



Baldwin, J., Reuben, A., Newbury, J., & Danese, A. (2019). Agreement between prospective and retrospective measures of childhood maltreatment: a systematic review and meta-analysis. *JAMA Psychiatry*, 76(6), 584-593.
<https://doi.org/10.1001/jamapsychiatry.2019.0097>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1001/jamapsychiatry.2019.0097](https://doi.org/10.1001/jamapsychiatry.2019.0097)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via American Medical Association at [dx.doi.org/10.1001/jamapsychiatry.2019.0097](https://doi.org/10.1001/jamapsychiatry.2019.0097). Please refer to any applicable terms of use of the publisher.

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

Agreement Between Prospective and Retrospective Measures of Childhood Maltreatment

A Systematic Review and Meta-analysis

Jessie R. Baldwin, PhD; Aaron Reuben, MEM; Joanne B. Newbury, PhD; Andrea Danese, MD, PhD

IMPORTANCE Childhood maltreatment is associated with mental illness. Researchers, clinicians, and public health professionals use prospective or retrospective measures interchangeably to assess childhood maltreatment, assuming that the 2 measures identify the same individuals. However, this assumption has not been comprehensively tested.

OBJECTIVE To meta-analyze the agreement between prospective and retrospective measures of childhood maltreatment.

DATA SOURCES MEDLINE, PsycINFO, Embase, and Sociological Abstracts were searched for peer-reviewed, English-language articles from inception through January 1, 2018. Search terms included *child* maltreatment*, *child* abuse*, *child* neglect*, *child bull**, *child* trauma*, *child* advers**, and *early life stress* combined with *prospective** and *cohort*.

STUDY SELECTION Studies with prospective measures of childhood maltreatment were first selected. Among the selected studies, those with corresponding retrospective measures of maltreatment were identified. Of 450 studies with prospective measures of childhood maltreatment, 16 had paired retrospective data to compute the Cohen κ coefficient.

DATA EXTRACTION AND SYNTHESIS Multiple investigators independently extracted data according to PRISMA and MOOSE guidelines. Random-effects meta-analyses were used to pool the results and test predictors of heterogeneity.

MAIN OUTCOMES AND MEASURES The primary outcome was the agreement between prospective and retrospective measures of childhood maltreatment, expressed as a κ coefficient. Moderators of agreement were selected a priori and included the measure used for prospective or retrospective assessment of childhood maltreatment, age at retrospective report, sample size, sex distribution, and study quality.

RESULTS Sixteen unique studies including 25 471 unique participants (52.4% female [SD, 10.6%]; mean [SD] age, 30.6 [11.6] years) contained data on the agreement between prospective and retrospective measures of childhood maltreatment. The agreement between prospective and retrospective measures of childhood maltreatment was poor, with $\kappa = 0.19$ (95% CI, 0.14-0.24; $P < .001$). Agreement was higher when retrospective measures of childhood maltreatment were based on interviews rather than questionnaires ($Q = 4.1521$; $df = 1$; $P = .04$) and in studies with smaller samples ($Q = 4.2251$; $df = 1$; $P = .04$). Agreement was not affected by the type of prospective measure used, age at retrospective report, sex distribution of the sample, or study quality.

CONCLUSIONS AND RELEVANCE Prospective and retrospective measures of childhood maltreatment identify different groups of individuals. Therefore, children identified prospectively as having experienced maltreatment may have different risk pathways to mental illness than adults retrospectively reporting childhood maltreatment. Researchers, clinicians, and public health care professionals should recognize these critical measurement differences when conducting research into childhood maltreatment and developing interventions.

JAMA Psychiatry. 2019;76(6):584-593. doi:10.1001/jamapsychiatry.2019.0097
Published online March 20, 2019.

← Editorial page 567

+ Supplemental content

Author Affiliations: Social, Genetic and Developmental Psychiatry Centre, King's College London, London, United Kingdom (Baldwin, Newbury, Danese); Department of Psychology and Neuroscience, Duke University, Durham, North Carolina (Reuben); Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom (Danese); National and Specialist CAMHS (Child and Adolescent Mental Health Services) Trauma, Anxiety, and Depression Clinic, South London and Maudsley NHS (National Health Service) Foundation Trust, London, United Kingdom (Danese).

Corresponding Author: Andrea Danese, MD, PhD, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, De Crespigny Park, Denmark Hill, London SE5 8AF, United Kingdom (andrea.danese@kcl.ac.uk).

Do prospective and retrospective measures of childhood maltreatment identify the same individuals? This question has captivated psychiatrists and psychologists since the inception of our discipline¹ and still permeates many aspects of our professions. Researchers use retrospective reports as a shortcut to better understand the consequences of childhood maltreatment without the significant time or financial investment needed to undertake cohort studies.² Clinicians use retrospective reports to swiftly identify individuals who are at heightened risk of mental illness by virtue of their exposure to childhood maltreatment.³ Public health professionals use retrospective reports to pragmatically estimate the health burden associated with exposure to childhood maltreatment.⁴ All these practices rely on the assumption that retrospective reports and prospective measures identify the same, or at least similar, groups of individuals. However, qualitative reviews^{5,6} have raised concerns about the validity of this assumption. Herein we present, to our knowledge, the first quantitative assessment of the agreement between retrospective reports and prospective measures of childhood maltreatment.

Methods

Data Sources

We performed a systematic review and meta-analysis in line with the PRISMA and MOOSE guidelines, following an a priori-defined protocol (eMethods and eTables 1 and 2 in the [Supplement](#)). We searched MEDLINE, PsycINFO, Embase, and Sociological Abstracts for peer-reviewed articles written in English and published from database inception to January 1, 2018, that included prospective assessments of childhood maltreatment. We used the following search terms: *child* maltreatment*, *child* abuse*, *child* neglect*, *child bull**, *child* trauma*, *child* advers**, and *early life stress* combined with *prospective** and *cohort*.

Study Selection

Two authors (J.R.B. and A.R.) independently screened titles and abstracts of all articles retrieved from the search before reviewing the full text of potentially eligible studies. We included original, peer-reviewed articles with prospectively collected information on childhood maltreatment (age <18 years). Measures of maltreatment (sexual abuse, physical abuse, emotional abuse, and neglect), domestic violence, bullying, institutionalization, and broader measures of adverse childhood experiences that included maltreatment were used to define overall childhood maltreatment. From the articles with prospective assessment of childhood maltreatment, we selected studies with data on corresponding retrospective measures (defined as subsequent assessment of the same individuals' exposure undertaken at any age).

Data Extraction

Three authors (J.R.B., A.R., and J.B.N.) independently extracted data from all studies with prospective assessment of childhood maltreatment on sample characteristics (cohort

Key Points

Question What is the agreement between prospective and retrospective measures of childhood maltreatment?

Findings This systematic review and meta-analysis of 16 unique studies and 25 471 unique participants found poor agreement between prospective and retrospective measures of childhood maltreatment, with Cohen $\kappa = 0.19$. On average, 52% of individuals with prospective observations of childhood maltreatment did not retrospectively report it, and likewise, 56% of individuals retrospectively reporting childhood maltreatment did not have concordant prospective observations.

Meaning Because findings from this meta-analysis demonstrated that prospective and retrospective measures of childhood maltreatment identify largely different groups of individuals, the 2 measures cannot be used interchangeably to study the associated health outcomes and risk mechanisms.

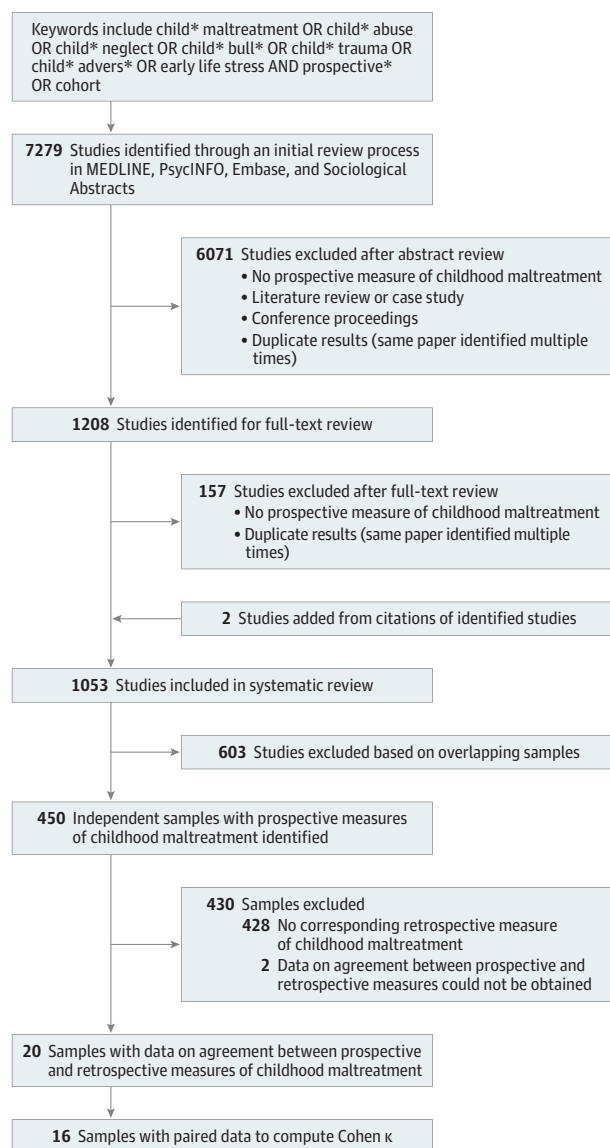
name, sample size, location, age at latest assessment, and sex distribution), childhood maltreatment type(s) assessed, prospective measure type(s) (official records, interview, and questionnaire), source (child protection services, hospital records, parent, child, teacher, or multiple informants), and availability of retrospective measures. If retrospective measures of childhood maltreatment were available, 2 authors (J.R.B. and A.D.) subsequently extracted data on the retrospective measurement type (interview or questionnaire) and source, agreement between prospective and retrospective measures, and study quality. Inconsistencies were resolved in consensus meetings and confirmed with the authors of the primary studies when necessary. Relevant missing information was requested from authors.

Statistical Analysis

The extracted data were converted to contingency tables comparing prospectively identified childhood maltreatment (yes or no) with retrospectively reported childhood maltreatment (yes or no). From the contingency tables, we derived estimates of prevalence, raw percentage of agreement between measures, and Cohen κ coefficient. Some studies only reported a κ . Prevalence and raw percentage of agreement estimates were used for descriptive purposes, and our primary outcome was the κ . The other extracted variables were used to explain the heterogeneity in the κ s across studies.

We first described the prevalence of childhood maltreatment based on prospective and retrospective measures of childhood maltreatment. We then examined the (1) prevalence of retrospective reports of childhood maltreatment among those with prospective observations, (2) prevalence of prospective observations among those with retrospective reports, and (3) raw percentage of agreement between the 2 measures through meta-analyses of proportions for different childhood maltreatment types with the metafor R package.⁷ Data from contingency tables were first converted using the Freeman-Tukey double arcsine transformation⁸ to normalize and stabilize the variance of the sampling distribution, then aggregated using random-effects model meta-analyses, and finally back-transformed using the inverse of the Freeman-Tukey double

Figure 1. Study Selection for Meta-analysis of the Agreement Between Prospective and Retrospective Measures of Childhood Maltreatment



A complete list of the studies included in the analysis with κ agreement is found in the Table.

arcsine transformation.⁸ To display the overlap between prospective and retrospective measures of child maltreatment based on these meta-analyses, we built Venn diagrams using the VennDiagram R package.⁹ To build Venn diagrams, we let the relative complements (the prevalence of retrospective reports without prospective observations [R – P] and the prevalence of prospective observations without retrospective reports [P – R]) vary while holding the intersection (R∩P or the prevalence of concordant retrospective reports and prospective observations) constant.

Because the raw percentage of agreement can be inflated by chance, we derived a measure of agreement based on the κ , which accounts for chance findings and provides an esti-

mate of variation in agreement in the population.¹⁰ The κ s for each study were derived from contingency tables using the `cohen.kappa()` command from the `psych` R package,¹¹ which computes CIs based on the variance estimates discussed by Fleiss et al.¹² The meta-analyses of κ s were undertaken with the `metafor` R package using a random-effects model. When a study reported multiple effect sizes for different types of maltreatment, we calculated the mean of multiple κ s to generate a single overall effect size for each study. We also undertook a sensitivity analysis selecting the largest κ from each study to assess the upper limit of agreement.

We next explored the effects of various possible sources of artifact or bias on κ estimates using the `metafor` R package.⁷ First, we assessed heterogeneity between studies using the I^2 statistic. Second, we assessed the presence of publication bias visually by funnel plot and formally by funnel plot-based tests, such as the Begg and Egger tests. Because these tests might be underpowered if only a few studies are available, we used a nonparametric trim-and-fill procedure to identify and correct for funnel plot asymmetry and reestimated the aggregate results. Third, we assessed the undue effect of individual studies on the meta-analysis results through jackknife sensitivity analyses, by testing changes in the estimate across permutations in which each study was omitted in turn.

Finally, we tested predictors of heterogeneity in κ s. We used subgroup analyses to test the contribution of measurement characteristics (ie, measure used for prospective or retrospective assessment of maltreatment, type of childhood maltreatment). We also used metaregression analyses to test the contribution of sample characteristics (ie, sex distribution, age at retrospective report, sample size, and study quality) (eTable 4 in the Supplement for coding). A 2-tailed $P < .05$ was considered statistically significant.

Results

Search Results

The study selection procedure is summarized in Figure 1, and further details are provided in the eResults in the Supplement. We identified 450 independent studies with prospective measures of childhood maltreatment (eTable 5 in the Supplement). Of these studies, we identified 20 studies (26 365 participants) with at least partial data on the agreement between prospective and retrospective measures of childhood maltreatment and 16 unique studies¹³⁻²⁸ with 25 471 unique participants (52.4% female [SD, 10.6%]; mean [SD] age, 30.6 [11.6] years) with direct measures or paired data sufficient to compute measures of κ s. Details of these studies are reported in the Table.

Overlap Between Individuals Identified by Prospective or Retrospective Measures of Childhood Maltreatment

eFigure 1 in the Supplement displays the range of prevalence estimates for childhood maltreatment based on 32 paired prospective and retrospective measures extracted from 15 studies.^{14-20,22-28,31} Capitalizing on the paired nature of the data, we next analyzed (1) the prevalence of retrospective reports of

Table. Description of Studies With Prospective and Retrospective Measures of Childhood Maltreatment

Source (Study Name)	No. of Participants	Female, %	Sample Description	Type of Maltreatment Assessed	Prospective Measure ^a		Retrospective Measure		Agreement Data		
					Type	Source	Type	Source	Age at Report, y	k	Paired Measures
Robins, ²⁹ 1966 (Deviant Children Grown Up)	411	NA	Children with antisocial behavior problems and matched controls from St Louis, Missouri	Maltreatment	Mixed	CPS, self	Interview	Self	45	No	No
Williams, ³⁰ 1994	129	100	Sexually abused females from a Northeastern US city examined at a hospital, 1973-1975	Sexual abuse	Mixed	Medical records, self, parent	Interview	Self	24.5	No	No
Widom and Shepard, ³¹ 1996; Widom and Morris, ¹³ 1997; Raphael et al, ³² 2001	1181-1196	48.7	Maltreated children and matched controls from a metropolitan area in the Midwestern United States	Maltreatment, physical abuse, and sexual abuse	Records	CPS	Interview	Self	28.7-29.2	Yes	Yes
Johnson et al, ¹⁴ 1999 (Children in the Community Study)	639	47.7	Children from 2 counties in northern New York State	Maltreatment	Records	CPS	Interview	Self	22.3	Yes	Yes
Goodman et al, ³³ 2003	175	80.6	Sexually abused children referred from DAS' offices in Denver, Colorado, 1985-1987	Sexual abuse	Records	CPS	Interview	Self	23	No	No
Tajima et al, ¹⁵ 2004 (Lehigh Longitudinal Study)	409	45.7	Children from 2 counties in eastern Pennsylvania with overrepresentation from low-income families	Physical abuse	Interview	Parent	Interview	Self	18	Yes	Yes
White et al, ²⁶ 2007 (Rutgers Health and Human Development Project)	359	50.7	Adolescents from New Jersey recruited through random telephone sampling	Physical abuse	Questionnaire	Self	Questionnaire	Self	30.5	Yes	Yes
Everson et al, ¹⁶ 2008 (Longitudinal Studies of Child Abuse and Neglect)	348-350	51	Children from Eastern and Southeastern United States identified to be high risk for poor medical or developmental outcomes	Physical, sexual, and emotional abuse	Records	CPS	Interview	Self	12	Yes	Yes
Shaffer et al, ¹⁷ 2008 (Minnesota Longitudinal Study of Parents and Children)	125-170	47.1	Children born to pregnant women with low socioeconomic status who attended prenatal care in Minnesota	Maltreatment, physical and sexual abuse, and neglect	Mixed	CPS, self, parent, teacher	Interview	Self	19	Yes	Yes
Barnes et al, ³⁴ 2009	179	100	Sexually abused females and matched controls in Washington, DC	Sexual abuse	Records	CPS	Interview	Self	24	No	No
Scott et al, ¹⁸ 2010 (The New Zealand Mental Health Survey)	2144	55.1	New Zealand, Maori, and Pacific adults in New Zealand	Maltreatment	Records	CPS	Interview	Self	21.9	Yes	Yes
Denholm et al, ¹⁹ 2013 (National Child Development Study) ^b	8461	NA	Children from Great Britain born in a single week in 1958	Neglect	Multiple	Parent, teacher	Questionnaire	Self	45	Yes	Yes
Elwyn and Smith, ²⁰ 2013 (Rochester Youth Development Study)	846	27.1	Children from Rochester, New York, with overrepresentation of boys and students from high-crime census tracts	Maltreatment	Records	CPS	Interview	Self	22.7	Yes	Yes
Patten et al, ²¹ 2015 (National Longitudinal Study of Children and Youth/National Population Health Survey)	1977	48.3	Children from Canada assessed as part of 2 linked surveys in childhood and adulthood	ACEs	Interview	Parent	Interview	Self	NA	Yes	No

(continued)

Table. Description of Studies With Prospective and Retrospective Measures of Childhood Maltreatment (continued)

Source (Study Name)	No. of Participants	Female, %	Sample Description	Type of Maltreatment Assessed	Prospective Measure ^a		Retrospective Measure		Agreement Data		
					Type	Source	Type	Source	Age at Report, y	Paired Measures	
Plant et al, ²⁷ 2015 (South London Child Development Study) ^b	97	52.4	Children born to pregnant women who attended antenatal clinics in South London, UK	Physical and sexual abuse	Interview	Parent, self	Questionnaire	Self	25	Yes	Yes
Mills et al, ²² 2016 (Mother-University of Queensland Study of Pregnancy)	3739	57.3	Children born to pregnant women who attended maternity services at a hospital in Brisbane, Australia	Sexual abuse	Records	CPS	Questionnaire	Self	21	Yes	Yes
Reuben et al, ²³ 2016 (Dunedin Multidisciplinary Health and Development Study)	950	49.9	Children born in Dunedin, New Zealand, in 1972-1973	Physical, sexual, and emotional abuse, neglect, and ACEs	Mixed	CPS, parent, research workers, teacher	Interview	Self	38	Yes	Yes
Shenk et al, ²⁴ 2016 (Female Adolescent Development Study) ^b	514	100	Maltreated females and matched controls from an area in the US Midwest	Maltreatment	Records	CPS	Interview	Self	15.7	Yes	Yes
Newbury et al, ²⁵ 2018 (E-Risk Longitudinal Twin Study)	2055	51	Twin children from England and Wales born in 1994-1995	Physical, sexual, and emotional abuse, neglect, and maltreatment	Interview	Parent, research worker	Interview	Self	18	Yes	Yes
Naicker et al, ²⁸ 2017 (Birth to Twenty Plus Cohort)	1506-1565	51.8	Children born in Soweto- Johannesburg, South Africa, during 7 weeks in 1990	Sexual, physical, and emotional abuse	Questionnaire	Self	Questionnaire	Self	23	Yes	Yes

Abbreviations: ACEs, adverse childhood experiences; CPS, child protection services; DA, district attorney; NA, not available.

^a "Mixed" refers to studies assessing childhood maltreatment using a combination of records and interviews or questionnaires; "multiple" refers to studies assessing childhood maltreatment using interviews and

questionnaires (excluding records).

^b Data on agreement between prospective and retrospective measures were obtained from study authors rather than from the published article.

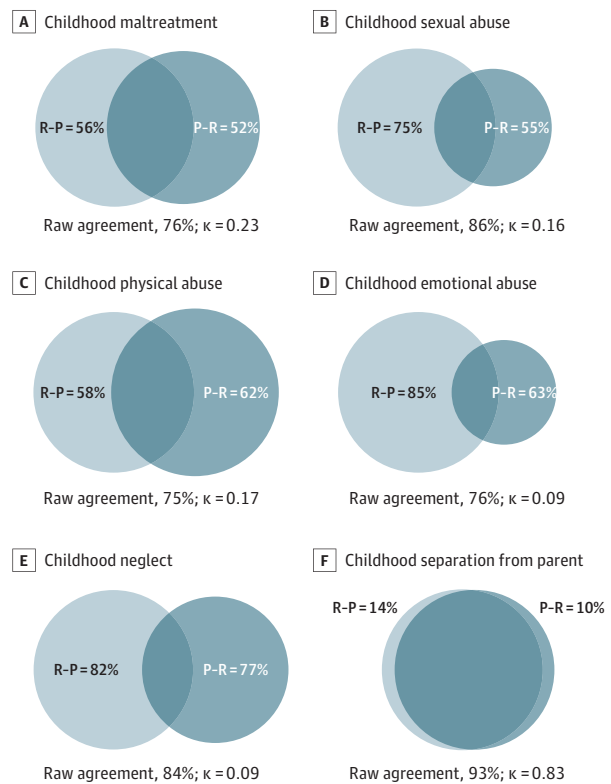
childhood maltreatment among those with prospective observations, (2) the prevalence of prospective observations among those with retrospective reports, and (3) the raw percentage of agreement between prospective and retrospective measures. A random-effects meta-analysis of 7 studies^{14,17,18,20,24,25,32} focusing on a broad measure of child maltreatment revealed that the prevalence of retrospective reports among those with prospective observations was 48% (95% CI, 34%-62%; $I^2 = 96%$); the prevalence of prospective observations among those who retrospectively reported childhood maltreatment was 44% (95% CI, 24%-65%; $I^2 = 99%$); and the percentage of agreement between prospective and retrospective measures of childhood maltreatment was 76% (95% CI, 67%-84%; $I^2 = 99%$). Therefore, on average, 52% of individuals with prospective observations of maltreatment did not retrospectively report it, and 56% of individuals retrospectively reporting maltreatment did not have concordant prospective observations (Figure 2A).

We next undertook sensitivity analyses to test whether the overlap between individuals identified as maltreated through prospective or retrospective measures varied as a function of the type of maltreatment (Figure 2B-E and eTable 3 in the Supplement). First, the prevalence of retrospective reports among those with prospective observations in 8 studies^{13,16,17,22,23,25,27,28} that included childhood sexual abuse was 45% (95% CI, 18%-75%; $I^2 = 97%$); the prevalence of prospective observations among those who retrospectively reported childhood sexual abuse was 25% (95% CI, 12%-41%; $I^2 = 96%$); and the percentage of agreement between prospective and retrospective measures of childhood sexual abuse was 86% (95% CI, 75%-94%; $I^2 = 99%$). Second, the prevalence of retrospective reports among those with prospective observations in the 9 studies^{15-17,23,25-28,31} that included childhood physical abuse was 38% (95% CI, 18%-60%; $I^2 = 98%$); the prevalence of prospective observations among those who retrospectively reported childhood physical abuse was 42% (95% CI, 19%-66%; $I^2 = 98%$); and the percentage of agreement between prospective and retrospective measures of childhood physical abuse was 75% (95% CI, 62%-86%; $I^2 = 99%$). Third, the prevalence of retrospective reports among those with prospective observations in the 4 studies^{16,23,25,28} that included childhood emotional abuse was 37% (95% CI, 23%-52%; $I^2 = 84%$); the prevalence of prospective observations among those who retrospectively reported childhood emotional abuse was 15% (95% CI, 4%-33%; $I^2 = 97%$); and the percentage of agreement between prospective and retrospective measures of childhood emotional abuse was 76% (95% CI, 57%-91%; $I^2 = 99%$). Finally, the prevalence of retrospective reports among those with prospective observations in the 4 studies^{17,19,23,25} that included childhood neglect was 23% (95% CI, 14%-34%; $I^2 = 81%$); the prevalence of prospective observations among those who recalled childhood neglect was 18% (95% CI, 13%-25%; $I^2 = 61%$); and the percentage of agreement between prospective and retrospective measures of childhood neglect was 84% (95% CI, 70%-94%; $I^2 = 99%$).

Agreement Between Prospective and Retrospective Measures of Childhood Maltreatment

Because the raw percentage of agreement can be inflated by chance, we next examined the agreement between prospective

Figure 2. Overlap Between Individuals Identified by Virtue of Prospective or Retrospective Measures of Childhood Maltreatment

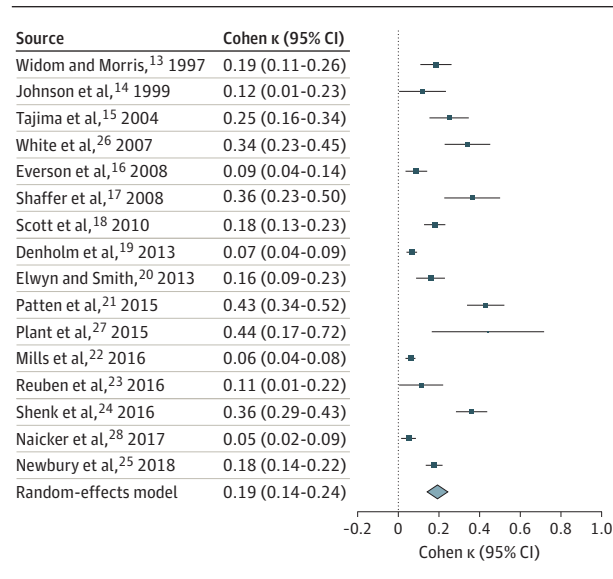


In the Venn diagrams, the light circles indicate retrospective recall, whereas the dark circles indicate prospectively identified childhood maltreatment. The light nonoverlapping section (R-P) shows the proportion of individuals who retrospectively reported a history of childhood maltreatment but were not prospectively identified as experiencing maltreatment in childhood. The dark nonoverlapping section (P-R) shows the proportion of individuals who were prospectively identified as experiencing maltreatment in childhood but did not retrospectively report a history of childhood maltreatment. The overlap between the 2 circles (RQP) shows the proportion of individuals who were prospectively identified as experiencing maltreatment in childhood and retrospectively reported a history of child maltreatment. Seven studies^{14,17,18,20,24,25,32} included childhood maltreatment; 8 studies,^{13,16,17,22,23,25,27,28} childhood sexual abuse; 9 studies,^{15-17,23,25-28,31} childhood physical abuse; 4 studies,^{16,23,25,28} childhood emotional abuse; and 4 studies,^{17,19,23,25} childhood neglect. An individual study by Reuben et al²³ investigated the overlap between groups identified by virtue of prospective or retrospective measures of childhood separation from from parents (due to separation, divorce, death, or removal from home; not included in the meta-analysis).

and retrospective measures based on the κ , which accounts for chance findings and provides an estimate of variation in agreement in the population. A random-effects model meta-analysis of the 16 studies that included any measure of maltreatment revealed that the agreement between prospective and retrospective measures of childhood maltreatment was poor, with $\kappa = 0.19$ (95% CI, 0.14-0.24; $P < .001$; $I^2 = 93%$). A forest plot displaying the meta-analytic findings is reported in Figure 3.

We found some evidence of publication bias, as suggested by slight asymmetry of the funnel plot (eFigure 2A in the Supplement) (Egger test, $z = 4.4273$; $P < .001$) and association between effect sizes and corresponding sampling variances (Begg test, $\tau = 0.37$; $P = .052$). To correct for funnel-plot asymmetry arising from publication bias, we used a

Figure 3. Forest Plot Depicting the Results of a Random-Effects Meta-analysis



Results are reported as Cohen κ agreement between prospective and retrospective measures of childhood maltreatment. When studies reported multiple effect sizes for different maltreatment types, the mean of the κ s was calculated to compute 1 overall effect size per study ($\kappa = 0.19$; 95% CI, 0.14-0.24; $P < .001$; $I^2 = 93\%$). Diamond marker indicates overall effect size and its variation; different sizes of markers are a function of the standard error for κ s in individual studies in the random-effects model.

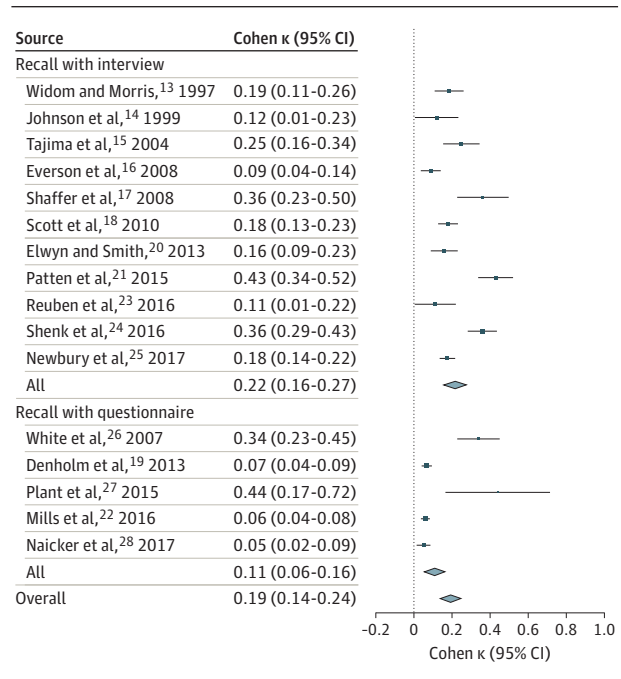
trim-and-fill procedure. The trim-and-fill results with 17 studies ($\kappa = 0.19$; 95% CI, 0.14-0.24; $P < .001$; $I^2 = 92\%$) (eFigure 2B in the Supplement) were similar to the results of our original meta-analysis, suggesting no substantial role of publication bias on the meta-analysis results.

Jackknife sensitivity analyses showed overall little evidence of undue effects of individual studies in the meta-analyses. The κ estimates in 16 automated permutations where each study was omitted in turn were similar and had overlapping CIs (eFigure 3 in the Supplement).

Predictors of Heterogeneity in Agreement Between Prospective and Retrospective Measures of Childhood Maltreatment

Finally, we tested predictors of heterogeneity across studies with subgroup and metaregression analyses. First, we considered whether the measure used for prospective assessment of maltreatment could explain heterogeneity in effect sizes. Agreement with retrospective reports was similar regardless of whether prospective assessment was based on records (eg, child protection records or medical records; $\kappa = 0.16$; 95% CI, 0.09-0.24), reports (eg, questionnaires or interviews by parents or young people; $\kappa = 0.22$; 95% CI, 0.14-0.31), or mixed measures (records and reports; $\kappa = 0.23$; 95% CI, -0.01 to 0.48). An overall test of moderation showed that prospective measure type did not explain the heterogeneity in agreement ($Q = 1.1755$; $df = 2$; $P = .56$). Second, we considered whether the measure used for retrospective assessment of maltreatment could explain heterogeneity in effect sizes. As shown

Figure 4. Forest Plot Depicting the Results of a Random-Effects Meta-analysis Stratified by the Type of Retrospective Measure Used



Results are reported as Cohen κ agreement between prospective and retrospective measures of childhood maltreatment. Retrospective measures included interview vs questionnaire ($Q = 4.1521$; $df = 1$; $P = .04$). Diamond marker indicates overall effect size and its variation; different sizes of markers are a function of the standard error for κ s in individual studies in the random-effects model.

in Figure 4, retrospective recall during interviews (eg, verbal assessment, including reading a questionnaire aloud) showed higher agreement with prospective measures ($\kappa = 0.22$; 95% CI, 0.16-0.27) compared with retrospective recall using questionnaires (eg, written assessment; $\kappa = 0.11$; 95% CI, 0.06-0.16; difference, -0.11; $P = .04$). An overall test of moderation showed that retrospective measure type explained the heterogeneity in agreement ($Q = 4.1521$; $df = 1$; $P = .04$). Third, we tested whether the type of childhood maltreatment could explain heterogeneity in effect sizes. As shown in eFigure 4 in the Supplement, broad measures of childhood adversity ($\kappa = 0.36$; 95% CI, 0.25-0.48) or maltreatment ($\kappa = 0.23$; 95% CI, 0.17-0.30) showed the strongest agreement, whereas measures of emotional abuse ($\kappa = 0.09$; 95% CI, 0.04-0.13) or neglect ($\kappa = 0.09$; 95% CI, 0.05-0.13) showed the weakest agreement. A formal test of moderation across type of childhood maltreatment was not possible because the subgroups were not independent (ie, different types of childhood maltreatment were measured in the same individuals). Fourth, we tested in meta-regression analyses whether characteristics of the samples could explain heterogeneity in effect sizes. As shown in eFigure 5 in the Supplement, sample size was negatively associated with the κ coefficient ($Q = 4.2251$; $df = 1$; $P = .04$), indicating that smaller samples had higher agreement between prospective and retrospective measures. However, we did not find that heterogeneity in agreement was explained by other characteristics of the samples, such as sex

($Q = 1.1653$; $df = 1$; $P = .28$) or age at retrospective report ($Q = 1.0561$; $df = 1$; $P = .30$). Variation in study quality also did not explain heterogeneity in effect sizes ($Q = 0.1632$; $df = 1$; $P = .69$) (eTable 4 in the [Supplement](#)). Finally, in sensitivity analyses where we selected the highest effect size for the 7 studies reporting multiple effect sizes for different abuse types (instead of calculating the mean as above), we found similar results (eResults in the [Supplement](#)).

Discussion

This meta-analysis is the first, to our knowledge, to examine the agreement between prospective and retrospective measures of childhood maltreatment. Across 16 studies that included 25 471 individuals, we found that prospective and retrospective measures of childhood maltreatment showed poor agreement. Notably, more than half of individuals with prospective observations of childhood maltreatment did not report it retrospectively, and likewise more than half of individuals retrospectively reporting childhood maltreatment did not have concordant prospective observations (Figure 2). This finding suggests that prospective and retrospective measures of childhood maltreatment identify largely different groups of individuals and, thus, cannot be used interchangeably.

Low agreement between prospective and retrospective measures of childhood maltreatment could be explained by multiple factors, such as motivation of reporters, measurement features, and memory biases. Motivation can reduce agreement if prospective or retrospective reporters may gain something by intentionally withholding information about childhood maltreatment (ie, nondisclosure, for example owing to embarrassment, feeling uncomfortable with the interviewer, not wanting to discuss upsetting events, or fear of referral to the authorities) or by fabricating information (ie, false disclosure, for example in the context of harassment, revenge, or family disputes).

Measurement features can also reduce agreement in several ways. First, all childhood maltreatment measures have imperfect test-retest reliability,³⁵ and constraints on reliability add error variance, ultimately reducing agreement between prospective and retrospective measures.³⁶ Second, low agreement may be due to systematic differences in the sensitivity of the measures (as reflected by the lower prevalence of childhood maltreatment identified by prospective vs retrospective measures) (eFigure 1 in the [Supplement](#)); for example, prospective measures through official records might capture only the most severe cases of maltreatment, whereas retrospective reports might detect more true cases. Third, low agreement may be owing to other systematic differences between prospective and retrospective measures, such as the reporter^{13-25,31,32} (eg, official records vs later self-reports), the reporting period^{15,19-21,23} (eg, prospective observation until 12 years of age vs retrospective recall of experiences from 0-18 years of age, or official records capturing maltreatment limited to early childhood owing to the focus of child protection services), or the definition of the maltreatment experience be-

tween prospective and retrospective measures¹⁹ (eg, neglect measured prospectively as lack of parental affection and retrospectively as lack of input or stimulation).

Finally, memory biases can reduce agreement by promoting underreporting and overreporting of actual experiences. On the one hand, underreporting may occur because of (1) deficits in encoding the maltreatment experience in early life owing to immature, delayed, or impaired brain development³⁷; (2) deficits in consolidating the maltreatment memory owing to low emotional valence (ie, not experiencing the event as distressing)³⁸ or disrupted sleep patterns³⁹; (3) deficits in reconsolidating the maltreatment memory owing to memory updating during subsequent reactivation if false feedback is given⁴⁰ (eg, being told that the experience was not abusive), if the memory is no longer associated with distressing emotions (eg, after successful psychotherapy),⁴¹ or if reappraisal is positively biased by personality features (eg, high agreeableness)²³; (4) deficits in memory storage owing to brain injury or aging⁴²; or (5) deficits in retrieving the maltreatment memory owing to infantile amnesia,⁴³ forgetting (eg, because of low contextual reinforcement or interference by competing memories),⁴⁴ or cognitive avoidance strategies to regulate affect.⁴⁵⁻⁴⁷ On the other hand, overreporting may occur because of (6) bias in memory encoding or reconsolidation owing to individual suggestibility (as shown in experimental paradigms of imagination inflation, false feedback, or memory implantation) or a source-monitoring error (eg, misinterpretation of internal images or dreams as lived experiences)^{40,48,49} or (7) inaccurate retrieval linked to negative bias in autobiographical memory (eg, in depression).⁵⁰

Our findings support some of these factors. First, we found that the agreement between prospective and retrospective measures of childhood maltreatment was higher in studies that used interviews rather than questionnaires to elicit retrospective recall (Figure 4). This finding is consistent with broader observations regarding the assessment of life stress and may occur because interviews enable provision of a more detailed definition of maltreatment, contextual or anchoring methods, and greater engagement of participants.⁵¹ Second, agreement was also higher in studies with smaller samples (eFigure 5 in the [Supplement](#)), which might reflect the presence of more detailed retrospective assessments. Finally, the agreement for any of the childhood maltreatment measures included was substantially lower than the agreement for more clear-cut forms of adversity, such as parental loss ($\kappa = 0.83$ in the study by Reuben et al²³; Figure 2F), suggesting that subjective interpretation of the childhood maltreatment measures may contribute to the observed heterogeneity. More research is clearly needed to disentangle factors contributing to the low agreement between prospective and retrospective measures of childhood maltreatment.

Limitations

Our findings should be interpreted in the context of some limitations. First, because of the high levels of heterogeneity, the average meta-analytical effect sizes for agreement should be interpreted with caution. However, we used random-effects models to minimize bias linked to high heterogeneity and note that the

meta-analytical CIs are narrow and consistent with the interpretation given.

Second, the results describe the agreement between prospective and retrospective measures of childhood maltreatment commonly used in the context of research studies. Therefore, the results cannot be extrapolated to infer agreement or validity of measures of childhood maltreatment used in other contexts (eg, retrospective allegations brought to the attention of the criminal justice system).

Third, although prospective measures are generally considered to be more valid (specific) indicators of the occurrence of maltreatment,⁵² the low agreement between prospective and retrospective measures cannot be interpreted to directly indicate poor validity of retrospective measures. For example, prospective measures may have lower sensitivity (ie, may identify a lower proportion of individuals who were maltreated), and the higher prevalence of retrospective measures could, thus, indicate greater ability to identify true cases of childhood maltreatment. If that was the case, predictions from retrospective measures should converge on the same outcomes as those of more specific prospective measures (convergent validity), and retrospective measures should not only be associated with outcomes assessed with the same method (ie, self-reports) but should also be associated with outcomes assessed with other methods, such as objective measures (eg, medical examinations or biomarkers [discriminant validity]).⁵³ A few studies^{13,15,23,25} have tested these questions and have observed that prospective and retrospective measures assessed in the same individuals are associated with

similar outcomes. However, retrospective measures showed stronger associations with self-reported outcomes than objectively assessed outcomes,^{23,31} raising concerns about potential common method bias.⁵⁴ Therefore, further research in other samples is needed to comprehensively evaluate the construct validity of retrospective measures. Regardless of any concerns regarding their validity, retrospective reports may still be pragmatically used in the clinic as risk indicators associated with incidence of psychopathology, its course of illness, or treatment response.^{3,55}

Conclusions

Our findings have implications for researchers and health care professionals. Although retrospective reports and prospective measures identify at-risk individuals, the groups of individuals identified with either measure are not the same (Figure 2). Therefore, assuming that the underlying risk mechanisms are the same in both groups may be inaccurate. That is, the mechanisms underlying disease risk in children identified as being maltreated through prospective assessments may be different from the mechanisms underlying disease risk in adults retrospectively reporting childhood maltreatment. If risk mechanisms are different, then the 2 groups will need different interventions to effectively prevent and treat disease. As such, our findings provide a new framework for etiologic research on childhood maltreatment and intervention development.

ARTICLE INFORMATION

Accepted for Publication: December 13, 2018.

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2019 Baldwin JR et al. *JAMA Psychiatry*.

Published Online: March 20, 2019.
doi:10.1001/jamapsychiatry.2019.0097

Author Contributions: Drs Baldwin and Danese had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Baldwin, Reuben, Danese.
Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Baldwin, Danese.
Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Baldwin, Danese.

Obtained funding: Danese

Administrative, technical, or material support: Newbury, Danese.

Supervision: Danese.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was supported by award U24AG047867 from the US National Institute on Aging (NIA) of the National Institutes of Health and award ES/M00919X/1 from the UK Economic and Social Research Council and the Biotechnology and Biological Sciences Research Council. Drs Baldwin and Newbury reported support from the UK Economic and Social Research Council. Mr Reuben reported support from grant AG032282 from the NIA, grant MR/P005918 from the UK Medical Research Council,

and grant F31ES029358 from the US National Institute of Environmental Health Sciences. Dr Danese reported funding from the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHS, the NIHR, or the Department of Health and Social Care.

REFERENCES

1. Masson JM. Freud and the seduction theory. *Atlantic*. <https://www.theatlantic.com/magazine/archive/1984/02/freud-and-the-seduction-theory/376313/>. February 1984. Accessed February 11, 2019.
2. Teicher MH, Samson JA, Anderson CM, Ohashi K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nat Rev Neurosci*. 2016;17(10):652-666. doi:10.1038/nrn.2016.111
3. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am J Psychiatry*. 2012;169(2):141-151. doi:10.1176/appi.ajp.2011.11020335
4. Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household

dysfunction to many of the leading causes of death in adults: the Adverse Childhood Experiences (ACE) Study. *Am J Prev Med*. 1998;14(4):245-258. doi:10.1016/S0749-3797(98)00017-8

5. Brewin CR, Andrews B, Gotlib IH. Psychopathology and early experience: a reappraisal of retrospective reports. *Psychol Bull*. 1993;113(1):82-98. doi:10.1037/0033-2909.113.1.82
6. Hardt J, Rutter M. Validity of adult retrospective reports of adverse childhood experiences: review of the evidence. *J Child Psychol Psychiatry*. 2004;45(2):260-273. doi:10.1111/j.1469-7610.2004.00218.x
7. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3). doi:10.18637/jss.v036.i03
8. Freeman MF, Tukey JW. Transformations related to the angular and the square root. *Ann Math Stat*. 1950;21(4):607-611. doi:10.1214/aoms/117729756
9. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. 2011;12(1):35. doi:10.1186/1471-2105-12-35
10. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46. doi:10.1177/001316446002000104
11. Revelle WR. psych: Procedures for psychological, psychometric, and personality research, R package 1.8.4. <https://cran.r-project.org/web/packages/psych/index.html>. Accessed February 11, 2019.
12. Fleiss JL, Cohen J, Everitt B. Large sample standard errors of kappa and weighted kappa. *Psychol Bull*. 1969;72(5):323-327. doi:10.1037/h0028106

13. Widom CS, Morris S. Accuracy of adult recollections of childhood victimization, II: childhood sexual abuse. *Psychol Assess*. 1997;9(1):34-46. doi:10.1037/1040-3590.9.1.34
14. Johnson JG, Cohen P, Brown J, Smiles EM, Bernstein DP. Childhood maltreatment increases risk for personality disorders during early adulthood. *Arch Gen Psychiatry*. 1999;56(7):600-606. doi:10.1001/archpsyc.56.7.600
15. Tajima EA, Herrenkohl TI, Huang B, Whitney SD. Measuring child maltreatment: a comparison of prospective parent reports and retrospective adolescent reports. *Am J Orthopsychiatry*. 2004;74(4):424-435. doi:10.1037/0002-9432.74.4.424
16. Everson MD, Smith JB, Hussey JM, et al. Concordance between adolescent reports of childhood abuse and Child Protective Service determinations in an at-risk sample of young adolescents. *Child Maltreat*. 2008;13(1):14-26. doi:10.1177/1077559507307837
17. Shaffer A, Huston L, Egeland B. Identification of child maltreatment using prospective and self-report methodologies: a comparison of maltreatment incidence and relation to later psychopathology. *Child Abuse Negl*. 2008;32(7):682-692. doi:10.1016/j.chiabu.2007.09.010
18. Scott KM, Smith DR, Ellis PM. Prospectively ascertained child maltreatment and its association with DSM-IV mental disorders in young adults. *Arch Gen Psychiatry*. 2010;67(7):712-719. doi:10.1001/archgenpsychiatry.2010.71
19. Denholm R, Power C, Li L. Adverse childhood experiences and child-to-adult height trajectories in the 1958 British Birth Cohort. *Int J Epidemiol*. 2013;42(5):1399-1409. doi:10.1093/ije/dyt169
20. Elwyn L, Smith C. Child maltreatment and adult substance abuse: the role of memory. *J Soc Work Pract Addict*. 2013;13(3):269-294. doi:10.1080/1533256X.2013.814483
21. Patten SB, Wilkes TC, Williams JV, et al. Retrospective and prospectively assessed childhood adversity in association with major depression, alcohol consumption and painful conditions. *Epidemiol Psychiatr Sci*. 2015;24(2):158-165. doi:10.1017/S2045796014000018
22. Mills R, Kisely S, Alati R, Strathearn L, Najman J. Self-reported and agency-notified child sexual abuse in a population-based birth cohort. *J Psychiatr Res*. 2016;74:87-93. doi:10.1016/j.jpsychires.2015.12.021
23. Reuben A, Moffitt TE, Caspi A, et al. Lest we forget: comparing retrospective and prospective assessments of adverse childhood experiences in the prediction of adult health. *J Child Psychol Psychiatry*. 2016;57(10):1103-1112. doi:10.1111/jcpp.12621
24. Shenk CE, Noll JG, Peugh JL, Griffin AM, Bensman HE. Contamination in the prospective study of child maltreatment and female adolescent health. *J Pediatr Psychol*. 2016;41(1):37-45. doi:10.1093/jpepsy/jsv017
25. Newbury JB, Arseneault L, Moffitt TE, et al. Measuring childhood maltreatment to predict early-adult psychopathology: comparison of prospective informant-reports and retrospective self-reports. *J Psychiatr Res*. 2018;96:57-64. doi:10.1016/j.jpsychires.2017.09.020
26. White HR, Widom CS, Chen PH. Congruence between adolescents' self-reports and their adult retrospective reports regarding parental discipline practices during their adolescence. *Psychol Rep*. 2007;101(3, pt 2):1079-1094.
27. Plant DT, Pariante CM, Sharp D, Pawlby S. Maternal depression during pregnancy and offspring depression in adulthood: role of child maltreatment. *Br J Psychiatry*. 2015;207(3):213-220. doi:10.1192/bjp.bp.114.156620
28. Naicker SN, Norris SA, Mabase M, Richter LM. An analysis of retrospective and repeat prospective reports of adverse childhood experiences from the South African Birth to Twenty Plus cohort. *PLoS One*. 2017;12(7):e0181522. doi:10.1371/journal.pone.0181522
29. Robins LN. *Deviant Children Grown Up*. Baltimore, MD: Williams & Wilkins; 1966.
30. Williams LM. Recall of childhood trauma: a prospective study of women's memories of child sexual abuse. *J Consult Clin Psychol*. 1994;62(6):1167-1176. doi:10.1037/0022-006X.62.6.1167
31. Widom CS, Shepard RL. Accuracy of adult recollections of childhood victimization, I: childhood physical abuse. *Psychol Assess*. 1996;8(4):412-421. doi:10.1037/1040-3590.8.4.412
32. Raphael KG, Widom CS, Lange G. Childhood victimization and pain in adulthood: a prospective investigation. *Pain*. 2001;92(1-2):283-293. doi:10.1016/S0304-3959(01)00270-6
33. Goodman GS, Ghetti S, Quas JA, et al. A prospective study of memory for child sexual abuse: new findings relevant to the repressed-memory controversy. *Psychol Sci*. 2003;14(2):113-118. doi:10.1111/1467-9280.01428
34. Barnes JE, Noll JG, Putnam FW, Trickett PK. Sexual and physical revictimization among victims of severe childhood sexual abuse. *Child Abuse Negl*. 2009;33(7):412-420. doi:10.1016/j.chiabu.2008.09.013
35. Colman I, Kingsbury M, Garad Y, et al. Consistency in adult reporting of adverse childhood experiences. *Psychol Med*. 2016;46(3):543-549. doi:10.1017/S0033291715002032
36. Anastasi A, Urbina S. *Psychology Testing*. New Jersey: Prentice Hall; 1997.
37. Eichenbaum H. Memory: organization and control. *Annu Rev Psychol*. 2017;68:19-45. doi:10.1146/annurev-psych-010416-044131
38. Roozendaal B, McEwen BS, Chattarji S. Stress, memory and the amygdala. *Nat Rev Neurosci*. 2009;10(6):423-433. doi:10.1038/nrn2651
39. Rasch B, Born J. About sleep's role in memory. *Physiol Rev*. 2013;93(2):681-766. doi:10.1152/physrev.00032.2012
40. Brewin CR, Andrews B. Creating memories for false autobiographical events in childhood: a systematic review. *Appl Cogn Psychol*. 2017;31(1):2-23. doi:10.1002/acp.3220
41. Lane RD, Ryan L, Nadel L, Greenberg L. Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: new insights from brain science. *Behav Brain Sci*. 2015;38:e1. doi:10.1017/S0140525X14000041
42. Frankland PW, Bontempi B. The organization of recent and remote memories. *Nat Rev Neurosci*. 2005;6(2):119-130. doi:10.1038/nrn1607
43. Travaglia A, Bisaz R, Sweet ES, Blitzer RD, Alberini CM. Infantile amnesia reflects a developmental critical period for hippocampal learning. *Nat Neurosci*. 2016;19(9):1225-1233. doi:10.1038/nn.4348
44. Wimber M, Alink A, Charest I, Kriegeskorte N, Anderson MC. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat Neurosci*. 2015;18(4):582-589. doi:10.1038/nn.3973
45. Dalgleish T, Rolfe J, Golden A-M, Dunn BD, Barnard PJ. Reduced autobiographical memory specificity and posttraumatic stress: exploring the contributions of impaired executive control and affect regulation. *J Abnorm Psychol*. 2008;117(1):236-241. doi:10.1037/0021-843X.117.1.236
46. Goodman GS, Quas JA, Ogle CM. Child maltreatment and memory. *Annu Rev Psychol*. 2010;61:325-351. doi:10.1146/annurev.psych.093008.100403
47. Shields GS, Sazma MA, McCullough AM, Yonelinas AP. The effects of acute stress on episodic memory: a meta-analysis and integrative review. *Psychol Bull*. 2017;143(6):636-675. doi:10.1037/bul0000100
48. Loftus EF. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn Mem*. 2005;12(4):361-366. doi:10.1101/lm.94705
49. Schacter DL. The seven sins of memory: insights from psychology and cognitive neuroscience. *Am Psychol*. 1999;54(3):182-203. doi:10.1037/0003-066X.54.3.182
50. Dalgleish T, Werner-Seidler A. Disruptions in autobiographical memory processing in depression and the emergence of memory therapeutics. *Trends Cogn Sci*. 2014;18(11):596-604. doi:10.1016/j.tics.2014.06.010
51. Monroe SM. Modern approaches to conceptualizing and measuring human life stress. *Annu Rev Clin Psychol*. 2008;4:33-52. doi:10.1146/annurev.clinpsy.4.022007.141207
52. Gilbert R, Kemp A, Thoburn J, et al. Recognising and responding to child maltreatment. *Lancet*. 2009;373(9658):167-180. doi:10.1016/S0140-6736(08)61707-9
53. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull*. 1959;56(2):81-105. doi:10.1037/h00046016
54. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol*. 2003;88(5):879-903. doi:10.1037/0021-9010.88.5.879
55. Agnew-Blais J, Danese A. Childhood maltreatment and unfavourable clinical outcomes in bipolar disorder: a systematic review and meta-analysis. *Lancet Psychiatry*. 2016;3(4):342-349. doi:10.1016/S2215-0366(15)00544-1