efpsa
Journal of European Psychology Students

McMurtrie, B. (2023). On the iatrogenic effect of sensitive-topics survey research: A scoping review. *Journal of European Psychology Students, 14*(1), 39–53. https://doi.org/10.59477/jeps.601

LITERATURE REVIEW

On the Iatrogenic Effect of Sensitive-Topics Survey Research: A Scoping Review

Brandon McMurtrie[1]

Self-report surveys are popular among researchers due to their convenience and, recently, the need for social distancing. However, there is concern that surveys investigating sensitive topics, such as suicide, may cause distress or harm to participants. Iatrogenic harm is any negative effect caused by an instrument, intervention, or treatment. I conducted a scoping review of the studies investigating the potential for iatrogenic harm from survey research. Thirty-seven studies were included. Results indicate that for a small subset of participants, sensitive-topics research can cause some distress. Usually, these participants also report that their participation was important and display a willingness to participate further. However, more randomised controlled trials based on power analyses and using validated psychometric outcome measures are needed.

Keywords: sensitive topics, iatrogenic, harm, survey research, ethics

It has long been known that all treatments and research programmes–medical or psychological–involving human participants carry the inherent risk of iatrogenic effects. These are negative effects induced by an intervention such as research or treatment (Krishnan & Kasthuri, 2005). The need to protect participants from iatrogenic harm and to maximise benefit has been enshrined in most ethical codes of conduct for research involving human participants (American Psychological Association, 2002; National Institutes of Health, 1979; World Medical Association, 2008). Consequently, Institutional Review Boards (IRBs) have become wary of approving research that involves sensitive topics such as suicide, abuse, sex, and trauma for fear that the research will cause distress, worsen symptoms, or otherwise cause harm to participants. The potential for harm as a result of paper-based or online surveys pertaining to sensitive topics is an important factor, which an ethical social scientist must consider before conducting research. Research using sensitive-topics surveys is vital to many services and bodies of literature that seek to lower suicide or to lower youth violence and abuse (Ybarra et al., 2009). Some claim that IRBs often inaccurately or incorrectly judge the potential harm of research based on a range of assumptions or socio-political concerns rather than empirical evidence as to the level of risk (Kuyper et al., 2012). Subsequently, many researchers are attempting to empirically investigate and quantify the harm and benefit experienced by participants in sensitive-topics survey research to provide a reliable research base upon which decisions can be based concerning the appropriateness of future research.

The Present Study

The coronavirus pandemic limited the opportunities for in-person research; therefore, researchers had to turn to remote methods for conducting research. Indeed, even before the heightened need for social distancing measures, paper-based surveys were one of the main tools used by social scientists. Such methods are efficient and cheap, and often, survey

---

[1] School of Psychology, Massey University
Corresponding author: Brandon McMurtrie (B.McMurtrie@massey.ac.nz)

research on sensitive topics is best self-reported rather than researcher-administrated, providing participants with a greater sense of anonymity to avoid demand characteristics, reactivity, and deception (Krumpal, 2013). Additionally, online participant pools provide researchers with an efficient means of recruiting participants and provide access to samples and data that are as good as those historically used by social scientists (Behrend et al., 2011).

Given that much of the research performed in psychology, both before the 2019 coronavirus pandemic and after, is performed remotely, it is relevant to focus the study of the iatrogenic harm in sensitive-topics research to that of pen-and-paper and online surveys. Before conducting such research, however, it may be valuable for researchers to have "a lay of the land"– an understanding of what research has already taken place, what common methods and measures are used, what type of samples were used, what type of sensitive-topics surveys were used, and the results that have been found so far. The aim of this study is to perform a full scoping review of the literature investigating the potential iatrogenic effects of paper sensitive-topics survey research. A scoping review methodology was chosen to provide an overview of the design, methods, and results of studies investigating iatrogenic harm from sensitive topics research in order to identify gaps in the literature and inform future studies.

Method

This scoping review is performed according to the five-step methodology laid out by Arksey and O'Malley (2005). The intent of the present study is based on two of the four reasons that Arksey and O'Malley identify for performing a scoping review: to examine the range and nature of research activity and to identify gaps in the existing literature.

Step 1: Identifying the Research Question(s)

This project aims to answer two primary research questions:

1) What have previous studies discovered regarding the effect of answering survey questions about mental health on distress and other emotional states?
2) What are the methodologies used to study this topic throughout the literature?

Step 2: Identifying Relevant Studies

The search for relevant studies began with an informal search of the literature, including database searches and hand searching, to identify the common keywords used in relevant studies. The most common keyword found throughout the search was the technical term "iatrogenic," meaning negative effects induced by an intervention, treatment, or study (Krishnan & Kasthuri, 2005). In the case of the present study, an "iatrogenic" effect refers to any harm caused to a participant in the process of taking part in mental health survey research.

The term iatrogenic is used frequently throughout the literature of clinical psychology and the medical and biological sciences. Likewise, terms such as "mental health", "distress", "negative effects", and "survey" are used throughout psychological research. This presented a problem in that any keyword combination involving "iatrogenic" paired with "harm", "distress", and "research" even in the presence of keywords more relevant to this review, such as "mental health" or "survey research", inevitably returned upwards of tens of thousands of results, often more than 30,000. A very large proportion of the results were studies from unrelated fields or subdisciplines that use the same terms. In the interest of feasibility and efficiency, I decided that for all keyword combinations, I would comb the first 20 pages of results for relevant studies. Once a body of relevant studies had been found from keyword searches, I would then continue the search via hand searching of the reference lists. Given that the body of literature was quite small and that many of the studies identified by keyword were reviews and meta-analyses, I concluded that a hand search following a partial keyword search would cover the available relevant literature well.

Google Scholar was used as the primary database for the search, as it searches the widest

Table 1.

Keywords used and returned results in the database search.

	Keywords	Number of results selected (Duplicates omitted)
1	“Mental Health Survey” and “Iatrogenic” and “Distress”	29
2	“Online Survey” and “Mental Health” and “Suicidality or Distress”	1
3	“Negative Effects” and “Mental Health Survey”	4
4	“Mental Health Survey” and “Effect” and “Distress”	2
5	Hand search of reference lists of previously selected studies.	69
	Total	105

array of databases and journals and covers the major journal repositories. The author’s University database search engine was used, which searches the university’s library as well as multiple online journal databases.

Because I was unable to comb through all returned studies from the keyword search and was instead collecting some of the relevant studies from the first 20 pages of results for a later hand search, I was very broad and generous with the selection criteria for this part of the search, by selecting studies that appeared even slightly relevant. Studies which appeared to be in the field of psychology and which appeared even nominally related to research, mental health and harm of some kind were selected and added to the list for later inspection and hand-searching of reference lists. Google Scholar returns approximately 10 results per page, so for each keyword combination I combed approximately 200 results (approximately 1,000 in total). Table 1 shows the search strategy: the keyword combinations and number of results selected for later inspection and reference list searching, and the subsequent studies found via hand-searching.

Step 3: Study Selection

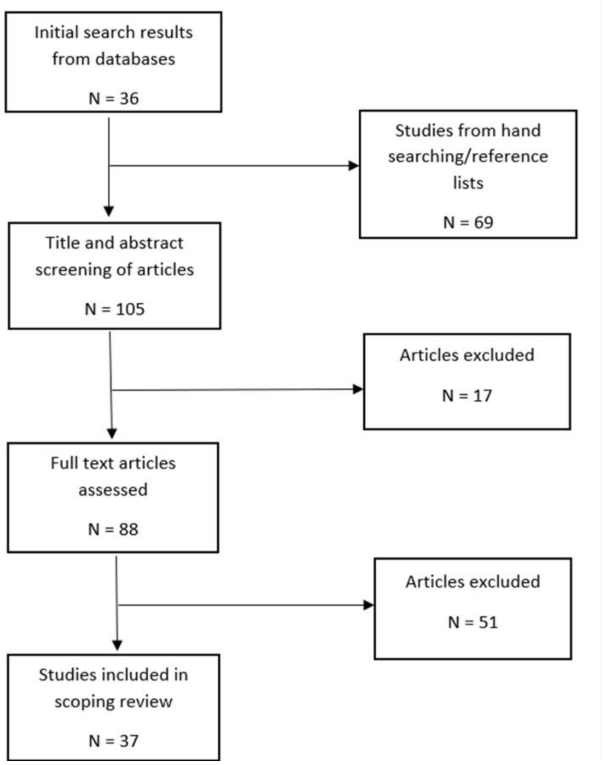
Step 2 resulted in a list of 105 articles that were selected based on the broad selection criteria for that stage. Any studies that appeared to investigate the potential effects of taking part in psychological research were selected. The criteria were necessarily broad for step 2 because most of the search was to be done by hand. Therefore, any slightly relevant studies were included because their reference lists might contain articles that are important to this review. Then, the list of 105 articles contained varied research investigating the iatrogenic effects of clinical screenings using psychological measures, interview research, pen-and-paper surveys, the iatrogenic effects of therapeutic interventions, the effect of school interventions, meta-analyses, and literature reviews.

In step 3, I established a set of exclusion/inclusion criteria. Given the small number of studies, these were similarly simple and broad, as I did not need to cull many studies to make the review more feasible. For studies to be included in this review, they were required to be based on pen-and-paper or online survey research. Given the small number of studies in the list, there was no limit placed on the year of publication. Similarly, no criteria were placed around research design (e.g., experimental vs. observational) or, outcome measures (e.g., focus on studies that measured distress specifically) or independent variables (e.g., I did not limit studies to those that use surveys on a particular sensitive topic such as depression).

Therefore, I selected studies that attempted to investigate the iatrogenic effects of pen-and-paper and online sensitive-topics surveys on participants. Figure 1 shows the PRISMA diagram of the selection process. In total, 68 studies were excluded. Of those excluded, 41 were excluded because they used interviews as the survey method, six were excluded because they investigated the iatrogenic effects of a type of therapy or therapeutic approach, five were excluded because they investigated the effect of an education or training programme, five were excluded because they used other types of stimuli, such as exposure to potentially distressing videos, and 11 because they were entirely unrelated.

Figure 1.

PRISMA diagram showing the number of articles retrieved and then excluded at each stage of the process.



Results

Thirty-seven studies were included in the final sample for this review. Of these, 10 were experimental studies, and 27 were observational. No meta-analyses were included in the sample because they all subsumed both interview and pen-and-paper/online surveys under the same meta-analyses, whereas this review is concerned with remote methods such as pen-and-paper/online surveys.

I was particularly interested in the number and results of experimental studies, as they provide a stronger basis for causal inferences about the iatrogenic effects of the surveys administered. Therefore, I have reviewed the experimental and observational studies separately.

Experimental Studies

Of the 37 studies selected for inclusion in this review, 10 were randomised controlled trials (RCTs) utilising random assignment and control conditions. Seven of the RCTs were performed in the USA (Cook et al., 2015; Ferrier-Auerbach et al., 2009; Gould et al., 2005; Muehlenkamp et al., 2014; Pedersen et al., 2014; Rinehart et al., 2017; Yeater et al., 2012), one was performed in Singapore (Harris & Goh, 2017), one in the Netherlands (de Beurs et al., 2016), and one in Australia (Robinson et al., 2011).

Independent Variables

These 10 RCTs investigated the iatrogenic effects of a range of different sensitive topics surveys such as suicide, non-suicidal self-injury (NSSI), sexual history, violence, substance abuse and PTSD (see Table 2). These studies almost always used batteries of measures– an amalgam of multiple surveys pertaining to different sensitive topics, often combined with non-validated items. For this reason, *Table 2* lists the subject matter of the independent measures used in the surveys rather than the specific measures, as listing all the individual instruments in the table was impractical. For example, Cook et al. (2015) used a survey composed of 300 questions pulled from 10 different instruments pertaining to many different sensitive topics. All studies used well-validated measures that are often used in psychological research, such as the Sexual Experiences Survey (Cook et al., 2015; Pedersen et al., 2014), the PTSD Checklist (Ferrier-Auerbach et al., 2009; Rinehart et al., 2017; Yeater et al., 2012), the Rape Myth Acceptance Scale (Rinehart et al., 2017; Yeater et al., 2012) and many others.

The use of batteries of multiple different sensitive surveys in these studies makes it difficult to ascertain which of the instruments or sensitive topics influenced the mood or distress level of participants. The studies which used a single independent measure were the study by Muehlenkamp et al. (2014), which investigated the effect that the Inventory of Statements About Self Injury (ISAS) had on participants, and the study by de Beurs et al. (2016) which used the Beck Suicidal Ideation scale.

Table 2.

Data table of the 10 experimental studies selected for review.

Author	Sensitive topic of focus	Outcome measured	Power Analysis	Sample
Cook et al. (2015)	Violence, Sex, Trauma	Positive and Negative Affect Scale (PANAS), State-Trait Anxiety, Traumatic Stress Checklist, Reaction to Research Participation Questionnaire (RRPQ)	Yes	559 American undergraduates
de Beurs et al. (2016)	Suicide	PANAS	No	301 Dutch undergraduates
Ferrier-Auerbach et al. (2009)	Trauma, PTSD, Combat Exposure	Self-Assessment Manikin (SAM), non-validated items measuring reactions to the study.	No	154 American adults
Gould et al. (2005)	Suicide, Substance abuse.	Profile of Mood States (POMS), Suicidal Ideation Questionnaire (SIQ)	Yes	2,342 US adolescents, between 13-19 years
Harris & Goh (2017)	Suicide, Death	PANAS, Centre for Epidemiological Studies Depression Scale (CES-D).	No	259 Singaporean adults
Muehlenkamp et al. (2014)	NSSI	Non-validated item rating distress, RRPQ.	No	704 American undergraduates
Pedersen et al. (2014)	Trauma, PTSD, Sex	PTSD Diagnostic Scale, Brief Symptom Inventory, Post-study satisfaction items.	No	202 American undergraduate women
Rinehart et al. (2017)	Sex, Trauma	Global Severity Index from the 90-Item Symptom Checklist.	No	504 American undergraduates
Robinson et al. (2011)	NSSI, Suicide, Psychological distress.	POMS, Suicide Prevention Program Rating Profile.	No	272 Australian high school boys
Yeater et al. (2012)	Trauma, Sex	PANAS, items measuring reaction to research.	No	504 American undergraduates

Outcome measure

There was less variability and complexity among the outcome instruments used to measure the effect of the sensitive surveys. The most common outcome measure used was the Positive and Negative Affect Scale (PANAS; Watson et al., 1988) used in four of the 10 studies, followed by the Profile of Mood States (POMS; Pollock et al., 1979), used in two of the studies (see Table 2).

Not all the studies focused only on the surveys' effect on mood or affect. Others investigated whether suicide research increased the rate of suicidal ideation on the Suicidal Ideation Questionnaire (SIQ) (Gould et al., 2005) or whether the trauma, sexual history, and PTSD survey battery affected PTSD symptom severity (Pedersen et al., 2014).

Samples and Power Analyses

As seen in *Table 2*, six of the studies used samples of undergraduate students (de Beurs et al., 2016; Cook et al., 2015; Muehlenkamp et al., 2014; Pedersen et al., 2014; Rinehart et al., 2017; Yeater et al., 2012), and two of the studies focused on a sample of adolescents (Gould et al., 2005; Robinson et al., 2011). Furthermore, one study focused on a female-only sample (Pedersen et al., 2014), and another one focused on adolescent boys (Robinson et al., 2011).

Only three of the 10 studies provided any indication of awareness or discussion of statistical power for the analyses performed (Cook et al., 2015; Gould et al., 2005; Rinehart et al., 2017). None of these studies describe an actual a priori power analysis. Still, they do show some awareness of the statistical power of their tests or at least describe an effort to maximise statistical power.

Methods

The most common design used in these experiments was a pretest-posttest design (de Beurs et al., 2016; Ferrier-Auerbach et al., 2009; Harris & Goh, 2017; Muehlenkamp et al., 2014). Rinehart et al. (2017) and Yeater et al. (2012) used a post-test-only design.

Three studies used variations of longitudinal designs: Gould et al. (2005) measured participants after an initial survey and then again before a second survey two days later; Cook et al. (2015) measured participants at four different time points, and Pedersen et al. (2014) used a longitudinal design composed of a pre-assessment, a 30-day monitoring period of daily surveys, and a post-assessment. Robinson et al. (2011) used a counterbalanced design in which the control and experimental conditions completed their respective measures on day one, and then the conditions were reversed on day two.

In addition to their main respective research designs, many of the studies used a post-test measure of reactions to research participation, asking participants about their experience with the research (Ferrier-Auerbach et al., 2009; Muehlenkamp et al., 2014; Pedersen et al., 2014; Yeater et al., 2012).

The length of some of the studies using multiple sensitive topics measures may be considered problematic. Yeater et al. (2012) subjected participants to a formidable battery of measures that took two hours or more to complete and covered a range of different topics. It is possible that engaging in sensitive topics research of such duration and complexity might cause distress by the sheer intensity and length of the survey. However, this effect of survey length has not been investigated. In contrast, upon completing such a long survey, participants may have returned to baseline or even been desensitised to the initial distress caused by the sensitive survey. Until the iatrogenic effects of validated measures of singular sensitive topics have been properly investigated, it does not seem appropriate to use multiple surveys of different measures administered simultaneously, as surveys such as these are not likely to be used often in reality and identifying which measure or what aspect of the study caused any harm or distress is extremely difficult.

Findings

While some of the studies did find a statistically significant increase in distress or reduction in positive emotion after completing sensitive topic surveys, the effect sizes were small. None of the studies concluded that sensitive-topics research poses a significant risk to participants or that the risks outweigh the benefits.

Gould et al. (2005) found no difference in distress levels between experimental and control groups at the time of the survey or two days later. The suicide screening survey did not increase suicide ideation or distress in high-risk students compared to the high-risk students in the control condition. There was some evidence that students with depression and those with suicide attempt histories were less distressed than others. Likewise, Harris and Goh (2017) found no significant differences in affect between study conditions and no pre- and post-test affect changes for condition or suicidal participants. Participants with depressive symptoms showed a decrease in positive affect in both sensitive topics and the control conditions. This indicates that the sensitive topics research was not any more distressing than participating in the control survey. Depressive symptoms and family support predicted negative affect changes. Robinson et al. (2011) found that exposure to the sensitive-topics screening questions did not increase distress, even among students previously identified as high-risk. Only 8.9% of students reported finding the questions about self-harm to be either moderately or very distressing. Over 70% of participants found the research to be moderately or very worthwhile, however, those who were identified as at-risk found the research to be less worthwhile than those who were not.

Perhaps somewhat counter-intuitively, Yeater et al. (2012) found that participants who completed trauma and sex surveys had higher positive affect and perceived the study as having greater benefits than those in the control condition. All the participants rated normal life stressors as more distressing than participating in the study. Similarly, Muehlenkamp et al. (2014) found that asking about NSSI did not produce iatrogenic effects and, in fact, may have lowered distress and produced a small decrease in regard to the urge to self-injure. Participants also listed a range of positive emotions toward the research and a willingness to participate again.

In contrast, de Beurs et al. (2016) found that answering surveys about suicide does produce distress in a minority of participants. Of those who did experience distress, 80% were in the experimental group. Similarly, Ferrier-Auerbach et al. (2009) found that after controlling for baseline affect, participants in the experimental group reported significantly higher sadness and tension than those in the control group. However, there was no difference in willingness to complete further research or perceived gain from participating in the research among the conditions. Rinehart et al. (2017) found that all participants had low distress after participation, though those exposed to sensitive surveys did have a significantly higher level of distress, but all participants perceived some benefit from participation. Pedersen et al. (2014) found that daily trauma surveys produced a small increase in distress and PTSD symptoms but that participants' responses to participation indicate that these changes are well-tolerated. Many participants reported no harm from participation, and few reported no benefit. The authors suggest, based on the data and participant-reported benefits, that for some participants, the short-term distress had a cathartic effect. Cook et al. (2015), when controlling for baseline PTSD symptoms and levels of distress, found a small decrease in positive affect immediately after responding to questions about sexual violation. Still, this effect had diminished at two weeks post-participation. These participants and others who responded to questions about stressful events reported greater perceptions of the benefits of research participation.

These studies tend to suggest that while some sensitive topics research may cause –for a minority of participants– a small increase in distress, a reduction in positive affect and/or an increase in negative affect, they do not increase symptoms of suicidal ideation or intent to self-harm. Likewise, even in the presence of increased distress, studies found that participants reported perceiving benefits from research participation, that the research was not more distressing than general life stress, and that willingness to participate was still generally high among participants.

Observational Studies

Of the 37 articles selected for this review, 27 were observational studies. That is, while some utilised statistical controls in their analyses, there were no control conditions and, therefore no random assignment to control/experimental conditions. The majority of these studies were performed in the USA (see Table 3).

Independent Variables

Like the experimental surveys reviewed above, the sensitive-topics surveys used in the observational studies are varied and often used batteries combining surveys of varied focus. Therefore, Table 4 shows the focus of the independent variables used rather than the actual measures used.

Table 3.

The countries in which the 27 observational studies included in this review were performed.

Country	Number of studies in this review
Australia	3
England	1
Estonia and Sweden	1
Netherlands	1
Norway	1
UK	1
USA	19

Table 4.

A simplified data chart of the 27 observational studies selected for review.

Author	Survey Focus	Dependent variable measures	Sample
Batterham et al. (2018)	Suicide	Distress Thermometer.	3,620 Australian adults.
Carter et al. (2020)	Suicide	Self-reported distress, nurse documented changes in mood, behavioural measures of distress, and medication use following the survey.	37 veterans, 18-69 years.
Coppersmith et al. (2021)	Suicide	Severity of suicide intent/desire.	101 adults recruited online.
Cromer et al. (2006)	Trauma	Non-validated items measuring distress and cost-benefit.	Sample 1: 166 women, 74 men. Sample 2: 203 women.
Daugherty & Lawrence (1996)	Personality	Lazarus Stress Questionnaire.	95 undergraduates in military college.
Deeley & Love (2010)	Suicide	Mood monitor, willingness to participate in future studies, qualitative comments.	Study 1: 129 adolescents. Study 2: 71 participants.

Author	Survey Focus	Dependent variable measures	Sample
Edwards et al. (2012)	Violence	RRPQ	232 men at Time 1, and 179 at Time 2. Undergraduates.
Edwards et al. (2013)	Violence	RRPQ	Time 1: 774 women. Time 2: 660 women. Undergraduates.
Edwards et al. (2014)	Domestic violence	RRPQ and open-ended questions pertaining to reaction to research.	940 undergraduate women.
Gibson et al. (2014)	Suicide	Seven open-ended items on survey experience. 18-item Likert questionnaire on suicidal feelings.	113 adults.
Halek et al. (2005)	PTSD	Qualitative responses at the end of the survey, health care utilisation for 8 weeks post-survey.	1,542 veterans.
Hasking et al. (2015)	NSSI, Alcohol abuse, Life events, Self-efficacy.	Two open-ended questions regarding enjoyment and level of upset from the survey.	2,637 adolescents completed the first survey; 1,973 adolescents completed the second.
Hom et al. (2018)	Suicide	Change in response to suicide items over time.	207 U.S. undergraduates.
Johnson & Benight (2003)	Domestic violence	RRPQ	55 women from domestic violence shelters.
Kuyper et al. (2012)	Sex	Distress was assessed with three items, need for help was examined by two items, and Positive feelings were assessed with four items.	889 sexually experienced young people. 14-26 years old.
Langhinrichsen-Rohling et al. (2006)	Suicide, abuse	One question pertaining to distress.	793 high schoolers, 87 middle schoolers, 340 truancy programme students, and 320 probation adolescents.
Mathias et al. (2012)	Suicide	Change in suicidal ideation over time.	170 adolescents completed the initial visit, 159 completed the second, 126 completed the third, 77 completed the fourth, and 54 completed all five.

Author	Survey Focus	Dependent variable measures	Sample
Newman et al. (1999)	Trauma	PTSD checklist. Benefit, expected upset, and regret were measured using the following 5-point Likert items: 1. I gained something positive from filling out this survey 2. Completing this survey upset me more than I expected 3. Had I known in advance what completing this survey would be like for me, I still would have agreed	1,174 women.
Priebe et al. (2010)	Sex, Abuse	Four questions about the participants' experience.	4,356 Estonian and Swedish adolescents.
Rojas & Kinder (2007)	Sex	State-Trait Personality Inventory	250 undergraduates.
Savell et al. (2006)	Sex	State-Trait Personality Inventory	207 female undergraduates.
Shorey, Febres, et al. (2013)	Domestic violence	RRPQ	282 female undergraduates.
Shorey, Zucosky, et al. (2013)	Domestic violence, Abuse	RRPQ	193 male undergraduates.
Skar et al. (2019)	Trauma	Self-reported levels of upset after the survey were scored on a scale ranging from 1 (not upsetting) to 7 (very upsetting).	10,517 youths aged 6-18.
Walker et al. (1997)	Sexual assault	Items pertaining to distress and reactions to the survey.	330 US women.
Whitlock et al. (2013)	NSSI, Suicide	Two Items: "Answering some of the questions on this survey caused me to feel upset" and "Answering some of the questions on this survey caused me to think more deeply about my life".	13,155 students.
Ybarra et al. (2009)	Violence	Five-point Likert scale measuring attitude toward participation.	1,588 American youths 10-15 years old.

The most common type of research investigated for iatrogenic effects was suicide research: 25% of studies focused solely on suicide surveys, and suicide featured in many of the studies which used a battery of surveys (see Table 5).

Table 5.

The focus of the batteries and surveys used as independent measures in the 27 observational studies.

Focus of survey/battery of surveys	Number of studies
Domestic violence	3
Domestic violence and abuse	1
NSSI, Alcohol abuse, life events and self-efficacy	1
NSSI, Suicide	1
Personality	1
PTSD	1
Sex	3
Sex and abuse	1
Sexual assault	1
Suicide	7
Suicide and abuse	1
Trauma	3
Violence	3

Dependent Variables

As seen in *Table 4*, many of the studies use non-validated or open-ended items to measure the effect that the survey had on participants rather than validated measures of distress, etc. Approximately 45% of studies used non-validated items to measure distress or reactions to research participation (see Table 4). Items such as these have not been assessed for validity or reliability in their use as questionnaires for measuring outcomes of research participation, however, they have strong face validity for their intended use. The majority of studies measured distress, level of upset, or change in mood from survey participation. Four studies investigated whether suicide/self-harm surveys increased or exacerbated suicide/self-harm symptoms or intentions (Coppersmith et al., 2021; Gibson et al., 2014; Hom et al., 2018; Mathias et al., 2012). Six studies used the Reactions to Research Participation Questionnaire (RRPQ) as the basis of their outcome measures (Edwards et al., 2012; Edwards et al., 2013; Edwards et al., 2014; Johnson & Benight, 2003; Shorey, Febres, et al., 2013; Shorey, Zucosky, et al., 2013), and others used well known and well-validated measures such as State-Trait Personality Inventory (Rojas & Kinder, 2007; Savell et al., 2006), the PTSD checklist (Newman et al., 1999), and the Lazarus Stress Questionnaire (Daugherty & Lawrence, 1996; see Table 4).

Samples and Power Analyses

None of the studies described any sort of a priori power analysis, and only two studies acknowledged a potential lack of statistical power in passing. Hom et al. (2018) acknowledged a potential lack of statistical power for their tests, and Carter et al. (2020) appear to have conducted a post hoc power analysis and likewise acknowledged a potential lack of statistical power. Most of these studies used reasonably large samples, however, without a power analysis and some idea of effect sizes expected, one cannot know how well the statistical tests performed.

As seen in *Table 4*, 10 of the studies used a sample of undergraduates, eight focused on adolescents and/or children, eight of the samples focused only on female participants and two focused on men.

Methods

Fourteen of the studies used a simple post-test survey design, while eight used a longitudinal/repeated measures design, and five used a pretest-posttest survey design (see Table 6).

Table 6.

Survey designs used in the 27 observational studies.

Author	Study Design
Batterham et al. (2018)	Pretest-Posttest
Carter et al. (2020)	Pretest-posttest
Coppersmith et al. (2021)	Longitudinal
Cromer et al. (2006)	Post-test
Daugherty & Lawrence (1996)	Post-test

Author	Study Design
Deeley & Love (2010)	Post-test, with a repeated measure
Edwards et al. (2012)	Longitudinal
Edwards et al. (2013)	Longitudinal
Edwards et al. (2014)	Post-test
Gibson et al. (2014)	Pretest-posttest
Halek et al. (2005)	Post-test
Hasking et al. (2015)	Post-test, with a repeated measure.
Hom et al. (2018)	Longitudinal
Johnson & Benight (2003)	Post-test
Langhinrichsen-Rohling et al. (2006)	Post-test
Mathias et al. (2012)	Longitudinal
Newman et al. (1999)	Post-test
Priebe et al. (2010)	Post-test
Rojas & Kinder (2007)	Pretest-posttest
Savell et al. (2006)	Pretest-posttest
Shorey, Febres, et al. (2013)	Post-test
Shorey, Zucosky, et al. (2013)	Post-test
Skar et al. (2019)	Post-test
Kuyper et al. (2012)	Longitudinal
Walker et al. (1997)	Post-test
Whitlock et al. (2013)	Post-test
Ybarra et al. (2009)	Post-test

Findings

There is a remarkable homogeneity of findings among most of the observational studies. In general, the studies find no evidence that surveys pertaining to suicide increase intent to self-harm or suicidal ideation (Coppersmith et al., 2021; Hom et al., 2018). However, while Mathias et al. (2012) found that half of the participants experienced a decrease in symptoms, 29% experienced no change, and 21% did experience a small increase in symptoms. Likewise, it seems that participating in the sensitive-topics surveys is sometimes associated with a minor increase in negative mood or distress *for a small portion of participants*, often those who have negative experiences relevant to the survey topic (Johnson & Benight, 2003; Langhinrichsen-Rohling et al., 2006). However, this small increase in negative mood or distress is usually accompanied by a feeling that the study was worthwhile or important (Cromer et al., 2006; Gibson et al., 2014). Usually, those participants who experienced distress did not display a reluctance to participate further, nor did they often regret their participation. On the contrary, participants, including those few who were distressed or upset, often reported benefits from participating in the research (Edwards et al., 2014: Newman et al., 1999). The longitudinal studies herein did not provide evidence that the distress of participating is long-lasting for participants.

While many studies did find a small increase in distress for a minority of participants, some also found no difference, and others found a decrease in distress or negative mood. It is therefore difficult to properly categorise the findings of these studies because most found a small increase in distress/negative affect for some, usually accompanied by feelings that participation was positive and important and that participants would willingly participate further. For example, Batterham et al. (2018) found that there was a significant decline in average stress levels between pre-post-tests. They found that 2.5% of people showed an increase in distress, while 5% experienced a decrease in distress. For Hasking et al. (2015), 74 and 72 percent of participants reported enjoying their participation, whereas 5% reported feeling negatively towards participation. Kuyper et al. (2012) did find that victims of sexual coercion reported higher distress levels after participation in their research; however, they also derived more positive experiences from participation than non-victimised participants. However, Savell et al. (2006) found that participants who reported a history of sexual abuse had no significant increase in distress following the completion of sexually explicit surveys.

Discussion

When researching sensitive topics such as suicide, sexual abuse, and traumatic life events, it is important to understand in advance the effect this may have on your research participants. It has become a priority of ethically conscious researchers and ethics review boards, in keeping with ethical codes of conduct, to know whether such research causes distress or exacerbates symptoms of stress or trauma.

This review focused on studies investigating the iatrogenic effects of pen-and-paper and online sensitive topics survey research. Thirty-seven studies were included for review. There was a clear lack of randomised controlled trials in the literature, with almost 75% of the studies being of observational design. Many of the studies –including the RCTs– used batteries of tests pertaining to a suite of sensitive topics. This makes it difficult for researchers to ascertain which sensitive topics surveys were causing the changes in affect and/or causing distress. There needs to be more investigation into the effect of survey length to clarify whether it was the length and brute force of such long sensitive topics surveys that caused distress.

There was a clear lack of a priori power analyses in the literature, even in the RCTs. While it is true that many of the studies had quite large samples, it is still necessary to understand the expected effect sizes and one's power to detect them.

The lack of experimental studies and the relative abundance of non-experimental studies is a major concern for researchers seeking to evaluate potential risk to participants in a course of study focused on a sensitive topic. These studies do not have the ability to make any causal claims about the surveys' effect on participants and, therefore, cannot make any strong claim as to the relative safety or danger of the surveys. In addition, the frequent use of non-validated outcome measures is a threat to the external validity of many of these studies.

Implications

With the limitations of the research in this review acknowledged, there was little evidence to suggest that online and pen-and-paper survey research on sensitive topics present a threat of harm to participants. While some sensitive topics surveys may cause minor distress, this distress was fleeting, and most participants felt that their participation was worthwhile. Therefore, there is little evidence in the existing literature to suggest that researchers or ethical review boards should be particularly hesitant to engage in sensitive topics research, as the research so far suggests that there is little to no harm associated with such research.

Future Directions

The lack of experimental studies on this topic presents a major gap in the literature and a clear direction for future researchers. There is a need for more RCTs to be performed based on full a priori power analyses and using validated measures of distress. These studies should be performed using single sensitive topics surveys, at least in the beginning, rather than entire batteries of multiple sensitive topics surveys. For experimental studies that find no evidence of an iatrogenic effect, it will be useful for researchers to perform analyses that are suitable for determining "no effect," such as Bayesian methods (Dienes, 2014).

Despite the conclusion that, in general, sensitive-topics survey research is likely safe, it is true that a minority of participants did experience mild increases in distress or symptomology. These results may represent a subset of participants who are particularly vulnerable to iatrogenic effects, and future research should focus on developing methods to identify and protect these particularly vulnerable participants.

Limitations

The main limitation of the present study is that the search method was confined to the first 20 pages of each search string or keyword combination. The fact that all keywords relevant to the present study were highly generic and relevant to most psychological research meant that the number of results returned was highly unwieldy and that the majority of them were irrelevant to the present study. Therefore, I found it appropriate to search the first 20 results pages for studies meeting inclusion criteria (a total of approximately 1,000 results) and, beyond that, to continue via hand searching the reference lists of studies for relevant cited literature. While I am confident that the present review provided adequate coverage of the relevant literature, this search method could fairly be perceived to be a limitation of the study.

Conflicts of Interest

The author declares no conflicts of interest.

Contributorship Statement

The author McMurtrie confirms sole responsibility for the paper.

Acknowledgements

This project was generously supported by Dr. Matt Williams from Massey University, who provided funding and guidance.

References

American Psychological Association. (2002). Ethical principles of psychologist and APA report and code of conduct. Retrieved from http://www.apa.org/ethics/code/code.pdf.

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19-32. https://doi.org/10.1080/1364557032000119616

Batterham, P. J., Calear, A. L., Carragher, N., & Sunderland, M. (2018). Prevalence and predictors of distress associated with completion of an online survey assessing mental health and suicidality in the community. *Psychiatry Research*, *262*, 348–350. https://doi.org/10.1016/j.psychres.2017.08.048

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800–813. https://doi.org/10.3758/s13428-011-0081-0

Carter, S. P., Ammerman, B. A., Gebhardt, H. M., Buchholz, J., & Reger, M. A. (2020). Participant reactions to suicide-focused research. *Crisis*, *41*(5), 367–374. https://doi.org/10.1027/0227-5910/a000650

Cook, S. L., Swartout, K. M., Goodnight, B. L., Hipp, T. N., & Bellis, A. L. (2015). Impact of violence research on participants over time: Helpful, harmful, or neither? *Psychology of Violence*, *5*(3), 314–324. https://doi.org/10.1037/a0038442

Coppersmith, D. D. L., Fortgang, R. G., Kleiman, E. M., Millner, A. J., Yeager, A. L., Mair, P., & Nock, M. K. (2021). Effect of frequent assessment of suicidal thinking on its incidence and severity: High-resolution real-time monitoring study. *The British Journal of Psychiatry*, *220*(1), 41–43. https://doi.org/10.1192/bjp.2021.97

Cromer, L. D., Freyd, J. J., Binder, A. K., DePrince, A. P., & Becker-Blease, K. (2006). What's the risk in asking? Participant reaction to trauma history questions compared with reaction to other personal questions. *Ethics & Behavior*, *16*(4), 347–362. https://doi.org/10.1207/s15327019eb1604_5

Daugherty, T. K., & Lawrence, J. W. (1996). Short-term effects of research participation on college men. *The Journal of Psychology*, *130*(1), 71–77. https://doi.org/10.1080/00223980.1996.9914989

de Beurs, D. P., Ghoncheh, R., Geraedts, A. S., & Kerkhof, A. J. F. M. (2016). Psychological distress because of asking about suicidal thoughts: A randomized controlled trial among students. *Archives of Suicide Research*, *20*(2), 153–159. https://doi.org/10.1080/13811118.2015.1004475

Deeley, S. T., & Love, A. W. (2010). Does asking adolescents about suicidal ideation induce negative mood state? *Violence and Victims*, *25*(5), 677–688. https://doi.org/10.1891/0886-6708.25.5.677

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Edwards, K. M., Gidycz, C. A., & Desai, A. D. (2012). Men's reactions to participating in interpersonal violence research. *Journal of Interpersonal Violence*, *27*(18), 3683–3700. https://doi.org/10.1177/0886260512447576

Edwards, K. M., Probst, D. R., Tansill, E. C., & Gidycz, C. A. (2013). Women's reactions to interpersonal violence research: A longitudinal study. *Journal of Interpersonal Violence*, *28*(2), 254–272. https://doi.org/10.1177/0886260512454721

Edwards, K. M., Sylaska, K. M., & Gidycz, C. A. (2014). Women's reactions to participating in dating violence research: A mixed methodological study. *Psychology of Violence*, *4*(2), 224–239. https://doi.org/10.1037/a0034339

Ferrier-Auerbach, A. G., Erbes, C. R., & Polusny, M. A. (2009). Does trauma survey research cause more distress than other types of survey research? *Journal of Traumatic Stress*, *22*(4), 320–323. https://doi.org/10.1002/jts.20416

Gibson, S., Boden, Z. V. R., Benson, O., & Brand, S. L. (2014). The impact of participating in suicide research online. *Suicide and Life-Threatening Behavior*, *44*(4), 372–383. https://doi.org/10.1111/sltb.12082

Gould, M. S., Marrocco, F. A., Kleinman, M., Thomas, J. G., Mostkoff, K., Cote, J., & Davies, M. (2005). Evaluating iatrogenic risk of youth suicide screening programs. *JAMA*, *293*(13), 1635. https://doi.org/10.1001/jama.293.13.1635

Halek, K., Murdoch, M., & Fortier, L. (2005). Spontaneous reports of emotional upset and health care utilization among veterans with posttraumatic stress disorder after receiving a potentially upsetting survey. *American Journal of Orthopsychiatry*, *75*(1), 142–151. https://doi.org/10.1037/0002-9432.75.1.142

Harris, K. M., & Goh, M. T. T. (2017). Is suicide assessment harmful to participants? Findings from a randomized controlled trial. *International Journal of Mental Health Nursing*, *26*(2), 181–190. https://doi.org/10.1111/inm.12223

Hasking, P., Tatnell, R. C., & Martin, G. (2015). Adolescents' reactions to participating in ethically sensitive research: a prospective self-report study. *Child and Adolescent Psychiatry and Mental Health*, *9*(1). https://doi.org/10.1186/s13034-015-0074-3

Hom, M. A., Stanley, I. H., Rogers, M. L., Gallyer, A. J., Dougherty, S. P., Davis, L., & Joiner, T. E. (2018). Investigating the iatrogenic effects of repeated suicidal ideation screening on suicidal and depression symptoms: A staggered sequential study. *Journal of Affective Disorders*, *232*, 139–142. https://doi.org/10.1016/j.jad.2018.02.022

Johnson, L. E., & Benight, C. C. (2003). Effects of trauma-focused research on recent domestic violence survivors. *Journal of Traumatic Stress*, *16*(6), 567–571. https://doi.org/10.1023/b:jots.0000004080.50361.f3

Krishnan, N. R., & Kasthuri, A. S. (2005). Iatrogenic disorders. *Medical journal, Armed Forces India, 61*(1), 2. https://doi.org/10.1016%2FS0377-1237(05)80107-8

Krumpal, I. (2013) Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant 47*, 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Kuyper, L., de Wit, J., Adam, P., & Woertman, L. (2012). Doing more good than harm? The effects of participation in sex research on young people in the Netherlands. *Archives of Sexual Behavior*, *41*(2), 497–506. https://doi.org/10.1007/s10508-011-9780-y

Langhinrichsen-Rohling, J., Arata, C., O'Brien, N., Bowers, D., & Klibert, J. (2006). Sensitive research with adolescents: Just how upsetting are self-report surveys anyway? *Violence and Victims*, *21*(4), 425–444. https://doi.org/10.1891/0886-6708.21.4.425

Mathias, C. W., Michael Furr, R., Sheftall, A. H., Hill-Kapturczak, N., Crum, P., & Dougherty, D. M. (2012). What's the harm in asking about suicidal ideation? *Suicide and Life-Threatening Behavior*, *42*(3), 341–351. https://doi.org/10.1111/j.1943-278x.2012.0095.x

Muehlenkamp, J. J., Swenson, L. P., Batejan, K. L., & Jarvi, S. M. (2014). Emotional and behavioral effects of participating in an online study of nonsuicidal self-injury. *Clinical Psychological Science*, *3*(1), 26–37. https://doi.org/10.1177/2167702614531579

National Institutes of Health. (1979). The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. Retrieved from http://ohsr.od.nih.gov/guidelines/belmont.html.

Newman, E., Walker, E. A., & Gefland, A. (1999). Assessing the ethical costs and benefits of trauma-focused research. *General Hospital Psychiatry*, *21*(3), 187–196. https://doi.org/10.1016/s0163-8343(99)00011-0

Pedersen, E. R., Kaysen, D. L., Lindgren, K. P., Blayney, J., & Simpson, T. L. (2014). Impact of daily assessments on distress and ptsd symptoms in trauma-exposed women. *Journal of Interpersonal Violence*, *29*(5), 824–845. https://doi.org/10.1177/0886260513505705

Pollock, V., Cho, D. W., Reker, D., & Volavka, J. (1979). Profile of Mood States: The factors and their physiological correlates. *Journal of Nervous and Mental Disease, 167*(10), 612–614. https://doi.org/10.1097/00005053-197910000-00004

Priebe, G., Bäckström, M., & Ainsaar, M. (2010). Vulnerable adolescent participants' experience in surveys on sexuality and sexual abuse: Ethical aspects. *Child Abuse & Neglect*, *34*(6), 438–447. https://doi.org/10.1016/j.chiabu.2009.10.005

Rinehart, J. K., Nason, E. E., Yeater, E. A., & Miller, G. F. (2017). Do some students need special protection from research on sex and trauma? New evidence for young adult resilience in "sensitive topics" research. *The Journal of Sex Research, 54*(3), 273–283. https://doi.org/10.1080/00224499.2016.1156047

Robinson, J., Pan Yuen, H., Martin, C., Hughes, A., Baksheev, G. N., Dodd, S., Bapat, S., Schwass, W., McGorry, P., & Yung, A. R. (2011). Does screening high school students for psychological distress, deliberate self-harm, or suicidal ideation cause distress – and is it acceptable? *Crisis*, *32*(5), 254–263. https://doi.org/10.1027/0227-5910/a000087

Rojas, A., & Kinder, B. N. (2007). Effects of completing sexual questionnaires in males and females with histories of childhood sexual abuse: Implications for institutional review boards. *Journal of Sex & Marital Therapy*, *33*(3), 193–201. https://doi.org/10.1080/00926230701267795

Savell, J. K., Kinder, B. N., & Young, M. S. (2006). Effects of administering sexually explicit questionnaires on anger, anxiety, and depression in sexually abused and nonabused females: Implications for risk assessment. *Journal of Sex & Marital Therapy*, *32*(2), 161–172. https://doi.org/10.1080/00926230500442326

Skar, A. M. S., Ormhaug, S. M., & Jensen, T. K. (2019). Reported levels of upset in youth after routine trauma screening at mental health clinics. *JAMA Network Open*, *2*(5). https://doi.org/10.1001/jamanetworkopen.2019.4003

Shorey, R. C., Febres, J., Brasfield, H., Zucosky, H., Cornelius, T. L., & Stuart, G. L. (2013). Reactions to dating violence research: Do difficulties with distress tolerance increase negative reactions? *Journal of Family Violence*, *28*(5), 479–487. https://doi.org/10.1007/s10896-013-9519-8

Shorey, R. C., Zucosky, H., Febres, J., Brasfield, H., & Stuart, G. L. (2013). Males' reactions to participating in research on dating violence victimization and childhood abuse. *Journal of Aggression, Maltreatment & Trauma*, *22*(4), 348–364. https://doi.org/10.1080/10926771.2013.775987

Walker, E. A., Newman, E., Koss, M., & Bernstein, D. (1997). Does the study of victimization revictimize the victims? *General Hospital Psychiatry*, *19*(6), 403–410. https://doi.org/10.1016/s0163-8343(97)00061-3

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Whitlock, J., Pietrusza, C., & Purington, A. (2013). Young adult respondent experiences of disclosing self-injury, suicide-related behavior, and psychological distress in a web-based survey. *Archives of Suicide Research, 17*(1), 20-32. https://doi.org/10.1080/13811118.2013.748405

World Medical Association. (2008). Declaration of Helsinki. Ethical principles for medical research involving human subjects. Retrieved from http://www.wma.net/en/30publications/10policies/b3/17c.pdf.

Ybarra, M. L., Langhinrichsen-Rohling, J., Friend, J., & Diener-West, M. (2009). Impact of asking sensitive questions about violence to children and adolescents. *Journal of Adolescent Health*, *45*(5), 499–507. https://doi.org/10.1016/j.jadohealth.2009.03.009

Yeater, E., Miller, G., Rinehart, J., & Nason, E. (2012). Trauma and sex surveys meet minimal risk standards. *Psychological Science*, *23*(7), 780–787. https://doi.org/10.1177/0956797611435131