Title

Abortion and mental health outcomes: A systematic review and meta-analysis.

Permalink

https://escholarship.org/uc/item/4c15g307

Journal

Campbell Systematic Reviews, 20(2)

Authors

Littell, Julia

Young, Sarah

Pigott, Therese

et al.

Publication Date

2024-06-01

DOI

10.1002/cl2.1410

Copyright Information

Peer reviewed

Abortion and mental health outcomes: A systematic review and meta-analysis

Julia H. Littell[1] | Sarah Young[2] | Therese D. Pigott[3] | M. Antonia Biggs[4] | Trine Munk-Olsen[5] | Julia R. Steinberg[6]

[1]Graduate School of Social Work and Social Research, Bryn Mawr College, Bryn Mawr, Pennsylvania, USA

[2]Hunt Library, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[3]College of Education and Human Development, Georgia State University, Atlanta, Georgia, USA

[4]Advancing New Standards in Reproductive Health, Department of Obstetrics, Gynecology & Reproductive Sciences, School of Medicine, University of California San Francisco, Oakland, California, USA

[5]Department of Psychiatry, University of Southern Denmark, Odense, Denmark

[6]Department of Family Science, School of Public Health, University of Maryland, College Park, Maryland, USA

Correspondence

Julia H. Littell, Graduate School of Social Work and Social Research, Bryn Mawr College, Bryn Mawr, PA, USA.
Email: jlittell@brynmawr.edu

Abstract

This is a protocol for a systematic review and meta-analysis of research on mental health outcomes of abortion. Does abortion increase the risk of adverse mental health outcomes? That is the central question for this review. Our review aims to inform policy and practice by locating, critically appraising, and synthesizing empirical evidence on associations between abortion and subsequent mental health outcomes. Given the controversies surrounding this topic and the complex social, political, legal, and ideological contexts in which research and reviews on abortion are conducted, it is especially important to conduct this systematic review and meta-analysis with comprehensive, rigorous, unbiased, and transparent methods. We will include a variety of study designs to enhance understanding of studies' methodological strengths and weaknesses and to identify potential explanations for conflicting results. We will follow open science principles, providing access to our methods, measures, and results, and making data available for re-analysis.

KEYWORDS

abortion, mental health, meta-analysis, systematic review

1 | BACKGROUND

1.1 | The problem, condition, or issue

Abortion, the voluntary termination of a pregnancy, has long been a subject of controversy. Ideological and legal debates on this issue reflect different views about when human life begins and the rights and autonomy of women versus the rights of an embryo or fetus. In recent decades, a newer controversy has centered on whether abortion has adverse effects on women's mental health (Lee, 2003; Major et al., 2009; Siegel, 2008). In 1989, a review by the U.S. Surgeon General concluded that the quality of available evidence was weak and "the data do not support the premise that abortion does or does not cause or contribute to psychological problems" (Koop, 1989, p. 174). Since then, many studies on this topic have been completed; these studies use divergent methodologies and reach different conclusions. Previous reviews of this body of research have not settled the controversy.

Claims that abortion is harmful to women's mental health have influenced policies around the globe, including abortion bans, gestational limits, mandated waiting periods, and counseling laws. Although these claims are contested, they have influenced recent landmark court decisions and legislation to restrict access to abortion, particularly in the USA. These claims appeared in arguments before

the US Supreme Court in the 2022 case (*Dobbs v Jackson Women's Health Organization*) that eliminated the constitutional right to abortion in the USA, and in a 2023 US District Court ruling (Kacsmaryk, 2023) that sought to overturn the US Food and Drug Administration's 2000 approval of mifepristone, one of the most widely used abortion medications. In the USA, some states require health care providers to inform patients that abortion has adverse mental health consequences (National Academies of Sciences, Engineering, and Medicine [NASEM], 2018).

Does abortion increase the risk of adverse mental health outcomes? That is the central question for this systematic review and meta-analysis. Our review aims to inform policy and practice, by locating, critically appraising, and synthesizing empirical evidence on associations between abortion and subsequent mental health outcomes. Given the controversies surrounding this topic and the complex social, political, legal, and ideological contexts in which research and reviews on abortion are conducted, it is especially important to conduct this systematic review and meta-analysis with rigorous, unbiased, and transparent methods. We will include a variety of study designs to enhance understanding of studies' methodological strengths and weaknesses and to identify potential explanations for conflicting results.

Throughout this protocol, we use the terms "women" and "persons" somewhat interchangeably, to refer to females, cisgender women, transgender men, and gender non-binary individuals who are capable of pregnancy and childbirth.

Because the vast majority of abortions result from unwanted or mistimed pregnancies (Kost et al., 2023), we begin with a brief discussion of pregnancy intentions and abortion. Then we describe legal restrictions on abortion.

1.1.1 | Pregnancy intentions and abortion

Unintended pregnancies are usually defined to include unplanned, mistimed, and unwanted pregnancies. The best available data suggest that almost half (48%) of all pregnancies worldwide are unintended and 61% of unintended pregnancies end in abortion, although rates vary considerably by region and demographic characteristics (Bearak et al., 2020, 2022). Unintended pregnancy and abortion rates tend to be higher in low- and middle-income countries and among people living in poverty in high income countries (Bearak et al., 2020; Finer & Zolna, 2016).

Unintended pregnancy is associated with an array of negative health, economic, social, and psychological outcomes for women and children (National Collaborating Centre for Mental Health [NCCMH], 2011; Nelson et al., 2022). Until recently, the incidence of unintended pregnancy has been treated as one of the "most essential health status indicators in the field of reproductive health" (Finer & Zolna, 2011, p. 478).

Measures of unintended pregnancy are usually binary and retrospective; they do not capture the complexity of pregnancy desires (including uncertainty and ambivalence) and they conflate pregnancy planning with wanted pregnancies (i.e., unplanned and mistimed pregnancies may or may not be wanted; Auerbach et al., 2023, Kost et al., 2023). For purposes of research on abortion, information on unwanted and mistimed pregnancies is more relevant than data on pregnancy planning or composite measures of pregnancy intentions. Nevertheless, many studies use imperfect measures of pregnancy intentions.

1.1.2 | Legal restrictions on abortion

Access to safe and legal abortion varies widely by policy context and whether abortion is banned, highly restricted, or stigmatized (Ganatra et al., 2017). Legal restrictions on abortion vary across jurisdictions and are often based on gestational duration and the reason for abortion. In some jurisdictions, abortions are only allowed to preserve the pregnant person's life or health (sometimes including mental health), for socioeconomic reasons, or in cases of rape, incest, or fetal anomaly (Remez et al., 2020). Some jurisdictions require parental consent for minors, require abortion procedures to be performed by a licensed physician, require involvement of a second physician, require abortions to be performed in a hospital after a certain gestational age, require waiting periods, and so forth. In some places, providers are mandated to provide counseling before an abortion, and sometimes this includes warning patients of physical and mental health consequences of abortion (National Academies of Sciences, Engineering, & and Medicine [NASEM], 2018).

The global trend toward reducing legal restrictions and expanding legal access to abortion continued over the past decade (Remez et al., 2020), but legal restrictions on abortion have increased in the USA for at least the past 10 years (Guttmacher Institute, 2021). Laws that restrict abortion reduce but do not eliminate its practice (Remez et al., 2020).

1.2 | The intervention: Abortion

When abortion is the result of an intervention, it is sometimes called "induced abortion." In contrast, miscarriage or pregnancy loss (sometimes referred to as "spontaneous abortion") occurs without any intervention in approximately 15% of all recognized pregnancies (Quenby et al., 2021). We use the term "abortion" to refer to any voluntary termination of pregnancy and not to miscarriage or pregnancy loss.

Abortion is safe and effective when carried out by methods recommended by the World Health Organization (Lohr et al., 2014; National Academies of Sciences, Engineering, and Medicine [NASEM], 2018; World Health Organization, 2022, 2023), including use of medications approved for this purpose (mifepristone and/or misoprostol, and methotrexate) and/or approved medical procedures such as uterine aspiration, dilation and curettage, dilation and evacuation, or dilation and extraction). The former are termed medication abortions and the later are called procedural abortions (Upadhyay et al., 2023). Both types of abortion are included in this review.

Self-managed abortions can be very safe and effective with the use of misoprostol or mifepristone medications (Moseson et al., 2020, 2022, 2023), but the same cannot be said for self-managed abortions that use physical force or other (potentially harmful) substances (Ralph et al., 2020; Remez et al., 2020). Therefore, this review will include studies of abortions that are self-managed with mifepristone and/or misoprostol, but it will not include studies of other self-managed or non-medical abortion procedures.

Safe and effective abortion methods depend, in part, on gestational duration. Thus, womens' experiences of abortion are likely to be affected by abortion method and gestational duration, along with characteristics of the socioeconomic, cultural, legal, health care, and relational contexts in which abortions occur (Russo, 2014). In high-income countries at least half of all abortions are medication abortions, at least 90% of all abortion occur within the first 13 weeks of pregnancy, and two-thirds occur within 9 weeks (Popinchalk & Sedgh, 2019). Global data on gestational duration at abortion are scarce.

Pregnant persons who do not have medication or procedural abortions may experience live births (full-term or premature deliveries) or fetal loss (miscarriages or stillbirths), or may end the pregnancy on their own without the use of approved medications or the support of a health professional.

1.3 | How and why abortion might relate to mental health outcomes

As illustrated in Figure 1, there are at least two competing explanations of how and why abortion might be related to subsequent mental health outcomes. We refer to the first as abortion-trauma theory and the second as a co-occurring risks perspective.

1.3.1 | Abortion-trauma theory and "post-abortion syndrome"

One theory is that "abortion is a uniquely traumatic experience because it involves a human death experience, specifically, the intentional destruction of one's unborn child and the witnessing of a violent death, as well as a violation of parental instinct and responsibility, the severing of maternal attachments to the unborn child, and unacknowledged grief" (Major et al., 2009, p. 866). Rue and Speckhard (Rue & Speckhard, 1992; Rue, 2014; Speckhard & Rue, 1992; Speckhard, 1985) suggested that the traumatic experience of abortion could lead to serious mental health problems, which they termed "post-abortion syndrome" (PAS). PAS was conceptualized as a specific form of posttraumatic stress disorder (PTSD) "comparable to the symptoms experienced by Vietnam veterans, including symptoms of trauma, such as flashbacks and denial, and symptoms such as depression, grief, anger, shame, survivor guilt, and substance abuse" (Major et al., 2009, p. 866). PAS is not recognized as a mental disorder in the DSM5 or ICD-10.

The abortion-trauma theory was originally developed from Speckhard's (1985) interviews with 30 women who had "highly stressful" abortions in the past 1 to 25 years, including some who had illegal abortions, 14 women who had second trimester abortions, and 1 who had an abortion in the third trimester (Major et al., 2009). As indicated above, some subsequent studies seemed to support this view. For example:

- lifetime measures of abortion were correlated with lifetime measures of mental health problems (Coleman et al., 2009; Steinberg & Finer, 2011);
- lifetime history of abortion or miscarriage were correlated with lifetime measures of substance use disorders and affective disorders in 21 year-old women (Dingle et al., 2008);
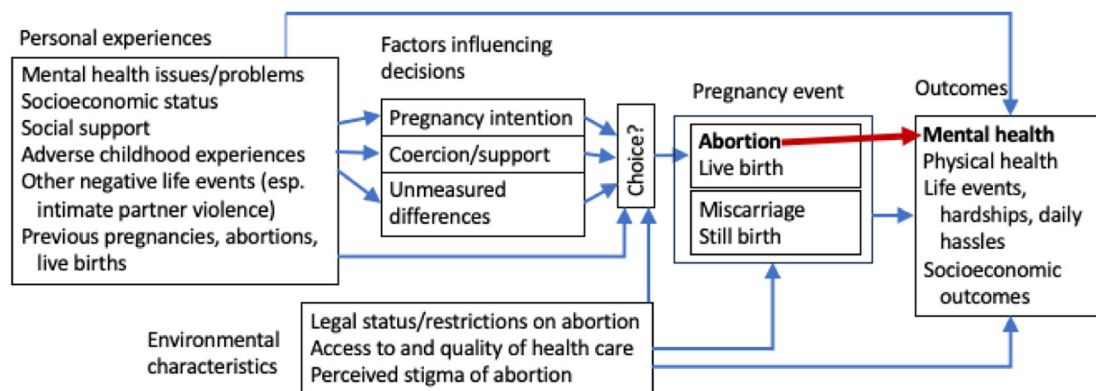

FIGURE 1 Abortion and mental health outcomes: two theories of change. Theory 1 (illustrated with a red arrow): Abortion is a traumatic event that increases risks of negative mental health outcomes. Theory 2 (illustrated with blue arrows): Abortion is confounded with other factors that affect mental health. Pregnancy events (including but not limited to abortion) may moderate or mediate influences of personal experiences and environmental characteristics on outcomes. Abortion does not have consistent, independent (direct) effects on mental health outcomes.

- after adjustment for some confounding variables (but not whether pregnancy was wanted or mistimed), abortion was associated with an increase in the risk of mental disorders (Fergusson et al., 2006, 2008); and
- one meta-analysis estimated that "women who had undergone an abortion experienced an 81% increased risk of mental health problems, and nearly 10% of the incidence of mental health problems was shown to be directly attributable to abortion" (Coleman, 2011, p. 183).

Abortion-trauma theory has been vigorously disputed by Stotland, 1992 and others (Charles et al., 2008; Major et al., 2009; National Academies of Sciences, Engineering, &and Medicine [NA-SEM], 2018; National Collaborating Centre for Mental Health [NCCMH], 2011; Steinberg & Russo, 2009) who noted that the evidence used to support a causal link between abortion and subsequent mental health disorders or symptoms is marred by serious methodological problems including use of highly selected samples, inadequate comparison groups, inadequate conceptualization and control of relevant variables, unreliable outcome measures, inappropriate statistical analyses, and errors of interpretation including misattribution of causal effects (Robinson et al., 2009; Steinberg & Finer, 2011).

Some evidence seems to contradict the abortion-trauma view. For example, a prospective, multi-site cohort study in Sweden found that few women developed PTSD or Posttraumatic Stress Symptom (PTSS) after abortion, and most cases of PTSD or PTSS were due to trauma experiences unrelated to induced abortion (Wallin Lundell et al., 2013). A recent, prospective longitudinal cohort study of patients who sought abortion at 30 abortion clinics in the USA (The Turnaway Study) showed that those who received abortions were at no higher risk of PTSD than patients who were denied abortions (Biggs et al., 2016); further, there were no long-term differences in perceived stress, although patients who were denied abortions experienced more perceived stress immediately afterward than did patients who had abortions (Harris et al., 2014).

1.3.2 | Co-occurring risks: A psychological and sociological perspective

A second explanation for how and why abortion might relate to mental health outcomes is that abortion is confounded with other factors that impact mental health, but abortion per se does not have an independent effect on mental health outcomes. This view is derived from three complementary conceptual frameworks, which Major et al. (2009) identified as the stress and coping, sociocultural, and co-occurring risk perspectives (we include all three within a broader co-occurring risks perspective).

The stress-and-coping perspective suggests that abortion is one of many stressful life events, such as childbirth, but abortion co-occurs with other potentially stressful events, such as unintended or unwanted pregnancy. Adaptations to such compounded, stressful events are highly varied and are likely influenced by the person's cognitive appraisal of their situation, their perceived ability to cope with it, and available social support (Major et al., 2009).

The sociocultural perspective suggests that the perceived stigma of experiences like abortion can lead to negative psychological outcomes. Perceived stigma and "internalized stigma" (shame and guilt) have been linked to a variety of cognitive, emotional, and behavioral problems (see Major et al., 2009). Further, societal messages about what women should do regarding unintended or unwanted pregnancies (the extent to which abortion is stigmatized or normalized) or should experience regarding abortion (e.g., relief or shame and guilt) can become self-fulfilling expectations. Thus, social stigma, internalized stigma, and sociocultural norms can have important mediating roles in the relationship between abortion and mental health outcomes.

Finally, a number of potent risk factors for mental health problems also increase the likelihood of unwanted pregnancy, and thus the likelihood of desiring an abortion. As mentioned above, unintended pregnancy is associated with an array of negative health, economic, social, and psychological outcomes for women and children (National Collaborating Centre for Mental Health [NCCMH], 2011; Nelson et al., 2022). Thus, "systemic, social, and personal factors that are precursors to unintended pregnancy ... place women at risk for having abortions and/or predispose them to experience mental health problems regardless of pregnancy and its resolution" (Major et al., 2009, p. 868). Epidemiological studies show that women who have abortions are disproportionately at risk of interpersonal violence and other types of violence (National Academies of Sciences, Engineering, and Medicine [NASEM], 2018).

Other confounding variables include socioeconomic status, race and ethnicity, smoking, and substance use (National Academies of Sciences, Engineering, and Medicine [NASEM], 2018). "Thus, the myriad of disadvantages, such as low socioeconomic status, lack of education, and violence, putting women at risk for unintended pregnancy are also associated with having an abortion" (Steinberg & Russo, 2009, p. 500). It may be the co-occurring risk factors, not abortion per se, that account for subsequent mental health outcomes. Indeed, one review found that:

- rates of mental health problems for women with an unwanted pregnancy were the same regardless of whether they had an abortion or gave birth,
- the most reliable predictor of post-abortion mental health problems was having a history of mental health problems before the abortion, and
- factors associated with increased rates of mental health problems for women in the general population following birth and following abortion were similar (National Collaborating Centre for Mental Health [NCCMH], 2011).

As depicted in Figure 1, pregnancy, pregnancy-related events, and other stressful life events can have diverse direct and/or indirect effects on mental health outcomes (e.g., postpartum depression) as well as outcomes related to patients' physical health, daily hardships,

stressors, and socioeconomic variables. When background and contextual factors are not adequately taken into account, abortion may act as a proxy for exposure to other risk factors (e.g., socioeconomic disadvantage, prior mental health issues, social stigma, exposure to violence including sexual assault), leading to spurious associations between abortion and mental health outcomes.

Russo noted that, although the predictors of problems after abortion or childbirth are similar, "it is reasonable to assume that there is something different about women who have unwanted pregnancies and opt for abortion over delivery" (Russo, 2014, p. 285). Unmeasured differences between these groups may be important.

Pregnancy and related events may mediate or moderate effects of personal background characteristics and environmental conditions on mental health outcomes. For example, pregnancy and childbirth or abortion might exacerbate or ease ongoing mental health issues.

It is reasonable to assume that there are complex interactions and reciprocal effects among predictors and outcome variables in this conceptual model, so that mental health outcomes affect and are affected by other outcomes. Further, the direction of influence between abortion and mental health is not always clear. Some studies show that mental health problems tend to peak immediately before an abortion, not afterward (Major et al., 2000). One study found that, compared with those who give birth, patients who had an abortion reported higher rates of mood disorders before the procedure (Steinberg et al., 2014). "Therapeutic abortion" is sometimes provided to preserve patients' mental health in cases with "psychological distress or mental suffering due to unwanted pregnancy and responsibility for childcare, or … anticipated serious fetal impairment" (Cook et al., 2006, p. 185). Thus, abortion can be viewed as both a predictor and an outcome of mental health problems.

In the overarching co-occurring risks model (which incorporates the stress and coping and sociocultural perspectives) shown in Figure 1, (1) confounding variables play an important role in understanding the relationship between abortion and mental health outcomes, and (2) pregnancy events may moderate or mediate effects of personal and environmental characteristics on outcomes, but (3) abortion itself does not have consistent or independent (direct) effects on mental health outcomes. This theory predicts that having an abortion does not increase a patient's risk of problems such as depression, anxiety, and/or PTSD (National Academies of Sciences, Engineering, and Medicine [NASEM], 2018). Criticisms of this approach might focus on its complexity.

1.4 | Why it is important to do this review

As described above, there are conflicting views about the relationship between abortion and subsequent mental health outcomes. Conflicting information has been provided to judicial and legislative bodies that are responsible for legal regulations on abortion, as well as to patients in some nations and states. A rigorous systematic review and meta-analysis is needed to provide a credible, up-to-date synthesis of research on this topic.

1.4.1 | Primary studies of abortion and mental health outcome

Many studies have examined associations between abortion and risks of post-traumatic stress and other anxiety disorders, depression and other mood disorders, suicidal ideation and suicidal behavior, substance use disorders, and/or a composite measure of any mental disorder (e.g., Biggs et al., 2015, 2016, 2017; Coleman et al., 2009; Crandell, 2012; Fergusson et al., 2006, 2008; Foster, 2020; Gissler et al., 1996; Munk-Olsen et al., 2011; Pedersen, 2007, 2008; Van Ditzhuijzen et al., 2017; Wallin Lundell et al., 2013). A few studies have looked at abortion in relation to eating disorders (Linna et al., 2013), sleep disorders (Reardon & Coleman, 2006), sexual dysfunction (Coleman et al., 2008), and psychotic disorders (Brewer, 1977; David et al., 1981; Gilchrist et al., 1995). Abortion has also been studied in relation to rates of outpatient mental health treatment (Coleman et al., 2002; Studnicki et al., 2023), psychiatric hospitalization (Reardon et al., 2003; Studnicki et al., 2023), and reported use of psychotropic medication for mental disorders (Steinberg et al., 2018).

Studies with parallel cohorts have included different types of comparison groups (sometimes called reference groups or control groups), including: people who had miscarriages (Broen et al., 2005; Dingle et al., 2008), people who sought but did not receive abortions (Biggs et al., 2017; Foster, 2020); people who had unintended pregnancies but carried them to term (Gilchrist et al., 1995); people who had live births (David et al., 1981; Studnicki et al., 2023), and birth cohorts from a general population including people who were never pregnant (Munk-Olsen et al., 2011; Pedersen, 2007, 2008; Van Ditzhuijzen et al., 2017).

1.4.2 | Previous reviews of research on this topic

Our preliminary search identified several nonsystematic, narrative reviews of research on abortion and women's mental health (American Psychological Association [APA] Task Force on Mental Health and Abortion, 2008; Coleman et al., 2005; Russo, 2014; Major et al., 2009; National Academies of Sciences, Engineering, and Medicine [NASEM], 2018; Robinson et al., 2009; Steinberg et al., 2009; Steinberg, 2011; Grupo Médico por el Derecho a Decidir – Columbia, 2011), as well as five systematic reviews and/or meta-analyses on this topic (Bellieni & Buonocore, 2013; Charles et al., 2008; Coleman, 2011; Fergusson et al., 2013; National Collaborating Centre for Mental Health [NCCMH], 2011).

In Table 1, we assess the adequacy of previous systematic reviews and meta-analyses, using the AMSTAR tool (for Assessment of Multiple SysTemAtic Reviews; Shea et al., 2007). We use the version of AMSTAR that was already available when these reviews were published (a latter version, AMSTAR 2, includes additional criteria that none of these reviews addressed; Shea et al., 2017). Our analysis of these reviews is based solely upon public *a priori* protocols (if available) and published reports.

TABLE 1 Assessment of previous systematic reviews and meta-analyses (adapted from AMSTAR; Shea et al., 2007).

	Charles et al. (2008)	Coleman (2011)	NCCMH (2011)	Fergusson et al. (2013)	Bellieni and Buonocore (2013)
Number of studies included	21 studies with comparison groups, 5 without comparison groups	22 studies (publications) based on 14 distinct databases	15 studies in section on comparative risks	4 studies (14 analyses, 8 publications; reanalysis of data from Coleman (2011) and NCCMH (2011)	30 studies with comparison groups
1. Was an *a priori* design provided? The research question and inclusion criteria should be established before the conduct of the review.	No	No	No	No	No
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.	Yes	No	Yes	No	No
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Keywords and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing references in studies found.	No	No	Yes	No	No
4. Was the status of publication (i.e., grey literature) avoided as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language, and so forth.	No (limited to published studies)	No (limited to studies published in English in peer-reviewed journals, 1995–2009)	Yes (searched for unpublished studies via contacts with experts; limited to English reports, 1990–2011)	No	No (limited to studies published 1995–2011)

TABLE 1 (Continued)

	Charles et al. (2008)	Coleman (2011)	NCCMH (2011)	Fergusson et al. (2013)	Bellieni and Buonocore (2013)
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided.	No (excluded studies not listed)	No	Yes	No	Yes
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions, and outcomes. The ranges of characteristics in all the studies analyzed, for example, age, race, sex, relevant socioeconomic data should be reported.	Yes	No	Yes	Yes	Yes
7. Was the scientific quality of the included studies assessed and documented? A priori methods of assessment should be provided.	Unclear (overall quality ratings were not documented)[a]	No	Yes (use of 3 NICE methodology checklists, overall quality ratings,[a] and GRADE)	Unclear (used overall study quality ratings from NCCMH review)[a]	Unclear (use of an overall design rating)
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.	No (use of overall quality ratings and vote counting)	No	Yes (used GRADE)	Yes	No
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to assess their homogeneity (i.e., Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?).	No (used vote counting)	No (meta-analysis included dependent effect sizes from overlapping samples)	Unclear (heterogeneity tests were used to determine whether meta-analysis was appropriate)	Unclear (used vote-counting in addition to meta-analysis)	No (used vote-counting)

(Continues)

TABLE 1 (Continued)

	Charles et al. (2008)	Coleman (2011)	NCCMH (2011)	Fergusson et al. (2013)	Bellieni and Buonocore (2013)
10. Was the likelihood of publication bias assessed? Assessment of publication bias should include a combination of graphical aids (e.g., funnel plot) and/or statistical tests (e.g., Egger regression test).	No	No	No	No	No
11. Was the conflict of interest (COI) stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.	**Partially** (support for the review was acknowledged, but there was no COI statement and COI in included studies was not reported)	No	**Partially** (interests of reviewers were documented, but COI in included studies was not)	**Partially** (support for the review was acknowledged, authors declared they had no conflicts, but they did not document COI in included studies)	**No** (authors reported no COI, but sources of support were not reported for the review or for included studies)
Reviewers' conclusions	"A clear trend emerges from this systematic review: the highest quality studies had findings that were mostly neutral, suggesting few, if any, differences between women who had abortions and their respective comparison groups in terms of mental health sequelae. Conversely, studies with the most flawed methodology found negative mental health sequelae of abortion" (p. 436).	"Women who had undergone an abortion experienced an 81% increased risk of mental health problems, and nearly 10% of the incidence of mental health problems was shown to be directly attributable to abortion" (p. 183).	"The rates of mental health problems for women with an unwanted pregnancy were the same whether they had an abortion or gave birth. An unwanted pregnancy was associated with an increased risk of mental health problems. The most reliable predictor of post-abortion mental health problems was having a history of mental health problems before the abortion. The factors associated with increased rates of mental health problems for women in the general population following birth and following abortion were similar. There were some additional factors associated with an increased risk of mental health problems specifically related to abortion, such as pressure from a partner to have an	"There is no available evidence to suggest that abortion has therapeutic effects in reducing the mental health risks of unwanted or unintended pregnancy. There is suggestive evidence that abortion may be associated with small to moderate increases in risks of some mental health problems" (p. 819).	"Fetal loss seems to expose women to a higher risk for mental disorders than childbirth; some studies show that abortion can be considered a more relevant risk factor than miscarriage..." (p. 301).

None of these reviews followed (or cited) any of the published guidelines for the conduct and reporting of systematic reviews and meta-analyses that were available at the time (e.g., Institute of Medicine, 2011; Moher et al., 1999, 2009; Stroup, 2000), nor did they reflect (or cite) the substantial body of theoretical, statistical, and methodological work upon which these guidelines were based (see Cooper & Hedges, 1994; Cooper et al., 2009; Hedges & Olkin, 1985; Higgins & Green, 2011; Light & Pillemer, 1984; Lipsey & Wilson, 2001; Littell et al., 2008; Petticrew & Roberts, 2006; Rothstein et al., 2005). Instead, most reviews cited prior reviews, a common practice in many fields of research, which leads to repetition of old practices, not necessarily best practices. Thus, these reviews did not reflect major developments in research synthesis methods through the early 2000s (or later), including efforts to discourage the use of publication status as a criterion for inclusion in a review (in light of evidence of systematic reporting, publication, and dissemination biases; Rothstein et al., 2005; Song et al., 2000, 2009, 2010; Stroup, 2000) and concerns about the use of overall study quality scores or ratings, which conflate potentially unrelated sources and types of biases (Jüni, 1999, 2001; Stroup, 2000; Wells & Littell, 2009). Indeed, several reviews systematically excluded unpublished studies and used overall study quality ratings (see Table 1).

As shown in Table 1, none of these reviews provided a public protocol in advance of conduct of the review to improve transparency and reduce bias (see Stewart et al., 2012). None of these reviews assessed the likelihood of publication bias or its potential impact on results (see Rothstein et al., 2005).

Only one review (National Collaborating Centre for Mental Health [NCCMH], 2011) conducted a comprehensive literature search (one that could retrieve relevant unpublished studies) and avoided systematic exclusion of unpublished studies. The other reviews shown in Table 1 used convenience samples of published studies, which are likely to produce biased results (Song et al., 2000, 2009, 2010).

The review conducted by the UK National Collaborating Centre for Mental Health (National Collaborating Centre for Mental Health [NCCMH], 2011) at the Royal College of Psychiatrists (RCP), underutilized meta-analysis. It relied on heterogeneity tests to determine whether meta-analysis was "appropriate," instead of using random effects models for heterogeneous sets of studies (see Borenstein et al., 2009, 2010). As a result, this review overutilized narrative synthesis, which is demonstrably less accurate than meta-analysis (Bushman & Wells, 2001; Cooper & Rosenthal, 1980).

Only two reviews (Charles et al., 2008; National Collaborating Centre for Mental Health [NCCMH], 2011) conducted duplicate study selection and data extraction, which can reduce bias and error (Buscemi et al., 2006; Gøtzsche et al., 2007).

Several reviews used overall study quality ratings that included undefined criteria (e.g., "appropriate" comparison groups) and "other methodological factors" that were "not … uniformly considered across studies" (Charles et al., 2008, p. 437). This approach was adopted in the National Collaborating Centre for Mental Health (NCCMH) (2011) review (alongside other critical appraisal methods) and overall quality

Charles et al. (2008)	Coleman (2011)	NCCMH (2011)	Fergusson et al. (2013)	Bellieni and Buonocore (2013)
		abortion and negative attitudes towards abortions in general and towards a woman's personal experience of the abortion" (p. 8).		

TABLE 1 (Continued)

aA priori methods for study quality assessment were not used. Overall quality ratings included categories that were not clearly defined (e.g., "appropriate" comparison groups) and "other methodological factors" that were "not … uniformly considered across studies" (Charles et al., 2008, p. 437).

ratings from the National Collaborating Centre for Mental Health (NCCMH) (2011) review were used in the Fergusson et al. (2013) review. As indicated above, overall quality scores conflate distinct sources and types of bias, which may have different effects on results; thus, these scores lack construct and discriminant validity.

As shown in Table 1, several reviews used vote-counting methods, which tally the number of studies according to the direction and statistical significance of their results. This is not a valid synthesis method, because such tallies are imprecise (uninformed by the magnitude of the effect size, sample size, or heterogeneity) and potentially misleading (Bushman, 1994; Grainger et al., 2022; Hedges & Olkin, 1985). Charles et al. (2008) used a vote-counting procedure to tally studies according to (a) the direction and significance of their findings and (b) overall quality ratings, thereby combining an invalid synthesis method (vote-counting) with unreliable and invalid ratings.

The Coleman (2011) meta-analysis (cited above) did not meet any of the AMSTAR criteria (Littell & Coyne, 2012). Further, it included multiple dependent effect sizes derived from the same data sets, and synthesized these effects as if they were independent estimates. This approach produces results that are not reliable, accurate, or meaningful.

Fergusson et al. (2013) reanalyzed data from the Coleman (2011) and National Collaborating Centre for Mental Health (NCCMH) (2011) reviews, limiting meta-analysis to effect sizes from non-overlapping samples (to manage dependent effect sizes properly).

Bellieni and Buonocore (2013) provided separate (vote-counting) analyses which compared abortion with different alternatives: child-birth, unplanned childbirth, or miscarriage.

Table 1 shows that these five reviews drew substantially different conclusions from research on associations between abortion and subsequent mental health outcomes. Some reviews concluded that abortion increased the risks of subsequent mental health problems, while others found that these problems were associated with unintended pregnancy and prior mental health problems, not abortion per se, and a third set of conclusions suggests that results vary based on study quality. In any case, serious methodological flaws in each of these reviews undermines confidence in their conclusions.

There are no reviews on this topic that meet current published guidelines for the conduct of systematic reviews and meta-analysis (e.g., Page et al., 2021). A rigorous systematic review and meta-analysis is needed to provide robust and reliable critical appraisal, analyses, and syntheses of the empirical evidence on the mental health outcomes of abortion. Our review aims to meet that need.

1.4.3 | Relevance for policy and practice

Better understanding of associations between abortion and mental health outcomes could inform policy and practice regarding screening and treatment for mental health problems in relation to reproductive health care.

1.5 | Our approach: Guiding principles and assumptions

To produce credible evidence on this controversial topic, our review must be as rigorous and transparent as possible. To ensure rigor, we will follow the Campbell Collaboration's standards for the conduct and

| Study | Time order | MH history | Pregn intent | Adequate design or statistical control for potential confounders (between-group differences) |||||||||||
| | | | | Socioeconomic status ||| Family background || Participant |||| Other ||
				Income	Educ	Employ	ACES	Exposure to Interpersonal violence	Age	Prior preg	Prior births	Race ethn	Service, care, attention	Perceived stigma
Study 1	+++	+++	+++	+++	?	?	+++	+++	+++	+++	+++	+++	+++	+++
Study 2	+++	+++	+++	?	?	?	?	+++	+++	+++	+++	+++	?	—
Study 3	+++	—	+++	+++	+++	+++	?	—	+++	+++	—	+++	+++	—
Study 4	?	?	?	—	—	+++	—	+++	+++	—	+++	+++	?	+++
Study 5	—	—	—	—	—	?	?	—	+++	—	—	+++	?	—
Study 6	—	—	—	?	?	?	?	?	?	?	?	?	?	?

+++ = adequate control, low risk of bias

? = insufficient information, unclear risk of bias

— = inadequate control, high risk

MH history = Mental health history; Pregn intent = Pregnancy intention; ACES = Adverse Childhood Experiences

FIGURE 2 Sample confounder matrix. ACES, Adverse Childhood Experiences; MH history, Mental health history; Pregn intent, Pregnancy intention. *Source*: adapted from Petersen et al., 2022.

reporting of systematic reviews and meta-analyses (The Methods Coordinating Group of the Campbell Collaboration, 2019a, 2019b), rely on empirical evidence on the best ways to minimize bias and error in systematic reviews and meta-analysis, and utilize the best available review methodologies. To ensure transparency, we will clearly define key constructs and measures. We will carefully document our procedures, including any changes in plans. We will follow open science principles, providing access to our methods, measures, and results on the website: https://osf.io/g6bju/. To minimize bias and error, we have developed and will follow steps (described in Section 3 below) that have been shown to reduce bias and error in research reviews; these plans include steps to report and manage any potential conflicts of interest.

In constructing this review, we have made use of the logic of causal inference and understanding of various counterfactual models (Shadish et al., 2002). For reasons described above, it is difficult to assess the effects of abortion on mental health outcomes. Randomized controlled studies on this topic are not ethical. However, we can synthesize data from observational studies that include the prerequisites for causal inference: correlation, time order, and ruling out other plausible explanations (Shadish et al., 2002). To assess correlations, abortion must be studied as a variable, not a constant (i.e., compared with a non-abortion group). To establish time order, abortion must precede the measurement of mental health outcomes. One of our challenges will be assessing the extent to which studies are able to establish time order. Another challenge will be assessing the extent to which studies are able to rule out other plausible explanations, including unmeasured pre-existing differences between groups of people who had abortions and those who did not (see Figures 1 and 2).

In this protocol, we refer to variables that may be associated with both abortion and mental health outcomes as potential *confounding* variables (or confounds), especially if there are uncontrolled *within-study* differences on these variables between abortion and non-abortion groups. Between-study differences (e.g., use of different types of comparison groups, use of statistical controls for prior measures of mental health) are treated as potential *moderators* of the association between abortion and mental health outcomes. We will use meta-analysis to synthesize study-level (aggregated) data on confounds and moderators; we will not use meta-analytic structural equation models to investigate whether moderators contribute to causal chains; that is, whether they *mediate* associations between abortion and mental health outcomes.

Our topic seems to meet three of the four criteria outlined by Bendersky et al. (2022) for conducting and maintaining a living systematic review (LSR): (1) the topic is relevant for policy and health decision making, (2) there is considerable uncertainty about the evidence for important outcomes, and (3) new evidence is likely to impact conclusions (also see Elliott et al., 2014, 2017). The fourth criterion—availability of sufficient resources to support a LSR—is unknown at this time. We will seek funding to support this systematic review and meta-analysis, along with important LSR components, if possible.

2 | OBJECTIVES

We will identify, critically appraise, and synthesize aggregate data from studies of associations between abortion and subsequent mental health outcomes to address the following objectives:

- Assess methodological qualities of studies that assess mental health outcomes of abortion, using a risk of bias framework.
  - o To what extent do studies establish time order in associations between abortion and mental health outcomes?
  - o To what extent do studies use design or statistical controls for mental health history?
  - o To what extent do studies use design or statistical controls for other confounding variables?
- Does abortion increase the risk of mental health outcomes?
  - o Is abortion associated with subsequent mental health symptoms or disorders and, if so, which ones?
  - o If associations are detected, do they persist after known confounding factors are taken into account?
  - o If detected, at what point(s) in time after abortion are associations with mental disorders evident (i.e., are there short-term, medium-term, and/or long-term associations)?
  - o If detected, are associations homogeneous or heterogeneous across studies?
- To what extent are associations between abortion and subsequent mental health outcomes moderated by the following study-level variables:
  - o Ability to establish time order in associations;
  - o Statistical or design controls for mental health history;
  - o Statistical or design controls for other confounding variables;
  - o Other risk of bias (RoB) variables (e.g., selective reporting of outcomes);
  - o Characteristics of the comparison group (e.g., do results differ depending on whether the comparison group includes a population of all persons capable of pregnancy, or pregnant persons, or people with unintended pregnancies, or people who sought but were denied abortions?);
  - o Gestational duration at the time of abortion (first trimester or later); and
  - o Characteristics of the geopolitical context (country, state, or province) at the time of the study (e.g., legal restrictions on abortion, country income level).

If sufficient resources are available, we will develop semi-automated search and screening procedures and workflows needed to create an LSR structure, so that results of this review can be easily updated as new information becomes available.

3 | METHODS

3.1 | Criteria for considering studies for this review

Eligible studies must (1) include parallel (non-overlapping) abortion and comparison groups and (2) obtain data on participants' mental health at least 1 day after the index abortion or after the end of the pregnancy (post-birth or post-pregnancy loss) or at a similar point in time on the calendar for cases in the comparison group. Eligibility criteria are described in greater detail below and are summarized in Supporting Information: Appendix 2.

3.1.1 | Types of studies

Eligible studies must include at least two parallel groups, including at least one group of people who had abortions and at least one non-overlapping comparison group (a reference group or control group).

Parallel groups must be non-overlapping at the time of observation (data collection). That is, for purposes of the study, an individual can only be a member of one group (the abortion group or comparison group) at any point in time, even though her reproductive status may change over time.

We expect some included studies to treat abortion as a status variable that is fixed at a certain point in time; this status is often used to identify an index abortion (if applicable) and determine membership in the abortion group or the comparison group. In these studies, a change in reproductive status (e.g., subsequent pregnancy, abortion, or live birth) should not change participants' initial group membership (just as group member does not change in an intent-to-treat analysis, regardless of receipt or completion of the services to which the participants were assigned). We will deal with deviations from this principle as a risk of bias issue (see "avoidance of selection bias, other criteria" below), not as a study exclusion criterion.

We will also include studies that treat abortion as a time-varying exposure (or status). This approach appears in some epidemiological studies, in which subsequent pregnancies, births, and abortions may be considered. The strengths of this approach often include its reliance on prospective longitudinal data with multiple pre- and post-measures of dependent variables (e.g., mental health).

To ensure that region-specific regulations related to abortion applied equally to both groups, we will exclude studies that compared historical cohorts (groups that were not observed at the same points in time) as well as studies where the abortion and comparison groups were located in separate jurisdictions (different states, provinces, or countries). We will include multi-site studies if each jurisdiction contains both abortion and comparison groups.

To avoid the ecological fallacy, we will exclude studies that were not based on individual participant data, such as those that assessed associations between state- or place-based abortion restrictions and aggregate data on mental health outcomes in those areas (e.g., Zandberg et al., 2023).

We will include cross-sectional, longitudinal, and case-control studies that meet our eligibility criteria. These three categories include a wide array of study designs; but, because study design labels are not used consistently across disciplines, our criteria rely on key features of eligible studies (as described above), not on study design labels.

3.1.2 | Types of participants

Participants include cisgender women, transgender men, and gender non-binary individuals who are capable of pregnancy and childbirth. Participants may or may not have had previous pregnancies, abortions, and/or live births. There are no restrictions on participants' age.

3.1.3 | Types of interventions

We will include abortions that involve the use of certain medications (mifepristone and/or misoprostal, and methotrexate) and/or medical procedures (uterine aspiration, dilation and curettage, dilation and evacuation, or dilation and extraction) that are approved for use in the termination of pregnancies at various stages (gestational durations). These interventions may or may not be legal at the times and/or locations in which they occurred. We will include abortions that occurred at any gestational age.

We assume that most abortions are unwanted pregnancies, most of which are the result of unintended pregnancies although some are also desired pregnancies that are later unwanted due to fetal indications (anomalies), health indications, or due to changes in social or other circumstances (Finer & Zolna, 2011).

Abortions may be self-reported or identified through any other means (e.g., hospital or administrative records). In studies that recruit patients outside of an abortion clinic setting, self-reported abortions are typically under-reported.

We will include self-managed medication abortions, which may or may not include some assistance from health professionals. We will not include other types of self-managed abortion or other nonmedical attempts to terminate pregnancy (such as the use of herbs, other substances, or physical methods).

We will exclude studies that focused on multifetal pregnancy reduction, where the aim is to reduce the number of implanted embryos.

3.1.4 | Types of comparisons

Because abortion is often legally restricted, care must be taken to ensure that comparison groups are truly comparable to abortion groups to avoid spurious associations and rule out other plausible explanations for associations between abortion and subsequent mental health outcomes. However, there is no consensus about the

ideal comparison group in this field. Some observers suggest that ideal comparison cases are people with unintended pregnancies who sought but did not have abortions, but much can be learned from different types of comparison groups. If, for example, mental health outcomes following an abortion are similar to those in the general population, then it is difficult to conclude that abortion increases the risk of mental health problems (especially when population-based studies provide data on mental health variables at multiple points in time).

Therefore, we will accept many types of comparison groups, but will make distinctions between subgroups of studies in terms of the strength of their counterfactual models, ability to support inferences about associations between abortion and subsequent mental health outcomes, and ability to rule out other plausible explanation for any between-group differences in outcomes (Shadish et al., 2002). To do this, we will assess risks of bias related to the comparability of abortion and comparison groups (e.g., use of similar inclusion and exclusion criteria, extent to which studies controlled for baseline differences between groups, potential confounding conditions, and differential attrition) in included studies.

Eligible studies include the following types of comparison groups:

- people who could become pregnant,
- pregnant persons (including those with intended pregnancies),
- people with unintended pregnancies, or
- people with pregnancies who sought but did not receive abortions.

We will not include studies that compared people who had medication abortions with those who had procedural abortions, unless the study also includes a non-abortion comparison group. Relevant comparisons will be limited to abortion versus non-abortion groups.

3.1.5 | Types of outcome measures

Eligible studies must have obtained data on measures of mental health; that is, data on participants' mental health at least 1 day after the abortion, or after the end of pregnancy, or at a similar point in time on the calendar for people who did not have abortions or pregnancies.

We will include studies with and without earlier (e.g., pre-pregnancy or pre-abortion) mental health measures. We will handle variations in the availability and use of initial mental health measures as a risk of bias issue, not an eligibility criterion.

We will include studies that used lifetime measures, other incidence and prevalence measures, and contemporaneous measures of mental health outcomes.

We will only include outcome measures that were defined and collected in identical ways (at the same points in time, with identical measurement instruments and data collection procedures) from the abortion and comparison groups. We will have more confidence in outcomes that were assessed with

measurement instruments that have evidence of reliability and validity (e.g., Cronbach's alpha or Cohen's kappa > 0.7) in the study sample or in similar samples, but will handle this as a risk of bias issue, not an inclusion criterion.

To address concerns about outcome reporting bias, we will include studies that *measured* one or more primary or secondary outcomes regardless of whether they *reported* results for these outcomes.

Primary outcomes

Primary indicators of mental health morbidity are (1) diagnosable conditions (mental disorders, as described in the DSM-5-TR or the ICD-10, or in previous versions of these classification systems), (2) measurable symptoms, (3) reported use of psychotropic medications, and (4) service utilization (outpatient visits, inpatient psychiatric hospitalization) for the following mental disorders that have been linked (in theory and/or past research) to abortion:

- PTSSs or PTSD,
- other anxiety symptoms and disorders,
- depression and other mood disorders,
- suicidal ideation, suicide attempts, and suicide,
- substance use and substance use disorders,
- indicators of any mental disorder, and
- overall mental health symptoms (e.g., scores on the Symptom CheckList [SCL-90-R], Derogatis, 1992).

These outcomes may be measured on continuous scales or as dichotomous variables (e.g., the proportion of participants who meet diagnostic criteria for mental disorders).

Secondary outcomes

Secondary outcomes include indicators of mental health morbidity (diagnosable conditions, measurable symptoms, reported use of psychotropic medications, outpatient treatments, and psychiatric hospitalization) for the following mental disorders that have more tenuous theoretical and empirical links to abortion:

- eating disorders,
- sleep disorders,
- sexual disorders, and
- psychotic disorders.

As above, secondary outcomes may include dichotomous and continuous measures.

3.1.6 | Duration of follow-up

Although included studies must have at least one set of outcome measures obtained after the end of the index pregnancy (or at a comparable point in time for comparison cases that did not involve pregnancy), we will include all available follow-up measures in our

analyses. We expect the strength of associations between abortion and outcomes to decline over time.

3.1.7 | Types of settings

There are no geographic, language, timeframe, or publication restrictions. We will use translation software with any potentially relevant non-English language materials to ensure that at least two reviewers can independently assess each document and we will rely on our team members' language expertise as appropriate (one member of our review team speaks Danish and one speaks Spanish).

3.2 | Search methods for identification of studies

The search methods include a variety of strategies designed to maximize the likelihood of finding potentially eligible studies (sensitivity) regardless of their publication status (Kugley et al., 2017; Lefebvre et al., 2019). We will carefully document search strategies so that they can be easily replicated and updated.

3.2.1 | Electronic searches

An initial strategy for Medline (Ovid) was adapted from the National Collaborating Centre for Mental Health (NCCMH) (2011) review and further modified based on this review's eligibility criteria and term harvesting. The search was tested against a set of 91 benchmark studies compiled from previous related systematic reviews and a search of Google Scholar, and was peer-reviewed by three information specialists with experience in evidence synthesis in health and psychology. The complete details of the search strategy development process can be found in Open Science Framework (https://osf.io/v8t92/). The full search strategy for Medline is in Supporting Information: Appendix 1. No date or language limits will be imposed on the search.

A variety of databases containing research in health, psychology, and social sciences, as well as a number of multi-disciplinary databases and databases with different geographic coverage were selected with the aim of being comprehensive and minimizing bias. The search will be adapted and run in these databases and the exact search strategies, search dates and numbers of results will be documented and reported.

- Medline (Ovid)
- Embase (Ovid)
- Global Health (Ovid)
- CINAHL (EBSCO)
- APA PsycInfo (EBSCO)
- Scopus (Elsevier)
- Social Science Citation Index (SSCI) (Web of Science)
- Science Citation Index Expanded (SCI-Expanded) (Web of Science)

- Conference Proceedings Citation Index-Science (CPCI-S) (Web of Science)
- Conference Proceedings Citation Index-Social Science & Humanities (CPCI-SSH) (Web of Science)
- Emerging Sources Citation Index (ESCI) (Web of Science)
- Applied Social Sciences Index and Abstracts (ASSIA) (ProQuest)
- International Bibliography of Social Sciences (IBSS) (ProQuest)
- Sociological Abstracts (including Social Services Abstracts) (ProQuest)
- GenderWatch (ProQuest)
- Latin American and Caribbean Health Science Literature (LILACS): https://lilacs.bvsalud.org/en/
- Africa-Wide Information (EBSCO)
- Netherland Research Portal, https://netherlands.openaire.eu/
- Research Portal Denmark, https://local.forskningsportal.dk/search/50274
- Cochrane Database of Systematic Reviews (Cochrane Library)
- Epistemonikos: https://www.epistemonikos.org

In addition, we will search databases of theses and dissertations, clinical trials, working papers, and preprints:

- Dissertations and Theses Global (Proquest)
- EThOS e-theses Online Service: https://ethos.bl.uk/
- Australian Theses and Dissertations: https://trove.nla.gov.au/landing/explore
- The Cochrane Central Register of Controlled Trials (CENTRAL) (Cochrane Library)
- *WHO International Clinical Trials Registry Platform (ICTRP)*: https://www.who.int/clinical-trials-registry-platform
- medRxiv: https://www.medrxiv.org/
- SSRN https://www.ssrn.com/index.cfm/en/
- IDEAS/RePEc: https://ideas.repec.org/

3.2.2 | Searching other resources

To identify other potentially eligible studies, we will search relevant gray literature sources (Kugley et al., 2017). A list of websites to be searched is in Supporting Information: Appendix 1.

The following journals and conferences proceedings will be hand-searched by reviewing the tables of contents for the last 6 months of journal issues or most recent years of proceedings:

- *Contraception*: https://www.contraceptionjournal.org/
- International Conference on Family Planning: https://icfp2022.org/
- *Population Association of America*: https://www.populationassociation.org/events-publications/future-past-meetings
- *Psychosocial Workshop*: https://www.populationassociation.org/events/event-description?CalendarEventKey=17901441-72e6-479d-9c37-0185f1add416
- Society for Family Planning: https://societyfp.org/learning/annual-meeting/

- Abortion and Reproductive Justice Conference: https://arjc2024.asap-asia.org/

The references of studies included in the review after full-text screening will be manually screened and relevant references (i.e., potentially eligible studies) will be harvested. Citing references (i.e., forward citations) will be identified using a combination of citation databases (e.g., Google Scholar, Web of Science). The references of other reviews of studies of abortion and mental health (including those cited in Section 1 and in Table 1) will be scanned for additional relevant studies.

We will use a form of snowball sampling to identify researchers and other key informants who are likely to know of relevant unpublished or ongoing studies. We will ask each informant to identify such studies and to identify others who might be able to so. We will continue to contact key informants until no new recommendations are provided. This process will also include calls for studies to listservs and newsletters (Society for Family Planning, WHO Implementing Best Practices Network [IBP], and International Campaign for Women's Right to Safe Abortion Newsletter). Additional information on potentially eligible studies will be sought from study authors as needed.

Google Scholar will also be searched and a set number of results screened for relevance.

3.3 | Data collection and analysis

3.3.1 | Description of methods used in primary research

A variety of methods have been used to assess potential associations between abortion and subsequent mental health outcomes. Studies have used in-person interviews and other survey methods—which vary in depth and detail—to obtain retrospective and/or contemporary self-reports on pregnancy intentions, abortions, and other pregnancy events, as well as past and current mental health status and symptoms. Some studies have relied on secondary analysis of survey data, medical records, or national registries. A few examples of the diverse methods employed in this field of research are provided below.

The Turnaway Study was a unique, prospective, longitudinal study that examined mental health, physical health, and socioeconomic consequences of abortion compared to being denied an abortion and carrying the unwanted pregnancy to term (Biggs et al., 2015, 2016, 2017; Dobkin et al., 2014; Foster, 2020). From 2008 to 2010, participants were recruited after they requested an abortion in 1 of 30 abortion facilities in the USA. Some participants received abortions because they were under the gestational limit of the clinic, while others were "turned away" and carried the pregnancy to term because they were past the gestational limit. Research assistants interviewed participants by phone every 6 months over a period of 5 years, ending in January 2016. Of the 1132 people who consented to participate, 956 (85%) were interviewed (research participation

rates were somewhat lower (79%) in the Turnaway group than in the abortion groups; Dobkin et al., 2014). The interviews covered a wide range of topics including mental health, physical health, employment, educational attainment, relationship status, contraceptive use, and emotions about pregnancy and abortion (Biggs et al., 2017; Foster, 2020).

Other comparison groups studies with primary data. Reardon and Ney (2000) mailed a reproductive history questionnaire to homes of a large sample of U.S. women. Analysis was restricted to White women who reported having had one or more abortions (n = 137) or no abortions (n = 395), and yes/no answers to the question, "Have you ever abused drugs or alcohol?" Broen et al. (2005) followed two groups of Norwegian women from 10 days to 5 years after a first trimester abortion (n = 80) or early miscarriage (n = 40), and compared results on anxiety, depression, and subjective well-being at multiple points in time while controlling for potential confounders. Gilchrist et al. (1995) conducted a large, prospective, longitudinal study of women with unplanned pregnancies in the UK (6410 had abortions, 6151 did not seek abortions, 379 sought and were denied abortions, and 321 sought abortions but changed their minds). Post-delivery/abortion psychiatric outcomes included psychosis, nonpsychotic illness, deliberate self-harm without other psychiatric illness (e.g., drug overdose), and no psychiatric illness. Between-group comparisons controlled for measures of prior mental health as well as other covariates (e.g., age, marital status, and prior history of abortion).

Secondary analysis of survey data. Relevant studies have been based on secondary analysis of national longitudinal data sets compiled in the USA, such as the National Longitudinal Survey of Youth (NLSY), the Health of American Women Survey (HAWS), the National Pregnancy and Health Survey (NPHS), National Longitudinal Study of Adolescent Health (Add Health), National Survey of Family Growth (NSFG), National CoMorbidity Survey (NCS), and on data sets from specific U.S. metropolitan areas (Major et al., 2009). These data sets provide opportunities for different teams of researchers to review and replicate or challenge the work of others (see, e.g., Steinberg & Finer, 2011; Steinberg et al., 2014).

Secondary analysis of medical records. Several papers by Reardon and colleagues were based on overlapping samples of medical records from California's state-funded medical insurance program for low-income families. These studies excluded women who had subsequent abortions from the delivery group, but not from the abortion group (Major et al., 2009).

Secondary analysis of national registries. Munk-Olsen and colleagues conducted a population-based cohort study, by linking information from three Danish national registries from 1995 to 2007 to obtain data on a sample of girls and women who had no record of mental disorders when they had either a first-trimester induced abortion or a first childbirth. They estimated rates of first-time psychiatric contact (inpatient admission or outpatient visit) for any type of mental disorder within 12 months after the abortion or childbirth, compared with the 9-month period preceding the event (Munk-Olsen et al., 2011). Other studies have been based on national registry data in Finland (Major et al., 2009).

3.3.2 | Selection of studies

As described below, the selection of studies for our review will occur in two stages: (1) screening of titles and abstracts and (2) eligibility decisions based on full text reading of all relevant study reports.

Search results (titles and abstracts) will be imported into Covidence for screening purposes. For titles that lack English language abstracts, we will retrieve and/or translate the abstract (or a portion of the document, if needed), and will upload this information into Covidence before screening. We will use Google translate to produce English versions of non-English abstracts and documents.

We will use the machine learning (ML) tool within Covidence to facilitate screening (Chappell et al., 2023). Informed by human screening decisions, Covidence uses machine ranking and continuous machine training to sort citations in the order of relevance.

Working independently, two reviewers will apply the screening tool (shown in Supporting Information: Appendix 2, Stage 1) to titles and abstracts. Screening decisions are Yes (potentially eligible), No (not eligible), and Maybe. We will sort these records into three groups: both reviewers voted Yes, both voted No, and conflicting opinions or Maybe votes. We will meet periodically to discuss conflicting opinions and resolve those that are clearly Yes or clearly No, leaving unresolved records as is until the end of screening.

Using ML to sort records by relevance, we will use a stopping rule to transition from two screeners per record to one screener as we approach less relevant records. Only one reviewer will screen the remaining records, which are expected to be not eligible or unclear. As the use of ML for this purpose is rapidly evolving, we will consult with experts on the pros and cons of various stopping rules and will select a stopping rule that is transparent, sensitive, and efficient.

At the end of screening, unresolved records will be promoted along with Yes (potentially eligible) records for full text retrieval.

Before study eligibility decisions are made, citations will be collated to determine which citations (reports) are associated with which unique studies. A unique study involves a sample that does not overlap with another study sample in this review. Studies may have multiple citations and citations may include information on multiple studies (non-overlapping samples).

Study eligibility decisions will be made by two reviewers working independently, using the second half (Stage 2) of the screening and eligibility tool shown in Supporting Information: Appendix 2. Reviewers will compare notes after they have completed their assessments and will discuss and attempt to resolve any discrepancies. Discrepancies that are not resolved by the first two reviewers will be discussed and resolved with a third reviewer. Reasons for exclusion will be documented for all studies excluded at this stage (based on full text analysis) and a full list of

these excluded studies with specific reasons for their exclusion will be provided in the final review.

3.3.3 | Data extraction and management

For each eligible study, data extraction will be conducted by at least two review team members who were not involved in the conduct of that study. Working independently, two reviewers will extract data using the form shown in Supporting Information: Appendix 3. As above, when finished with their independent assessments, reviewers will compare notes, resolve discrepancies, and consult with a third reviewer when necessary to resolve discrepancies.

For each study, we will extract information on:

- Study or data set characteristics;
- Sub-study characteristics (to distinguish multiple analyses derived from shared data sets, such as the U.S. National Comorbidity Study): comparison groups, study design, timing, locations;
- Participant characteristics: demographics, socioeconomic characteristics, initial mental health status;
- Measurement tools: reliability and validity of outcome measures;
- Effect size data: raw data, valid subgroup Ns, attrition, analytic models, covariates (e.g., statistical controls for initial differences between groups); and
- Study-level risks of bias.

As shown in Supporting Information: Appendix 3, data extraction will be structured hierarchically, with up to five nested levels: study-level data, sub-study-level data, report-level data, measurement-level data, and effect-size-level data. For purposes of data analysis, we will generate an effect-size-level file (a spreadsheet) with all relevant study-, sub-study-, report-, and measurement-level data for each unique effect size (ES). This structure will handle repeated measures, by treating them as separate (unique) ES.

We will try to obtain data on the nature of any legal restrictions on abortion that were in place at the time and location of each included study. If this information is not provided in the study reports, we will search the Center for Reproductive Rights (2023) database on the world's abortion laws for information relevant for included studies (https://reproductiverights.org/maps/worlds-abortion-laws/).

Meta-Reviewer (https://www.metareviewer.org/) will be used for data extraction because it is designed to support duplicate extraction and a hierarchical data structure. Online data extraction forms will mirror the forms shown in Supporting Information: Appendix 3 and will be used to store all (duplicate) coding decisions, along with all resolutions (of initial coding differences) and validation checks.

All coders will independently pilot test the screening and data extraction forms and revisions will be made as needed to improve the clarity, efficiency, and inter-rater reliability of these tools. We will apply all criteria, questions, and coding uniformly across all included studies.

3.3.4 | Assessment of risk of bias in included studies

Included studies will be judged against the following criteria, adapted from the Cochrane Risk of Bias tool (version 1.0, Higgins & Green, 2011), the What Works Clearinghouse standards for initial group equivalence and attrition (What Works Clearinghouse, 2018a, 2018b), previous Campbell systematic reviews (Littell et al., 2005, 2008, 2021, 2023; Valentine et al., 2023), and the confounder matrix for assessing observational studies of etiology (Petersen et al., 2022). Risk of bias ratings will not be used as criteria for inclusion in the synthesis but may be used in subgroup and moderator analyses.[1]

Time order: study design establishes a clear temporal order for abortion and mental health measures.

- Yes = Low risk: longitudinal study collected data on mental health measures at specific time points/intervals (e.g., present time, within last 3 months, within last 12 months) before and/or after abortion or other pregnancy events (or at similar points in time on the calendar for comparison cases).
- Unclear risk = insufficient information.
- No = High risk: use of correlational data, lifetime measures of abortion, and/or lifetime measures of mental health outcomes.

Avoidance of confounding (initial equivalence on mental health history): initial differences between groups on measures of mental health history were small or moderate ($d \leq 0.25$) or researchers used statistical controls (e.g., propensity score matching, regression covariates) for initial differences.

- Yes = Low risk
- Unclear risk: insufficient information (e.g., group-level data were not provided, d cannot be computed, unclear if statistical controls were sufficient to create comparable groups)
- No = High risk: there were initial differences between groups with $d > 0.25$, no/inadequate statistical controls for these differences, or no measures of mental health history.

Avoidance of confounding (pregnancy intention): comparison groups were matched on pregnancy intention (e.g., pregnancy intention was used as an inclusion criterion, used in matching designs and/or the proportion of cases with unintended or unwanted pregnancies was equivalent ($d \leq 0.25$) across groups).

- Yes = Low risk.
- Unclear risk = insufficient information.

- No = High risk: between group differences on pregnancy intentions ($d > 0.25$).

Avoidance of confounding (initial equivalence on background characteristics): initial differences between groups on socioeconomic variables (e.g., income, education, employment) and family background characteristics (e.g., adverse childhood experiences, interpersonal violence) were small or moderate ($d \leq 0.25$) or researchers used statistical controls (e.g., propensity score matching, regression covariates) for baseline differences.

- Yes = Low risk
- Unclear risk: insufficient information (e.g., group-level data were not provided, d cannot be computed, unclear if statistical controls were sufficient to create comparable groups)
- No = High risk: there were initial differences between groups with $d > 0.25$, and no/inadequate statistical controls for these differences.

Avoidance of confounding (other criteria): comparison groups were constructed and maintained with similar inclusion and exclusion criteria (e.g., related to age, prior pregnancies, live births, abortions, and prior mental health status), taking into account any country- or region-specific regulations limiting access to abortion to certain groups.

- Yes = Low risk.
- Unclear risk = insufficient information.
- No = High risk: inclusion or exclusion criteria were applied differently to different comparison groups.

Avoidance of confounding (performance bias): there were no systematic differences ($d \leq 0.25$) between groups (or use of statistical controls for differences between groups) in terms of levels of service, care, attention, perceived stigma, or social support.

- Yes = Low risk
- Unclear (insufficient information)
- No = High risk: one group received more services, care, attention, perceived stigma, or support and these factors were not accounted for in the analysis.

Avoidance of detection bias (blinding): assessor is unaware of group membership when collecting outcome data.

- Yes for all outcomes = Low risk
- Yes for some outcomes = Unclear
- Unclear (insufficient information)
- No = High risk

Avoidance of attrition bias: Losses to follow up were less than or equal to 25% and equally distributed (≤10% difference in response rates) across groups. Group equivalence on important initial characteristics was retained after losses to follow-up ($d \leq 0.25$).

---

[1]Our approach differs from the ROBINS-I tool which was developed for assessment of risks of bias in nonrandomized studies of intervention effects (Sterne et al., 2016). ROBINS-I recommends *post hoc* exclusion of studies with a critical risk of bias rating on any one (or more) items. In our view, different types of risk of bias are likely to have different impacts on effect sizes (ES), and these relationships should be explored empirically when possible. Further, the ROBINS-I approach often leads to *post hoc* exclusion of otherwise-eligible studies without transparent documentation of reasons for exclusion (specific risks of bias). We think that full critical appraisal of all studies that meet our eligibility criteria is important, given controversies over the interpretations and uses of results of these studies.

- Yes for all outcomes = Low risk
- Yes for some outcomes = Unclear
- Unclear (insufficient information)
- No = High risk: loss of initial equivalence (d > 0.25), losses to follow up >25% overall, or losses were unequally distributed (>10% difference) across groups.

Standardized observation periods: follow-up data were collected from each case at a fixed point in time (e.g., after the end of pregnancy), or analyses included statistical controls for variable observation periods (e.g., event history analysis).

- Yes for all outcomes = Low risk
- Yes for some outcomes = Unclear
- Unclear (insufficient information)
- No = High risk

Validated measures of mental health history: use of instruments with demonstrated reliability (e.g., alpha/kappa > 0.7) or validity in the study sample or in similar samples (from similar countries, socioeconomic and racial/ethnic groups), or use of administrative data on mental health events (e.g., prior psychiatric hospitalizations).

- Yes = Low risk: reliable mental health history measures were collected before, during, or within 14 days after pregnancy and at the same point(s) in time for all groups.
- Unclear risk = insufficient information on reliability, validity, or timing of baseline mental health measures.
- No = High risk: reliance on retrospective reporting, single item self-reports, or unreliable measures (alpha/kappa < 0.7) or no measures of mental health history.

Validated mental health outcome measures: use of instruments with demonstrated reliability (e.g., alpha/kappa > 0.7) or validity in the study sample and/or similar samples from (similar countries, socioeconomic, and racial/ethnic groups), or use of external administrative data on events (e.g., psychiatric hospitalization).

- Yes for all outcomes = Low risk
- Yes for some outcomes = Unclear
- Unclear (insufficient information)
- No = High risk

Free of selective reporting: a prospective study protocol is available and all pre-specified outcomes are reported in the pre-specified way; all expected outcomes are reported in full and for all cases (e.g., no systematic exclusion of subgroups of cases), regardless of the direction and statistical significance of results.

- Yes = Low risk
- Unclear (e.g., prospective protocol is not available, or changes in the protocol were made after the study began)

- No = High risk: some outcomes are not reported or are reported incompletely (e.g., non-significant results are mentioned, but data are not provided; data are provided for selected subgroups only).

All risk of bias assessments will be documented in a section on Characteristics of Included Studies, with support for each judgment, including verbatim passages from study reports with clear citations. Results will be summarized in a Risk of Bias Graph and a Risk of Bias Summary with data on each rating for all included studies (e.g., see Littell et al., 2021, 2023).

3.3.5 | Measures of treatment effect

Certain ES metrics will be used to compare results obtained from two parallel groups (an abortion and comparison group) on continuous and dichotomous outcome measures.

Continuous outcomes. Between-group differences on continuous outcome measures, such as scores on standardized measures of symptoms of mental disorders, will be analyzed using Hedges' g, the small-sample adjusted standardized mean difference (SMD) between the two groups. We will follow the US What Works Clearinghouse (WWC, 2022) guidelines for computing Hedges' g using unstandardized regression coefficients and the pooled, unadjusted sample standard deviations (see appendix E in WWC, 2022). If only unadjusted summary statistics (means and standard deviations, t-tests, etc.) are reported, we will extract these estimates to compute Hedges' g, but will treat unadjusted estimates separately from covariate-adjusted estimates in the analysis (as described below).

Dichotomous outcomes. Studies may report the odds or risks of a mental health condition in the two groups or the time-to-event/hazards ratios for both groups. Risk ratios will be converted to odds ratios (ORs), following the guidance in Borenstein and Hedges (2019), using the risk of the event in the comparison group as the baseline risk. For purposes of analyses, ORs and their associated standard errors will be converted to log units (LORs) and their associated standard errors. For studies that use logistic regression models to estimate a covariate-adjusted OR, we will extract the adjusted OR and its associated SE, and then convert these statistics to the log-odds ratio (LOR) and its associated SE for purposes of meta-analysis. Then, LORs and their SE will be converted back to ORs and their SEs for purposes of reporting and interpretation.

Studies may report hazard ratios for the odds of a particular outcome occurring during a certain time period. We will extract hazard ratios if no other information is available. We will explore the use of methods Watkins and Bennett (2018) provided for converting binomial counts/proportions from two groups allowing the synthesis of studies reporting either hazard ratios or binomial counts/proportions.

Adjusted and unadjusted effect sizes. For both continuous and dichotomous outcomes, effect sizes that are adjusted for initial between-group differences will be prioritized and treated separately from unadjusted estimates in the meta-analysis models. Studies may

present the results of multiple statistical models that control for various differences between the two groups. We will prioritize effect sizes that adjust for the following initial between-group differences:

- mental health history,
- age,
- exposure to violence or trauma,
- socioeconomic status (income, education, poverty status),
- marital or relationship status, and
- number of children.

As we extract data, we will use coding to capture the covariates used in each model (for each adjusted ES). If studies only present unadjusted ES, we will extract these estimates, but will treat them separately in the meta-analysis (see Section 3.4 below).

When reports provide insufficient data for effect size calculations, we will attempt to retrieve additional information from the study authors.

After computing effect sizes, we will examine outliers and check to make sure that our data accurately reflect study reports.

3.3.6 | Unit of analysis issues

It is possible that some participants will have multiple pregnancies (events) during the study time period. Abortion and comparison groups are defined by the conclusion of an index pregnancy (usually the first pregnancy that occurs in the study timeframe) or the passage of a fixed amount of time. We assume that included studies will treat individual participants (not pregnancies or abortions) as the primary unit of analysis (i.e., an individual should not be double-counted if they had two pregnancy-related events).

Studies with more than two comparison groups may provide useful opportunities to examine multiple contrasts (e.g., abortions vs. live births and abortions vs. miscarriages). We will extract all relevant data from all eligible comparison groups, taking care to keep these estimates separate in pairwise meta-analysis (data on one group cannot be used twice in the same pairwise meta-analysis). As explained below, all outcome data will be included in correlated and hierarchical effects (CHE) models, but different contrasts will be coded as such.

Similarly, we expect to find multiple analyses (and multiple reports) based on shared data sets. For purposes of our review, a data set is a unique sample of participants that does not overlap with another data set; thus, a data set is considered a "study" in our review. Sub-studies are based on shared data sets, but may use different comparison groups (i.e., different inclusion and exclusion criteria), different time frames, covariates, and outcome variables. We will assess the overlap between sub-studies that relied upon the same data sets to determine if they share participants; methods for handling overlapping samples and other sources of dependencies among effect sizes are described in the next section.

Multiple measures of the same outcome (e.g., multiple scales and/or dichotomous measures) will be included in the analysis, following plans detailed in the Data Analysis section.

If studies use clustered samples and did not adjust for clustering, we will adjust the standard errors of the SMDs using the methods described in the WWC 5.0 Procedures and Standards Handbook (What Works Clearinghouse, 2022). For ORs reported from unadjusted clustered samples, we will adjust the standard error of the LOR using methods recommended in the Cochrane Handbook.

3.3.7 | Criteria for determination of independent findings

We will include data from all available study reports on all relevant outcomes, contrasts, and endpoints in our analysis. We are interested in mean effects and in exploring the heterogeneity of effects across studies (i.e., both convergent and divergent approaches), so we will deal with effect size multiplicity (multiple statistically dependent effect sizes) by using both reductionist and integrative meta-analytic techniques (López-López et al., 2018). As described below, we will conduct two separate types of synthesis: pairwise meta-analysis (using a reductionist approach) and CHE models (an integrative approach).

We will construct as many pairwise meta-analyses as are needed to capture conceptually distinct outcomes and distinct time intervals. To avoid multiplicity, each study can contribute only one effect size to each pairwise meta-analysis. We will select one ES per study for each pairwise meta-analysis, using the following selection rules.

We will collapse endpoints into common non-overlapping intervals, for example: less than 90 days after pregnancy, 3–6 months, 7–11 months, and 12–23 months, 24–35 months, 3–5 years, and so forth. If a study has more than one relevant data point within a particular time interval, we will use the latest endpoint within that interval in that pairwise meta-analysis.

When studies report multiple measures of the same outcome (e.g., depression) at the same point in time, we will select the most comprehensive measure (e.g., overall depression score) over single items or single/selected symptoms for pairwise meta-analysis. If outcome data are provided by multiple sources (self-reports, collateral interviews, and official records), we will use the most direct or reliable measure in pairwise meta-analysis; self-reports on validated measures are considered more direct than official records, and collateral reports are considered less direct that self-reports and official records.

Data from multiple sub-studies or multiple reports with overlapping samples will not be included in the same pairwise meta-analysis to avoid duplication of cases and dependent effect sizes. Instead, we will display results from overlapping samples as separate subgroups in forest plots to show results of different analyses based on the same (or overlapping) sample of cases; results will not be pooled (averaged) across subgroups with overlapping samples.

We will use CHE models with robust variance estimates (RVE) to combine multiple dependent effect sizes across studies (Pustejovsky &

Tipton, 2022). Described below (in Section 3.4), these analyses allow us to include multiple effect sizes per study (including multiple measures of the same outcome, multiple endpoints, multiple subgroups, and multiple comparisons) in statistical models that adjust for multiplicity.

3.3.8 | Dealing with missing data

We will request missing data from the primary authors of included studies when needed to understand study characteristics, complete risk-of-bias assessments, or calculate effect sizes, paying particular attention to outcome data that were collected but not reported or under-reported.

We will record data on attrition and differential attrition for each effect size (each outcome and each endpoint). If there is missing data for more than 50% of one or both comparison groups, we will not include that effect size in meta-analysis.

Where possible, we will use Cochrane's revman-calculator to calculate missing standard deviations.

We expect eligible studies to differ in the reporting of moderators that we plan to use in exploring heterogeneity. Before using statistical methods to explore the sensitivity of results to missing covariates, we will conduct and report an exploratory analysis of the amount and patterns of missing data as described by Schauer et al. (2022). These exploratory methods can highlight what covariates are incompletely reported as well as patterns of incomplete data across studies. Research is ongoing regarding the application of standard missing data methods such as multiple imputation to meta-analysis (see e.g., Lee & Beretvas, 2023; Schauer, 2019). A concern raised by this research is the association between effect size and likelihood of missing values, a situation unique to meta-analysis. We will consider the use of multiple imputation as a method of checking the sensitivity of our results to missing data depending on the pattern and prevalence of missing covariates.

3.3.9 | Assessment of heterogeneity

We will evaluate heterogeneity with the Chi-square test of heterogeneity, I^2 (in pairwise meta-analysis), and tau-squared (τ^2). For CHE models estimating the mean effect size, we will present the square-root of the within-study (ω^2) and between-study components of variance (τ^2), and the prediction interval. We will also consider presenting the multivariate version of I^2 as suggested by Viechtbauer (2022).

With sufficient studies, we will use meta-regression to explore associations between study and participant characteristics (potential moderators of effects) and the direction and magnitude of an effect size.

3.3.10 | Assessment of reporting biases

We will use funnel plots to assess the risk of publication bias and other potential sources of bias. If there are at least 10 studies (independent

effect sizes) in a funnel plot, we will use statistical tests for asymmetry (e.g., Egger's test) for dependent effect sizes as discussed by Rodgers and Pustejovsky (2021) and an extension to selection models using cluster wild bootstrapping methods (Pustejovsky & Joshi, 2023). We will also consider the use of other methods for exploring the sensitivity of results to publication/small sample bias such as those developed by Mathur and VanderWeele (2020).

We will use all available reports on included studies (included registered protocols, if available) to track reporting of outcomes within studies, across all endpoints, and by outcome domain. Results of these analyses will be arrayed graphically and used to support judgments regarding risks of selective reporting (e.g., see Littell et al., 2021, 2023).

3.4 | Data synthesis

We will use pairwise meta-analysis to synthesize data from multiple, non-overlapping studies on comparable outcome measures at similar points in time. We will use CHE meta-analysis models to synthesize data on all available outcomes within conceptually distinct outcome domains. These two different kinds of meta-analysis are explained below.

We do not expect all studies to produce estimates of the same population parameters (i.e., risks of subsequent mental health problems associated with abortion), given the likely differences between studies in research designs, settings, participant characteristics, and outcome measures. Therefore, we will use random effects meta-analysis models whenever possible (i.e., in pairwise meta-analysis and in CHE models with $df > 4$).

3.4.1 | Pairwise meta-analysis

We will conduct separate pairwise meta-analyses for conceptually distinct outcomes, at specific endpoints, and within the following outcome domains:

- PTSSs or PTSD,
- other anxiety symptoms and disorders,
- depression and other mood symptoms or disorders,
- suicidal ideation and suicide attempts,
- substance use and substance use disorders,
- any mental disorder or mental health symptoms,
- eating disorders,
- sleeping disorders,
- sexual disorders, and
- psychotic disorders.

We plan to present data in subgroups by study design variables, and display results in forest plots alongside study ROB ratings. If possible, contrasts between cases with abortions and live births will be kept separate from contrasts between cases with abortions and other pregnancy outcomes (e.g., miscarriages) or no pregnancy outcome. In

some analyses we may compare abortion with all other outcomes (including comparisons with people who were not pregnant).

Pairwise meta-analysis will be conducted in the R program *metafor* (Viechtbauer, 2010). We will use Hedges' *g*, the small-sample adjusted SMD for continuous outcomes. Inverse variance methods will be used to pool SMDs, so that each effect size is weighted by the inverse of its variance in an overall estimate of effect size. Confidence intervals (CIs) of 95% will be used for individual study data and for pooled estimates.

3.4.2 | CHE models

Observational studies often report multiple dependent outcomes, including multiple measures of the same construct, reports on the same measure from multiple data sources, and repeated measures from the same sources over time. We will use the CHE model described by Pustejovsky and Tipton (2022) to handle these dependencies, and RVE to guard against model mis-specification. The CHE model assumes that effect sizes are correlated and nested within studies, because they are derived from the same sample or from the same study. This approach provides "valid point estimates, standard errors, and hypothesis tests even when the degree and structure of dependence between effect sizes is unknown" (Fisher & Tipton, 2015, p. 1; also see: Hedges et al., 2010; Tanner-Smith & Tipton, 2014; Tanner-Smith et al., 2016).

We will use small sample corrections for RVE with meta-analysis (Tipton, 2015).

Studies can report similar outcomes in different ways (e.g., days of drug use vs. days of abstinence from drug use). Therefore, before performing the CHE analysis, we will reverse score some outcomes so that higher scores always represent more severe mental health problems.

We will use all available data on primary and secondary outcomes in the CHE models, including multiple measures of the same outcome at different points in time. We will assume a correlation of 0.6 for effect sizes measured within the same study, but will test this assumption with sensitivity analysis, assessing results for rho (ρ) = 0.4, 0.6, and 0.8. Results will show whether different values of rho produce consistent estimates of mean ES coefficients, standard errors, and tau-squared (τ^2).

We will estimate effect size models (both the mean effect size model and any moderator models) using the R programs *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2023).

For dichotomous outcomes, our synthesis will be conducted using the LOR, and we will convert results back to ORs for ease of interpretation. Then, to increase statistical power, we will conduct an analysis combining SMDs and studies reporting dichotomous outcomes. For studies reporting dichotomous outcomes, we will compute the Cox index, an effect size comparable to Hedges' *g* and used by the What Works Clearinghouse (2022). If there are too few studies with valid effect sizes for meaningful analyses within each outcome domain, we will collapse all primary mental health outcomes (the first six domains) and all secondary outcomes (the last four domains) into larger CHE analyses.

Estimates from CHE models with fewer than four degrees of freedom are unreliable (Tanner-Smith & Tipton, 2014). In these instances (when there are fewer than five studies reporting on an outcome in the analysis), we will examine forest plots of all relevant ES. We will use forest plots that reflect the dependent effect size structure, such as those included in Winters et al. (2022). We will then consider aggregating these ES within studies (computing a study-level ES) using the aggregate function in metafor and following recommendations by Pustejovsky (2019). We will assume a correlation of 0.6 for associations among effect sizes within studies. We will then use fixed effect (FE) models to estimate the mean ES across studies. We will also consider using cluster wild bootstrapping methods for estimating dependent effect size models with a small number of studies (Joshi et al., 2022).

Where possible, we will provide 95% prediction intervals (PIs) as well as 95% CIs around point estimates of main effects. PIs show the range of values within which results of future studies are likely to fall. Fixed effect models assume that tau-squared (τ^2) is zero; thus, there are no PIs for fixed effect models.

3.4.3 | Subgroup analysis and investigation of heterogeneity

We plan to fully document and transparently report the decisions we make to explore effect size heterogeneity. We anticipate overlap and multicollinearity among the planned moderators. We will explore these correlations and the distribution of moderator levels both within and across studies. Power for tests of moderators likely relates not only to the number of studies/ effect sizes within each subgroup but also their balance across studies/effect sizes.

If available data support moderator analysis (e.g., at least 5 effect sizes per subgroup/moderator level)), we will examine the following variables as potential moderators of associations between abortion and subsequent mental health outcomes:

- Study design characteristics including:
  ◦ establishment of time order (abortion precedes mental health outcomes),
  ◦ design/statistical controls for mental health history and other confounding variables,
  ◦ characteristics of the comparison group, and
  ◦ other risk-of-bias ratings;
- Gestational duration at the time of abortion (e.g., first trimester or later); and
- Context at the time and location of the study:
  ◦ type(s) of legal restrictions on abortion (Center for Reproductive Rights, 2023), and
  ◦ country income level (low, middle, high; World Bank, 2023).

We expect to find larger effect sizes (associations) when studies do not clearly establish time order, when ES were not adjusted for

mental health history and other known confounding variables, in studies with more dissimilar comparison groups (e.g., where the groups differ in pregnancy intention), and when other risks of bias are present (e.g., lack of initial group comparability). We expect larger ES in studies where a higher proportion of participants had abortions after the first trimester. Finally, we expect larger ES in contexts with more stringent legal restrictions on abortion and in settings where there is less access to health care.

If possible, we will also perform exploratory analyses with the following potential moderators:

- Type of outcome measure (e.g., short standardized self-report, registry data, in-depth interview)
- Year(s) of data collection (calendar effects)
- Type of abortion: procedural or medication abortion,
- Lag time between abortion and mental health outcome measures,
- Mean age of pregnant persons,
- Participants' socioeconomic status (income, education, poverty),
- Number of prior live births, and
- Perceived stigma related to abortion.

We will explore how groups of moderators relate to effect size heterogeneity, starting with methodological characteristics of studies, then average gestational duration, context, and (if possible) type of outcome measure and other potential moderators listed above. We will treat these effect size models as exploratory, using caution in interpreting the results. We are likely to have both direct and indirect evidence about effect size heterogeneity as detailed by Pustejovsky and colleagues (2023), and will ensure that results are interpreted appropriately.

If possible, we will estimate ES for specific outcome domains (see Section 3.1.4).

3.4.4 | Sensitivity analysis

We will use sensitivity analysis to examine the potentially biasing effects of outliers (e.g., studies with unusually large sample sizes, and those with extremely high or low ES). Sensitivity analyses will be performed by removing studies one at a time from a forest plot or from a CHE analysis and comparing results with and without a study. We will use caution in defining outliers as they are dependent on context. We will start with Tukey's (1977) suggestion to define outliers as effect size estimates falling more than three times the interquartile range below the first quartile or above the third quartile.

As mentioned above, we will assess the sensitivity of CHE and FE models to various assumptions about the size of the correlations between effect sizes within studies.

If our plans change in ways that deviate from this protocol, we will document these changes and report them—along with our rationale for making changes—in the final report (in a section on Deviations from the Protocol). Where possible, we will conduct sensitivity analysis to assess the possible impact of deviations from the protocol on the results.

3.4.5 | Treatment of qualitative research

We do not plan to include qualitative research.

3.4.6 | Summary of findings and assessment of the certainty of the evidence

We will use the GRADE guidelines (gdt.gradepro.org) to assess the certainty of evidence regarding primary outcomes in a Summary of Findings (SoF) table. If possible, the SoF table will include measures of each of our six primary outcomes at approximately 1 year after pregnancy ends.

3.5 | Data sharing plans

Data sharing is not applicable to this protocol as no data sets were generated or analyzed for during the development of the protocol.

We will publish the raw data from the completed systematic review and meta-analysis (i.e., the aggregate data that we extracted from included studies) as a csv file, along with a codebook, and the R syntax used to generate all statistical analyses.

Before publication of the final report on this project, we will seek independent replication of the results of our CHE analysis by a meta-analyst who is not associated with our review and is not aware of the meaning of the direction of effect sizes (i.e., does not know which groups is favored by positive or negative ES).

If possible, we will introduce crowd-sourcing and semi-automation of critical procedures (e.g., searching and screening) and workflows to develop and maintain an LSR (Bendersky et al., 2022; Elliott et al., 2014, 2017).

CONTRIBUTIONS OF AUTHORS
Julia H. Littell (JHL), Trine Munk-Olson (TMO), and Julia R. Steinberg (JRS) conceived the idea for this project, along with Chelsea B. Polis. JHL drafted the protocol with Sarah Young (SY) and Therese D. Pigott (TDP). All other co-authors provided input, reviewed, and commented on multiple drafts of the protocol; these co-authors (Content experts) are listed in alphabetical order by last name.

Content: M. Antonia Biggs (MAB), TMO, and JRS are experts in the content area.

Systematic review methods: JHL, SY, and TDP are experts in systematic review methods.

Information retrieval: SY is an expert in information retrieval.

Statistical analysis: TDP is an expert in meta-analysis.

DECLARATIONS OF INTEREST

MAB, TMO, and JRS have been involved in the conduct and reporting of primary research on the topic. As explained in the Methods section, under Data Extraction and Management, any author of a study considered for potential inclusion in this review will not be involved in the study eligibility decision, data extraction, or coding of that study.

PRELIMINARY TIMEFRAME

We plan to submit the final review by December 2026.

PLANS FOR UPDATING THIS REVIEW

The review will be updated by every 5 years or more often, if sufficient resources are available.

SOURCES OF SUPPORT

Internal sources: None.
External sources: None.

ORCID

Julia H. Littell http://orcid.org/0000-0002-2623-9728

REFERENCES

American Psychological Association (APA) Task Force on Mental Health and Abortion. (2008). *Report of the task force on mental health and abortion*. American Psychological Association. Retrieved May 10, 2023, from https://Www.Apa.Org/Pi/Women/Programs/Abortion/Mental-Health.Pdf

Auerbach, S. L., Coleman-Minahan, K., Alspaugh, A., Aztlan, E. A., Stern, L., & Simmonds, K. (2023). Critiquing the unintended pregnancy framework. *Journal of Midwifery & Women's Health*, 68(2), 170–178. https://doi.org/10.1111/jmwh.13457

Bearak, J. M., Popinchalk, A., Beavin, C., Ganatra, B., Moller, A. B., Tunçalp, Ö., & Alkema, L. (2022). Country-specific estimates of unintended pregnancy and abortion incidence: A global comparative analysis of levels in 2015–2019. *BMJ Global Health*, 7(3), e007151. https://doi.org/10.1136/bmjgh-2021-007151

Bearak, J., Popinchalk, A., Ganatra, B., Moller, A. B., Tunçalp, Ö., Beavin, C., Kwok, L., & Alkema, L. (2020). Unintended pregnancy and abortion by income, region, and the legal status of abortion: Estimates from a comprehensive model for 1990–2019. *The Lancet Global Health*, 8(9), e1152–e1161. https://doi.org/10.1016/S2214-109X(20)30315-6

Bellieni, C. V., & Buonocore, G. (2013). Abortion and subsequent mental health: Review of the literature. *Psychiatry and Clinical Neurosciences*, 67(5), 301–310. https://doi.org/10.1111/pcn.12067

Bendersky, J., Auladell-Rispau, A., Urrútia, G., & Rojas-Reyes, M. X. (2022). Methods for developing and reporting living evidence synthesis. *Journal of Clinical Epidemiology*, 152, 89–100. https://doi.org/10.1016/j.jclinepi.2022.09.020

Biggs, M. A., Neuhaus, J. M., & Foster, D. G. (2015). Mental health diagnoses 3 years after receiving or being denied an abortion in the United States. *American Journal of Public Health*, 105(12), 2557–2563. https://doi.org/10.2105/AJPH.2015.302803

Biggs, M. A., Rowland, B., McCulloch, C. E., & Foster, D. G. (2016). Does abortion increase women's risk for post-traumatic stress? Findings from a prospective longitudinal cohort study. *BMJ Open*, 6(2), e009698. https://doi.org/10.1136/bmjopen-2015-009698

Biggs, M. A., Upadhyay, U. D., McCulloch, C. E., & Foster, D. G. (2017). Women's mental health and well-being 5 years after receiving or being denied an abortion: A prospective, longitudinal cohort study. *JAMA Psychiatry*, 74(2), 169.

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In L. V. H. H. Cooper, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 207–243). Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis* (1st ed.). John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. https://doi.org/10.1002/jrsm.12

Brewer, C. (1977). Incidence of post-abortion psychosis: A prospective study. *BMJ*, 1, 476–477.

Broen, A. N., Moum, T., Bødtker, A. S., & Ekeberg, Ø. (2005). The course of mental health after miscarriage and induced abortion: A longitudinal, five-year follow-up study. *BMC Medicine*, 3(1), 18. https://doi.org/10.1186/1741-7015-3-18

Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., & Klassen, T. P. (2006). Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology*, 59(7), 697–703. https://doi.org/10.1016/j.jclinepi.2005.11.010

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. M. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (1st ed., pp. 193–213). Russell Sage Foundation.

Bushman, B. J., & Wells, G. L. (2001). Narrative impressions of literature: The availability bias and the corrective properties of meta-analytic approaches. *Personality and Social Psychology Bulletin*, 27, 1123–1130.

Center for Reproductive Rights. (2023). The world's abortion laws. Retrieved August 23, 2023, from https://Reproductiverights.Org/Maps/Worlds-Abortion-Laws/

Chappell, M., Edwards, M., Watkins, D., Marshall, C., & Graziadio, S. (2023). Machine learning for accelerating screening in evidence reviews. *Cochrane Evidence Synthesis and Methods*, 1(5), e12021. https://doi.org/10.1002/cesm.12021

Charles, V. E., Polis, C. B., Sridhara, S. K., & Blum, R. W. (2008). Abortion and long-term mental health outcomes: A systematic review of the evidence. *Contraception*, 78(6), 436–450.

Coleman, P. K. (2011). Abortion and mental health: Quantitative synthesis and analysis of research published 1995–2009. *British Journal of Psychiatry*, 199(3), 180–186.

Coleman, P. K., Coyle, C. T., Shuping, M., & Rue, V. M. (2009). Induced abortion and anxiety, mood, and substance abuse disorders: Isolating the effects of abortion in the national comorbidity survey. *Journal of Psychiatric Research*, 43(8), 770–776. https://doi.org/10.1016/j.jpsychires.2008.10.009

Coleman, P. K., Reardon, D. C., Rue, V. M., & Cougle, J. (2002). State-funded abortions versus deliveries: A comparison of outpatient mental health claims over 4 years. *American Journal of Orthopsychiatry*, 72(1), 141–152. https://doi.org/10.1037/0002-9432.72.1.1410155

Coleman, P. K., Reardon, D. C., Strahan, T., & Cougle, J. R. (2005). The psychology of abortion: A review and suggestions for future research. *Psychology & Health*, 20(2), 237–271.

Coleman, P. K., Rue, V. M., Spence, M., & Coyle, C. T. (2008). Abortion and the sexual lives of men and women: Is casual sexual behavior more appealing and more common after abortion? *International Journal of Clinical and Health Psychology*, 8(1), 77–91.

Cook, R. J., Ortega-Ortiz, A., Romans, S., & Ross, L. E. (2006). Legal abortion for mental health indications. *International Journal of Gynaecology and Obstetrics*, 95(2), 185–190. https://doi.org/10. 1016/j.ijgo.2006.07.002

Cooper, H., & Hedges, L. V. (1994). *Handbook of research synthesis*. Russell Sage Foundation.

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87(3), 442–449.

Crandell, L. (2012). Psychological outcomes of medical versus surgical elective first trimester abortion. *Nursing for Women's Health*, 16(4), 296–307. https://doi.org/10.1111/j.1751-486X.2012.01748.x

David, H. P., Rasmussen, N. K., & Holst, E. (1981). Postpartum and postabortion psychotic reactions. *Family Planning Perspectives*, 13(2), 88. https://doi.org/10.2307/2134703

Derogatis, L. R. (1992). *SCL-90-R: Administration, scoring & procedures manual-II, for the R (Revised) version and other instruments of the psychopathology rating scale series* (2nd ed.). Clinical Psychometric Research Inc.

Dingle, K., Alati, R., Clavarino, A., Najman, J. M., & Williams, G. M. (2008). Pregnancy loss and psychiatric disorders in young women: An Australian birth cohort study. *British Journal of Psychiatry*, 193(6), 455–460. https://doi.org/10.1192/bjp.bp.108.055079

Dobkin, L. M., Gould, H., Barar, R. E., Ferrari, M., Weiss, E. I., & Foster, D. G. (2014). Implementing a prospective study of women seeking abortion in the United States: Understanding and overcoming barriers to recruitment. *Women's Health Issues*, 24, e115–e123. https://doi.org/10.1016/j.whi.2013.10.004

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., Thomas, J., Agoritsas, T., Hilton, J., Perron, C., Akl, E., Hodder, R., Pestridge, C., … Pearson, L. (2017). Living systematic review: 1. Introduction—The why, what, when, and how. *Journal of Clinical Epidemiology*, 91, 23–30. https://doi.org/10.1016/j.jclinepi. 2017.08.010

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P. T., Mavergames, C., & Gruen, R. L. (2014). Living systematic reviews: An emerging opportunity to narrow the evidence-practice gap. *PLoS Medicine*, 11(2), e1001603. https://doi.org/10.1371/journal.pmed. 1001603

Fergusson, D. M., Horwood, L. J., & Boden, J. M. (2008). Abortion and mental health disorders: Evidence from a 30-year longitudinal study. *British Journal of Psychiatry*, 193(6), 444–451. https://doi.org/10. 1192/bjp.bp.108.056499

Fergusson, D. M., Horwood, L. J., & Boden, J. M. (2013). Does abortion reduce the mental health risks of unwanted or unintended pregnancy? A re-appraisal of the evidence. *The Australian and New Zealand Journal of Psychiatry*, 47(9), 819–827.

Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2006). Abortion in young women and subsequent mental health. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 47(1), 16–24. https://doi.org/10.1111/j.1469-7610.2005.01538.x

Finer, L. B., & Zolna, M. R. (2011). Unintended pregnancy in the United States: Incidence and disparities, 2006. *Contraception*, 84(5), 478–485. https://doi.org/10.1016/j.contraception.2011.07.013

Finer, L. B., & Zolna, M. R. (2016). Declines in unintended pregnancy in the United States, 2008–2011. *New England Journal of Medicine*, 374(9), 843–852. https://doi.org/10.1056/NEJMsa1506575

Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in metaanalysis. https://Cran.R-project.Org/ Web/Packages/Robumeta/Vignettes/robumetaVignette.Pdf

Foster, D. G. (2020). *The Turnaway Study: Ten years, a thousand women, and the consequences of having—Or being denied—An abortion*. Scribner.

Ganatra, B., Gerdts, C., Rossier, C., Johnson, B. R., Tunçalp, Ö., Assifi, A., Sedgh, G., Singh, S., Bankole, A., Popinchalk, A., Bearak, J., Kang, Z., & Alkema, L. (2017). Global, regional, and subregional classification of abortions by safety, 2010–14: Estimates from a Bayesian hierarchical model. *The Lancet*, 390(10110), 2372–2381. https:// doi.org/10.1016/S0140-6736(17)31794-4

Gilchrist, A. C., Hannaford, P. C., Frank, P., & Kay, C. R. (1995). Termination of pregnancy and psychiatric morbidity. *British Journal of Psychiatry*, 167(2), 243–248. https://doi.org/10.1192/bjp.167.2.243

Gissler, M., Hemminki, E., & Lonnqvist, J. (1996). Suicides after pregnancy in Finland, 1987–94: Register linkage study. *BMJ*, 313(7070), 1431–1434. https://doi.org/10.1136/bmj.313.7070.1431

Gøtzsche, P. C., Hróbjartsson, A., Marić, K., & Tendal, B. (2007). Data extraction errors in meta-analyses that use standardized mean differences. *Journal of the American Medical Association*, 298(4), 430–437. https://doi.org/10.1001/jama.298.4.430

Grainger, M., Stewart, G., & Haddaway, N. R. (2022). Why "vote-counting" is never acceptable in evidence synthesis. *Open Science Framework*. https://doi.org/10.31219/osf.io/c49uh

Grupo Médico por el Derecho a Decidir – Columbia. (2011). *Unwanted pregnancy, forced continuation of pregnancy and effects on mental health: Position paper*. Global Doctors for Choice (GDC) Network. https://Globaldoctorsforchoice.Org/Wp-Content/Uploads/ Unwanted-Pregnancy-Forced-Continuation-of-Pregnancy-and- Effects-on-Mental-Health-v2.Pdf

Guttmacher Institute. (2021). For the first time ever, U.S. States enacted more than 100 abortion restrictions in a single year. Retrieved April 13, 2023, from https://Www.Guttmacher.Org/Article/2021/10/ First-Time-Ever-Us-States-Enacted-More-100-Abortion-Restrictions- Single-Year?gclid=CjwKCAjw_L6LBhBbEiwA4c46uuN12Tdm-gyBG_ hYzfNMmA5HwHleBK8_-w6XdjULZ3iWrPswbrmWVRoC608QAv D_BwE

Harris, L. F., Roberts, S. C., Biggs, M. A., Rocca, C. H., & Foster, D. G. (2014). Perceived stress and emotional social support among women who are denied or receive abortions in the United States: A prospective cohort study. *BMC Women's Health*, 14(1), 76. https:// doi.org/10.1186/1472-6874-14-76

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions, version 5.1.0*. John Wiley & Sons.

Institute of Medicine. (2011). *Finding what works in health care: Standards for systematic reviews*. The National Academies Press. http://Www. Nap.Edu/Catalog.Php?Record_id=13059

Joshi, M., Pustejovsky, J. E., & Beretvas, S. N. (2022). Cluster wild bootstrapping to handle dependent effect sizes in meta-analysis with a small number of studies. *Research Synthesis Methods*, 13(4), 457–477. https://doi.org/10.1002/jrsm.1554

Jüni, P. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association*, 282, 1054–1060.

Juni, P. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*, 323(7303), 42–46. https://doi.org/10. 1136/bmj.323.7303.42

Kacsmaryk, M. J. (2023). Memorandum Opinion and Order, Alliance for Hippocratic Medicine et al. V. U.S. Food and Drug Administration et al. United States District Court for the Nothern District of Texas Amarillo

Division, Case 2:22-CV-00223-Z, Document 137. Retrieved June 8, 2023, from https://Storage.Courtlistener.Com/Recap/Gov.Uscourts.Txnd.370067/Gov.Uscourts.Txnd.370067.137.0_8.Pdf

Koop, C. E. (1989). The US surgeon general on the health effects of abortion. *Population and Development Review*, *15*(1), 172–175. https://doi.org/10.2307/1973425

Kost, K., Zolna, M., & Murro, R. (2023). Pregnancies in the United States by desire for pregnancy: Estimates for 2009, 2011, 2013, and 2015. *Demography*, *60*(3), 837–863. https://doi.org/10.1215/00703370-10690005

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A. M. K., Hammerstrøm, K., & Sathe, N. (2017). Searching for studies: A guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews*, *13*(1), 1–73. https://doi.org/10.4073/cmg.2016.1

Lee, E. (2003). *Abortion, motherhood, and mental health: Medicalizing reproduction in the United States and Great Britain*. Aldine de Gruyter.

Lee, J., & Beretvas, S. N. (2023). Comparing methods for handling missing covariates in meta-regression. *Research Synthesis Methods*, *14*(1), 117–136. https://doi.org/10.1002/jrsm.1585

Lefebvre, C., Glanville, J., Briscoe, S., Littlewood, A., Marshall, C., Metzendorf, M. I., Noel-Storr, A., Rader, T., Shokraneh, F., Thomas, J., Wieland, L. S., & the Cochrane Information Retrieval Methods Group. (2019). Searching for and selecting studies. In J. P. T. Higgins, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 67–107). Wiley. https://doi.org/10.1002/9781119536604.ch4

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Linna, M. S., Raevuori, A., Haukka, J., Suvisaari, J. M., Suokas, J. T., & Gissler, M. (2013). Reproductive health outcomes in eating disorders. *International Journal of Eating Disorders*, *46*(8), 826–833. https://doi.org/10.1002/eat.22179

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.

Littell, J. H., & Coyne, J. C. (2012). Abortion and mental health: Guidelines for proper scientific conduct ignored. *British Journal of Psychiatry*, *200*(1), 75–76. https://doi.org/10.1192/bjp.200.1.75

Littell, J. H., Pigott, T. D., Nilsen, K. H., Green, S. J., & Montgomery, O. L. K. (2021). Multisystemic Therapy® for social, emotional, and behavioural problems in youth age 10 to 17: An updated systematic review and meta-analysis. *Campbell Systematic Reviews*, *17*(4), e1158. https://doi.org/10.1002/cl2.1158

Littell, J. H., Pigott, T. D., Nilsen, K. H., Roberts, J., & Labrum, T. K. (2023). Functional family therapy for families of youth (age 11–18) with behaviour problems: A systematic review and meta-analysis. *Campbell Systematic Reviews*, *19*, e1324. https://doi.org/10.1002/cl2.1324

Littell, J. H., Popa, M., & Forsythe, B. (2005). Multisystemic therapy for social, emotional, and behavioral problems in youth aged 10–17. *Campbell Systematic Reviews*, *1*(1), 1–63. https://doi.org/10.4073/csr.2005.1

Lohr, P. A., Fjerstad, M., DeSilva, U., & Lyus, R. (2014). Abortion. *BMJ*, *348*, f7553.

López-López, J. A., Page, M. J., Lipsey, M. W., & Higgins, J. P. T. (2018). Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research Synthesis Methods*, *9*(3), 336–351. https://doi.org/10.1002/jrsm.1310

Major, B., Appelbaum, M., Beckman, L., Dutton, M. A., Russo, N. F., & West, C. (2009). Abortion and mental health: Evaluating the evidence. *American Psychologist*, *64*(9), 863–890.

Major, B., Cozzarelli, C., Cooper, M. L., Zubek, J., Richards, C., Wilhite, M., & Gramzow, R. H. (2000). Psychological responses of women after first-trimester abortion. *Archives of General Psychiatry*, *57*(8), 777. https://doi.org/10.1001/archpsyc.57.8.777

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *The Lancet*, *354*, 1896–1900. https://doi.org/10.1016/s0140-6736(99)04149-5

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097.

Moseson, H., Bullard, K. A., Cisternas, C., Grosso, B., Vera, V., & Gerdts, C. (2020). Effectiveness of self-managed medication abortion between 13 and 24 weeks gestation: A retrospective review of case records from accompaniment groups in Argentina, Chile, and Ecuador. *Contraception*, *102*(2), 91–98. https://doi.org/10.1016/j.contraception.2020.04.015

Moseson, H., Jayaweera, R., Egwuatu, I., Grosso, B., Kristianingrum, I. A., Nmezi, S., Zurbriggen, R., Bercu, C., Motana, R., & Gerdts, C. (2023). Effectiveness of self-managed medication abortion between 9 and 16 weeks of gestation. *Obstetrics and Gynecology*, *142*(2), 330–338. https://doi.org/10.1097/AOG.0000000000005266

Moseson, H., Jayaweera, R., Egwuatu, I., Grosso, B., Kristianingrum, I. A., Nmezi, S., Zurbriggen, R., Motana, R., Bercu, C., Carbone, S., & Gerdts, C. (2022). Effectiveness of self-managed medication abortion with accompaniment support in Argentina and Nigeria (SAFE): A prospective, observational cohort study and non-inferiority analysis with historical controls. *The Lancet Global Health*, *10*(1), e105–e113. https://doi.org/10.1016/S2214-109X(21)00461-7

Munk-Olsen, T., Laursen, T. M., Pedersen, C. B., Lidegaard, Ø., & Mortensen, P. B. (2011). Induced first-trimester abortion and risk of mental disorder. *New England Journal of Medicine*, *364*(4), 332–339. https://doi.org/10.1056/NEJMoa0905882

National Academies of Sciences, Engineering, and Medicine (NASEM). (2018). *The safety and quality of abortion care in the United States*. National Academies Press. https://doi.org/10.17226/24950

National Collaborating Centre for Mental Health (NCCMH). (2011). *Induced abortion and mental health: A systematic review of the mental health outcomes of induced abortion, including their prevalence and associated factors*. Academy of Medical Royal Colleges. Retrieved May 10, 2023, from https://Www.Aomrc.Org.Uk/Wp-Content/Uploads/2016/05/Induced_Abortion_Mental_Health_1211.Pdf

Nelson, H. D., Darney, B. G., Ahrens, K., Burgess, A., Jungbauer, R. M., Cantor, A., Atchison, C., Eden, K. B., Goueth, R., & Fu, R. (2022). Associations of unintended pregnancy with maternal and infant health outcomes: A systematic review and meta-analysis. *Journal of the American Medical Association*, *328*(17), 1714. https://doi.org/10.1001/jama.2022.19097

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pedersen, W. (2007). Childbirth, abortion and subsequent substance use in young women: A population-based longitudinal study. *Addiction*, *102*(12), 1971–1978. https://doi.org/10.1111/j.1360-0443.2007.02040.x

Pedersen, W. (2008). Abortion and depression: A population-based longitudinal study of young women. *Scandinavian Journal of Public Health*, *36*(4), 424–428. https://doi.org/10.1177/1403494807088449

Petersen, J. M., Barrett, M., Ahrens, K. A., Murray, E. J., Bryant, A. S., Hogue, C. J., Mumford, S. L., Gadupudi, S., Fox, M. P., & Trinquart, L. (2022). The confounder matrix: A tool to assess confounding bias in systematic reviews of observational studies of etiology. *Research Synthesis Methods*, 13(2), 242–254. https://doi.org/10.1002/jrsm.1544

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing, Ltd.

Popinchalk, A., & Sedgh, G. (2019). Trends in the method and gestational age of abortion in high-income countries. *BMJ Sexual & Reproductive Health*, 45(2), 95–103. https://doi.org/10.1136/bmjsrh-2018-200149

Pustejovsky, J. E. (2019). Sometimes, aggregating effect sizes is fine. *Blog Post*. https://Www.Jepusto.Com/Sometimes-Aggregating-Effect-Sizes-Is-Fine/

Pustejovsky, J. (2023). clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. R package version 0.5.10, https://CRAN.R-project.org/package=clubSandwich

Pustejovsky, J. E., Zhang, J., & Tipton, E. (2023). Equity-related moderator analysis in syntheses of dependent effect sizes: Conceptual and statistical considerations. Presentation at the Society for Research on Educational Effectiveness, Arlington, VA. https://www.jepusto.com/talk/sree-2023-equity-related-moderator-analysis/

Pustejovsky, J. E., & Joshi, M. (2023). Cluster-bootstrapping a meta-analytic selection model. https://www.jepusto.com/cluster-bootstrap-selection-model/

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, 23, 425–438. https://doi.org/10.1007/s11121-021-01246-3

Quenby, S., Gallos, I. D., Dhillon-Smith, R. K., Podesek, M., Stephenson, M. D., Fisher, J., Brosens, J. J., Brewin, J., Ramhorst, R., Lucas, E. S., McCoy, R. C., Anderson, R., Daher, S., Regan, L., Al-Memar, M., Bourne, T., MacIntyre, D. A., Rai, R., Christiansen, O. B., ... Coomarasamy, A. (2021). Miscarriage matters: The epidemiological, physical, psychological, and economic costs of early pregnancy loss. *The Lancet*, 397(10285), 1658–1667. https://doi.org/10.1016/S0140-6736(21)00682-6

Ralph, L., Foster, D. G., Raifman, S., Biggs, M. A., Samari, G., Upadhyay, U., Gerdts, C., & Grossman, D. (2020). Prevalence of self-managed abortion among women of reproductive age in the United States. *JAMA Network Open*, 3(12), e2029245. https://doi.org/10.1001/jamanetworkopen.2020.29245

Reardon, D. C., & Coleman, P. K. (2006). Relative treatment rates for sleep disorders and sleep disturbances following abortion and childbirth: A prospective record-based study. *Sleep*, 29(1), 105–106. https://doi.org/10.1093/sleep/29.1.105

Reardon, D. C., Cougle, J. R., Rue, V. M., Shuping, M. W., Coleman, P. K., & Ney, P. G. (2003). Psychiatric admissions of low-income women following abortion and childbirth. *CMAJ: Canadian Medical Association Journal = journal de l'Association medicale canadienne*, 168(10), 1253–1256.

Reardon, D. C., & Ney, P. G. (2000). Abortion and subsequent substance abuse. *The American Journal of Drug and Alcohol Abuse*, 26(1), 61–75. https://doi.org/10.1081/ADA-100100591

Remez, L., Mayall, K., & Singh, S. (2020). Global developments in laws on induced abortion: 2008–2019. *International Perspectives on Sexual and Reproductive Health*, 46(Suppl. ment 1), 53–65. https://doi.org/10.1363/46e0920

Robinson, G. E., Stotland, N. L., Russo, N. F., Lang, J. A., & Occhiogrosso, M. (2009). Is there an "Abortion Trauma Syndrome"? Critiquing the evidence. *Harvard Review of Psychiatry*, 17(4), 268–290.

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, 26(2), 141–160. https://doi.org/10.1037/met0000300

Rothstein, H., Sutton, A. J., & Bornstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. John Wiley & Sons Ltd.

Rue, V. M. (2014). Clinical dimensions of post-abortion trauma. In *Post-abortion trauma: Possible psychological and existential aftermaths* (pp. 19–146). Pontifical Academy for Life.

Rue, V., & Speckhard, A. (1992). Post-abortion trauma: Incidence and diagnostic considerations. *Medicine & Mind*, 6(1), 57–73.

Russo, N. F. (2014). Abortion, unwanted childbearing, and mental health. *Salud Mental*, 37, 283–291.

Schauer, J. M. (2019). Challenges for imputing missing covariates in meta-regression. Presentation at the annual meeting of the Society for Research Synthesis Methods, Chicago, IL. https://www.jmschauer.com/talk/srsm_2019/

Schauer, J. M., Diaz, K., Pigott, T. D., & Lee, J. (2022). Exploratory analyses for missing data in meta-analyses and meta-regression: A tutorial. *Alcohol and Alcoholism*, 57(1), 35–46. https://doi.org/10.1093/alcalc/agaa144

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, 7, 10. https://doi.org/10.1186/1471-2288-7-10

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358, j4008. https://doi.org/10.1136/bmj.j4008

Siegel, R. B. (2008). The right's reasons: Constitutional conflict and the spread of woman-protective anti-abortion argument. *Duke Law Journal*, 57(6), 1641–1692.

Song, F., Eastwood, A. J., Gilbody, S., Duley, L., & Sutton, A. J. (2000). Publication and related biases. *Health Technology Assessment*, 4(10), 1–115. https://doi.org/10.3310/hta4100

Song, F., Parekh-Bhurke, S., Hooper, L., Loke, Y. K., Ryder, J. J., Sutton, A. J., Hing, C. B., & Harvey, I. (2009). Extent of publication bias in different categories of research cohorts: A meta-analysis of empirical studies. *BMC Medical Research Methodology*, 9(1), 79. https://doi.org/10.1186/1471-2288-9-79

Song, F., Parekh, S., Hooper, L., Loke, Y., Ryder, J., Sutton, A., Hing, C., Kwok, C., Pang, C., & Harvey, I. (2010). Dissemination and publication of research findings: An updated review of related biases. *Health Technology Assessment*, 14(8), 1–220. https://doi.org/10.3310/hta14080

Speckhard, A. C. (1985). *The psycho-social aspects of stress following abortion* [Unpublished PhD dissertation. ProQuest Dissertations and Theses]. University of Minnesota.

Speckhard, A. C., & Rue, V. M. (1992). Postabortion syndrome: An emerging public health concern. *Journal of Social Issues*, 48(3), 95–119. https://doi.org/10.1111/j.1540-4560.1992.tb00899.x

Steinberg, J. R. (2011). Later abortions and mental health: Psychological experiences of women having later abortions—A critical review of research. *Women's Health Issues*, 21(Suppl 3), S44–S48.

Steinberg, J. R., & Finer, L. B. (2011). Examining the association of abortion history and current mental health: A reanalysis of the National Comorbidity Survey using a common-risk-factors model. *Social science & medicine (1982)*, 72(1), 72–82.

Steinberg, J. R., Jordan, B., & Wells, E. S. (2009). Science prevails: Abortion and mental health. *Contraception*, 79(2), 81–83.

Steinberg, J. R., Laursen, T. M., Adler, N. E., Gasse, C., Agerbo, E., & Munk-Olsen, T. (2018). Examining the association of antidepressant

prescriptions with first abortion and first childbirth. *JAMA Psychiatry*, *75*(8), 828. https://doi.org/10.1001/jamapsychiatry.2018.0849

Steinberg, J. R., McCulloch, C. E., & Adler, N. E. (2014). Abortion and mental health: Findings from The National Comorbidity Survey-Replication. *Obstetrics and Gynecology*, *123*(2), 263–270. https://doi.org/10.1097/AOG.0000000000000092

Steinberg, J. R., & Russo, N. F. (2009). Evaluating research on abortion and mental health. *Contraception*, *80*(6), 500–503.

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., ... Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, *355*, i4919. https://doi.org/10.1136/bmj.i4919

Stewart, L., Moher, D., & Shekelle, P. (2012). Why prospective registration of systematic reviews makes sense. *Systematic Reviews*, *1*, 7. https://doi.org/10.1186/2046-4053-1-7

Stotland, N. L. (1992). The myth of the abortion trauma syndrome. *JAMA: The Journal of the American Medical Association*, *268*(15), 2078–2079. https://doi.org/10.1001/jama.1992.03490150130038

Stroup, D. F. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Journal of the American Medical Association*, *283*(15), 2008–2012.

Studnicki, J., Longbons, T., Fisher, J., Reardon, D. C., Skop, I., Cirucci, C. A., Harrison, D. J., Craver, C., Tsulukidze, M., & Ras, Z. (2023). A cohort study of mental health services utilization following a first pregnancy abortion or birth. *International Journal of Women's Health*, *15*, 955–963. https://doi.org/10.2147/IJWH.S410798

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and spss. *Research Synthesis Methods*, *5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. https://doi.org/10.1007/s40865-016-0026-5

The Methods Coordinating Group of the Campbell Collaboration. (2019a). *Methodological expectations of Campbell Collaboration intervention reviews: Conduct standards*. The Campbell Collaboration. https://doi.org/10.4073/cpg.2016.3

The Methods Coordinating Group of the Campbell Collaboration. (2019b). *Methodological expectations of Campbell Collaboration intervention reviews: Reporting standards*. The Campbell Collaboration. https://doi.org/10.4073/cpg.2016.4

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley, Reading.

Upadhyay, U. D., Coplon, L., & Atrio, J. M. (2023). Society of family planning committee statement: Abortion nomenclature. *Contraception*, *126*, 110094. https://doi.org/10.1016/j.contraception.2023.110094

Valentine, J. C., Littell, J. H. & Young, S., (Eds.). (2023). *Systematic reviews and meta-analysis: A Campbell Collaboration online course*. Open Learning Initiative, Carnegie Mellon University. https://Oli.Cmu.Edu/Courses/Systematic-Reviews-and-Meta-Analysis/

Van Ditzhuijzen, J., Ten Have, M., De Graaf, R., Lugtig, P., Van Nijnatten, C. H. C. J., & Vollebergh, W. A. M. (2017). Incidence and recurrence of common mental disorders after abortion: Results from a prospective cohort study. *Journal of Psychiatric Research*, *84*, 200–206. https://doi.org/10.1016/j.jpsychires.2016.10.006

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(1), 1. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W. (2022). I^2 for multilevel and multivariate models. https://www.metafor-project.org/doku.php/tips:i2_multilevel_multivariate

Wallin Lundell, I., Georgsson Öhman, S., Frans, Ö., Helström, L., Högberg, U., Nyberg, S., Sundström Poromaa, I., Sydsjö, G., Östlund, I., & Skoog Svanberg, A. (2013). Posttraumatic stress among women after induced abortion: A Swedish multi-centre cohort study. *BMC Women's Health*, *13*(1), 52. https://doi.org/10.1186/1472-6874-13-52

Watkins, C., & Bennett, I. (2018). A simple method for combining binomial counts or proportions with hazard ratios for evidence synthesis of time-to-event data. *Research Synthesis Methods*, *9*, 352–360. https://doi.org/10.1002/jrsm.1301

Wells, K., & Littell, J. H. (2009). Study quality assessment in systematic reviews of research on intervention effects. *Research on Social Work Practice*, *19*(1), 52–62. https://doi.org/10.1177/1049731508317278

What Works Clearinghouse. (2018a). WWC standards brief: Attrition standard. https://Ies.Ed.Gov/Ncee/Wwc/Docs/Referenceresources/Wwc_brief_attrition_080715.Pdf

What Works Clearinghouse. (2018b). *WWC standards brief: Baseline equivalence*. US Institute of Education Sciences. https://Ies.Ed.Gov/Ncee/Wwc/Docs/Referenceresources/Wwc_brief_baseline_080715.Pdf

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). Retrieved August 24, 2023, from https://Ies.Ed.Gov/Ncee/Wwc/Handbooks

Winters, K. L., Jasso, J., Pustejovsky, J. E., & Byrd, C. T. (2022). Investigating narrative performance in children with developmental language disorder: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research: JSLHR*, *65*(10), 3908–3929. https://doi.org/10.1044/2022_JSLHR-22-00017

World Bank. (2023). Historical and current classification by income. https://Datacatalog.Worldbank.Org/Search/Dataset/0037712/World-Development-Indicators

World Health Organization. (2022). *Abortion care guideline*. World Health Organization.

World Health Organization. (2023). *Clinical practice handbook for quality abortion care*. World Health Organization.

Zandberg, J., Waller, R., Visoki, E., & Barzilay, R. (2023). Association between state-level access to reproductive care and suicide rates among women of reproductive age in the United States. *JAMA Psychiatry*, *80*(2), 127. https://doi.org/10.1001/jamapsychiatry.2022.4394

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.