 (5.1)	36.2 (5.2)	30.5 (5.6)	30.7 (5.8)	26.1 (5.6)	26 (5.6)	20.7 (3.8)	20.7 (3.7)
Age range	(28, 51)	(28, 51)	(20, 49)	(20, 49)	(14, 39)	(14, 39)	(15, 29)	(15, 29)
CESD score	3.24 (3.79)	4.38 (4.5)	3.93 (3.86)	4.88 (4.58)	4.1 (3.61)	4.95 (4.1)	3.85 (3.43)	4.44 (3.73)
High-CESD symptoms	0.12 (0.33)	0.2 (0.4)	0.14 (0.35)	0.24 (0.43)	0.15 (0.36)	0.22 (0.42)	0.13 (0.34)	0.17 (0.38)
Housework hours/ week	6.9 (2.34)	25 (6.88)	10.1 (0.93)	20.3 (2.57)	9 (0.39)	16.6 (0.28)	9.1 (0.4)	16.6 (0.32)
Housework ratio*		4.41 (2.46)		1.99 (0.11)		1.85 (0.09)		1.84 (0.08)
College degree %	0.26 (0.01)	0.21 (0.01)	0.25 (0.01)	0.26 (0.03)	0.33 (0.02)	0.29 (0.02)	0.29 (0.01)	0.38 (0.01)
College degree ratio*		0.82 (0.06)		1.02 (0.05)		1.23 (0.05)		1.2 (0.02)
Employed %	0.8 (0.02)	0.41 (0.01)	0.74 (0.01)	0.5 (0.03)	0.73 (0.01)	0.58 (0.01)	0.68 (0.01)	0.55 (0)
Employed ratio*		0.6 (0.04)		0.72 (0.03)		0.79 (0.01)		0.82 (0)

SD=Standard deviation; CESD=Center for Epidemiologic Studies Depression scale; *Ratio of women to men; a ratio >1 means women comprise the majority

Supplementary Table 2.3. Significance testing the fit of models with nested polynomial age terms

	Model 1. Linear age	Model 2. M1 + age^2	Model 3. M2 + age^3
Random effects			
Individual intercept	5.68	5.68	5.68
Residual	10.21	10.22	10.21
Model fit statistics			
BIC	236853	236853	236848
Log Likelihood	-118423	-118422	-118418
ANOVA testing model fit vs. model A*			
Chi-squared statistic		1.89	7.13
p-value		0.169	0.008

*null hypothesis=the more parsimonious model is a better fit of the data

Supplementary Table 2.4. Hierarchical mixed model results estimating the risk of high-CESD symptoms for birth cohort, gender, and their interaction

	Model 1*	Model 2*
Fixed-effects parameters	*RD (95% CI)*	
Intercept	-2.83 (-2.95, -2.70)	0.08 (0.065, 0.096)
Cohort (ref=1955)	-0.25 (-0.32, -0.18)	-0.013 (-0.025, -0.002)
Gender (ref=men)	0.63 (0.56, 0.69)	0.075 (0.066, 0.084)
Cohort gender interaction (ref=men)		-0.007 (-0.013, -0.001)
Random-effects variance		
Individual intercept	1.85	0.03
Residual	7.51	0.11
Fit statistics		
BIC	69399	65804
Log likelihood	-34665	-32845

Note: all continuous variables are mean centered; *adjusted for age polynomials and cohort mean age

Supplementary Table 2.5. Hierarchical mixed model results estimating cohort effects in the risk of high-CESD symptoms stratified by gender

	Women*	Men*
Fixed Effects parameters	*RD (95% CI)*	
Intercept	-1.51 (-1.61, -1.41)	-2.27 (-2.38, -2.16)
Cohort (ref=1955)	-0.32 (-0.42, -0.23)	-0.17 (-0.27, -0.08)
Random-effects variance		
Individual intercept	1.85	1.85
Residual	7.63	7.69
Fit statistics		
BIC	38137	31292
Log likelihood	-19042	-15619

Note: all continuous variables are mean centered; * adjusted for age polynomials and cohort mean age

Supplementary Table 3.1. Distribution of industries in the total analytic sample

	n	%
1. Agriculture, Forestry, Fishing, and Hunting	48	0.4
2. Mining	17	0.14
3. Utilities	67	0.56
4. Construction	172	1.44
5. Manufacturing	1044	8.74
6. Wholesale Trade	276	2.31
7. Retail Trade	932	7.8
8. Transportation and Warehousing	423	3.54
9. Information	267	2.23
10. Finance and Insurance	783	6.55
11. Real Estate and Rental and Leasing	186	1.56
12. Professional, Scientific, and Technical Services	500	4.18
13. Management, Administrative and Support, and Waste Management Services	409	3.42
14. Educational Services	1519	12.71
15. Health Care and Social Assistance	2774	23.22
16. Arts, Entertainment, and Recreation	126	1.05
17. Accommodations and Food Services	492	4.12
18. Other Services (Except Public Administration)	366	3.06
19. Public Administration and Active Duty Military	1013	8.48
20. Armed Forces (for CPS)	12	0.1
21. No code	813	6.64

Methodological Appendix

Hierarchical Mixed Modeling

Model specification and model fitting

Hierarchical mixed models allowed for the specification of random intercepts to compare intra- vs. inter-individual CESD scores at baseline, and random slopes to compares CESD score changes over time. To determine the best initial model, I tested the fit of three models with random intercepts estimating the association between age and CESD score. Each model contained an additional age polynomial and model fit was assessed using log likelihood, BIC, and chi-squared test statistics. The best-fitting model contained linear, quadratic and cubic age terms (chisq=7.13, p-value=0.008) (see supplementary table 2.3). This model was then compared with a model that additionally estimated random slopes for individuals. The random slopes variance accounted for less than 1% of the total model variance and thus was deemed not meaningful. The random intercepts accounted for 33-36% of the total model variance in all models. Therefore, all subsequent models only include a random intercept to account for the intra-individual variation in initial CESD scores.

Mediation

The mediation analysis proceeded in two steps. First, we specified two statistical models: the mediator model estimated the conditional distribution of the mediator given the exposure and the set of observed covariates (model 2), and the outcome model estimated the conditional distribution of the outcome given the observed exposure, mediator, and covariates (model 3).

[2] $$\mathrm{M_j} = \gamma_{00} + \gamma_{01}\mathrm{G_j} + \gamma_{02}\mathrm{C_j} + \gamma_{03}\mathrm{cohA_j} + \gamma_{04}\mathrm{GC} + \mathrm{A_{ti}}(\gamma_{10} + \gamma_{11}\mathrm{G_j} + \gamma_{12}\mathrm{C_j} + \mathrm{w_{1i}}) + \mathrm{e_{ti}} + \mathrm{w_{0i}}$$

[3] $$\mathrm{CESD_{ti}} = \mathrm{M_j} + \textstyle\sum_q \gamma_{1q}\,\mathrm{Z_{qi}} + \mathrm{w_{1i}}) + \mathrm{e_{ti}} + \mathrm{w_{0i}}$$

Where G is gender, C is birth cohort, A is age, and cohA is the mean age within each cohort, to account for inter-cohort differences in age. In equation 1, γ_{00} is the estimated CESD score with all covariates at their reference levels, γ_{01} is the mean gender difference in CESD score, γ_{02} is the mean cohort difference in CESD scores, γ_{03} is a covariate to adjust for mean age differences across cohorts, γ_{04} is the between-cohort differences in the gender differences in mean CESD scores, γ_{10} is the expected change in CESD score with age, γ_{11} is the between-cohort gender differences in CESD score changes, γ_{12} is the change with age between cohorts. In addition to the notation described for equation 1, $\mathrm{M_j}$ is one parameter each for education, employment, and housework ratios, and Z is confounders of the mediator-outcome association. Similar models were run for subsamples of women and men. The random coefficients include: $\mathrm{e_{ti}}$ which is the within-individual variation in CESD scores, $\mathrm{w_{0i}}$ is the between-person variation in initial CESD score, and $\mathrm{w_{1i}}$ is the between-individual variation in changes in CESD scores by age, after controlling for cohort differences. All variance terms are assumed $\mathrm{e_{ti}} \sim \mathrm{N}(0,\sigma^2)$, $\mathrm{w_{0i}} \sim \mathrm{N}(0, \mathrm{r_{u0}})$, and $\mathrm{w_{1i}} \sim \mathrm{N}(0, \mathrm{r_{u1}})$. The assumptions necessary for unbiased estimation of mediation models include no residual confounding of either the mediator-outcome pathway and the exposure-outcome pathway. The DAG in Supplementary Figure 2.1 details the covariates that were included in each model.