

Bond University
Research Repository



Disordered eating measures validated in pregnancy samples: a systematic review

Bannatyne, Amy Jean; McNeil, Elyse; Stapleton, Peta; MacKenzie-Shalders, Kristen; Watt, Bruce

Published in:
Eating Disorders

DOI:
[10.1080/10640266.2019.1663478](https://doi.org/10.1080/10640266.2019.1663478)

Licence:
Other

[Link to output in Bond University research repository.](#)

Recommended citation(APA):
Bannatyne, A. J., McNeil, E., Stapleton, P., MacKenzie-Shalders, K., & Watt, B. (2021). Disordered eating measures validated in pregnancy samples: a systematic review. *Eating Disorders*, 29(4), 421-446.
<https://doi.org/10.1080/10640266.2019.1663478>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

Disordered Eating Measures Validated in Pregnancy Samples: A Systematic Review

Amy Jean Bannatyne*^{1,2}, Elyse McNeil^{1,3}, Peta Stapleton², Kristen MacKenzie-Shalders¹, &
Bruce Watt²

¹ Faculty of Health Sciences and Medicine, Bond University

² School of Psychology, Bond University

³ School of Psychology & Counselling, Queensland University of Technology

* Corresponding Author

Email: abannaty@bond.edu.au

Contact telephone number: +61 7 5595 4447

Address: 14 University Drive, Robina, QLD Australia, 4229

Abstract

Although disordered eating in pregnancy has been linked to numerous negative consequences, there is currently no published instrument specifically devised to identify or measure such symptoms in pregnancy. As such, this study systematically reviewed the literature to evaluate the performance of general measures of disordered eating in pregnancy samples. A systematic search of the following electronic databases was undertaken from inception to April 2019: Scopus, Medline, PsycINFO, Embase, ProQuest Dissertations and Theses, and the Cumulative Index to Nursing and Allied Health Literature. From 1724 citations, 8 publications met the inclusion criteria and were included in the review. Most of the included studies (6/8) were of reasonable quality. Overall, three self-report inventories (EDE-Q, EDI-2, and DEBS) and one semi-structured clinical interview (EDE) had some form of psychometric information available. Most studies reported reliability, with only two reporting validity. No studies assessed screening accuracy. Other than the EDE-Q, which had preliminary evidence to suggest possible utility in pregnancy, the findings of this review revealed little to no evidence to support the use of general measures of disordered eating in pregnancy. A strong need for research exploring the validity of existing measures in pregnancy samples, including the EDE-Q, was also evident.

Disordered Eating Measures Validated in Pregnancy Samples: A Systematic Review

Although mental health concerns are one of the most common morbidities during pregnancy and in the postnatal period (Howard et al., 2014), assessing maternal mental health and wellbeing during pregnancy has often been less salient than ensuring optimal physical health for the mother and improving birth outcomes (Bauer, Parsonage, Knapp, Lemmi, & Adelaja, 2014; beyondblue, 2008; Davies, 2015; Hogg, 2013; Naylor et al., 2016). It is, however, well established that poor maternal mental health during the pregnancy and the postnatal period has negative effects and consequences for the mother, child, partner (if present), and immediate family (Bauer et al., 2015; beyondblue, 2008; Gavin et al., 2005; Gray, 2013; Lovestone & Kumar, 1993; Meltzer-Brody & Stuebe, 2014; Oates, 2015).

Pregnancy is a significant biopsychosocial event that often marks the beginning of a new stage in a women's lifespan development (Bulik et al., 2007). The transition to motherhood entails a multitude of rapid changes to a woman's body, eating patterns, social functioning, and self-identity, most of which are largely outside her control (Darvill et al., 2010; Larsson & Andersson-Ellström, 2003). Adjusting to these morphological, endocrinological, and psychological changes may be a relatively uncomplicated process for some women; however, for other women, the adjustment may be more challenging (Easter, 2015). These changes, combined with the age-related vulnerability of a woman's prime childbearing years to eating disturbances (Abebe, Lien, & von Soest, 2012; Hsu, 1989; Leddy, Jones, Morgan, & Schulkin, 2009; Stice, Marti, & Rohde, 2013) means pregnancy may represent a period of increased risk for the onset, resurgence, or exacerbation of disordered eating symptomatology, even for women with no history of such symptoms (Tierney, Fox, Butterfield, Stringer, & Furber, 2011).

Disordered eating in pregnancy has been linked to numerous negative consequences such as miscarriage, prematurity, low birth weight, increased need for caesarean section, and

other obstetric and postpartum difficulties (Linna et al., 2014; Watson et al., 2014). The estimated prevalence of disordered eating during pregnancy varies considerably across studies, ranging from 0.6% to 27.8% (Broussard et al., 2012; Easter et al., 2013; Micali, Treasure, & Simonoff, 2007; Pettersson, Zandian, & Clinton, 2016; Soares et al., 2009; Turton, Hughes, Bolton, & Sedgwick, 1999). These rates differ depending on the characteristics of the sample (i.e., pregnancy stage), the component of disordered eating being investigated (e.g., cognitive vs. affective vs. behavioral), the psychometric instrument employed (e.g., screening tool vs. self-report inventory vs. clinical interview), and the various instrument thresholds used to determine clinically significant scores. Given these varying prevalence estimates and the short- and long- term adverse health consequences for mothers and children, it is crucial to identify valid and reliable instruments that can be used to measure and discern disordered eating in prenatal care and clinical research. In the current review, disordered eating was conceptualized as including subclinical levels of ED symptoms. Alternate forms of disordered eating such external eating, disinhibited eating, or emotional eating, were not considered.

According to Meades and Ayers (2011), two broad approaches can be undertaken to measure disordered eating symptomatology in the perinatal period: (1) use disordered eating measures developed in other populations and validate them for use with pregnant women; or (2) develop pregnancy-specific measures of disordered eating. To date, research and screening for disordered eating in pregnancy has adopted the former approach, with most researchers using instruments developed and validated in non-pregnant populations, and/or suggesting use of these instruments in antenatal care. Examples of tools used and/or suggested by researchers (e.g., Astrachan-Fletcher et al., 2008; Easter et al., 2013; Harris, 2010; Hawkins & Gottlieb, 2013) include formal self-report inventories such as the Eating Disorder Examination Questionnaire (EDE-Q; Fairburn & Beglin, 1994, 2008), the Eating

Disorder Diagnostic Scale (EDDS; Stice, Fisher, & Martinez, 2004), the Eating Disorders Inventory-2 (EDI-2; Garner, 1991), the Eating Attitudes Test (EAT; Garner, Olmsted, Bohr, & Garfinkel, 1982), and the Bulimia Investigatory Test (BITE; Henderson & Freeman, 1987).

Acknowledging the more time-consuming nature of self-report inventories, other researchers have recommended use of brief screening instruments, which typically have 15 items or less and use a simple cut-off score to identify clinical levels of symptomatology, making them ideal for busy clinical settings (Marquer et al., 2012). These instruments do not seek to diagnose a particular condition, rather they aim to identify individuals who may be experiencing concerning symptoms and possibly require further monitoring and/or assessment (Jacobi, Abascal, & Taylor, 2004). For example, many authors and organisations (e.g., Andersen & Ryan, 2009; Harris, 2010; Hawkins & Gottlieb, 2013; Lowes et al., 2012; Micali, 2010; Mitchell & Bulik, 2006; National Eating Disorders Collaboration [NEDC], 2015) have recommended the use of the Sick Control One Fat Food (SCOFF) screening questionnaire. Some researchers (e.g., Andersen & Ryan, 2009; Chizawsky & Newton, 2006; Micali, 2010; NEDC, 2015; Ward, 2008) have also suggested unstructured, opportunistic questions should be used. Extended versions of these informal, opportunistic screening questions, covering both cognitive and behavioural symptomatology, have also been recommended by certain researchers (Chizawsky & Newton, 2006; Ward, 2008). Overall, the SCOFF questionnaire appears to be the most frequent recommendation for detecting disordered eating in pregnancy (NEDC, 2015).

The validity of any instrument, particularly self-report measures, requires re-adjustment for the specific population being examined (Geisinger, 1994). Self-report measures developed for use in a specific population may produce flawed or erroneous results when administered in a different population (Myers & Winters, 2002). Data distribution,

normative values, and cut-offs may deviate from the original population. As such, self-report instruments must be evaluated in new populations to ensure any variability in measurement is minimised or is similar to the original validation population (Myers & Winters, 2002).

Despite the unique nature of pregnancy, psychological scales developed in non-pregnant populations are frequently used without sufficient evidence to suggest these instruments are suitable or effective (Meades & Ayer, 2011). This could lead to inaccurate interpretations of instrument data. For example, the overlap between pregnancy-related symptomatology and disordered eating pathology could potentially increase the percentage of false positives (i.e., over-identifying pregnancy symptoms as ‘disordered’) or false negatives (i.e., under-identifying cases of disordered eating by attributing symptoms to pregnancy) on an instrument. Furthermore, recent research (e.g., Bannatyne, Hughes, Stapleton, Watt, & MacKenzie-Shalders, 2018) has indicated the expression of disordered eating in pregnancy may include unique pregnancy-specific features that are not assessed in traditional instruments such as overvaluation of the offspring’s weight and shape (e.g., desire for the baby to be “small” or “petite”), rationalisation of self-induced vomiting as pregnancy-appropriate, and emotional detachment from the pregnancy.

At the current time, there is no published instrument specifically devised to identify or measure disordered eating in pregnancy. Robust and psychometrically sound measures of disordered eating are needed for screening and research purposes in pregnancy. As such, the aim of this study was to systematically review the literature to identify and evaluate the performance of general measures of disordered eating in pregnancy samples. Performance was explored by examining the reliability (consistency) and validity (accuracy) of each instrument in pregnancy samples. An additional aim of this review was to identify traditional measures of disordered eating that demonstrate adequate performance (i.e., are valid and

reliable) in pregnancy, and those that require further validation in pregnancy. Performance adequacy was evaluated using a standardised tool.

Method

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher, Alessandro, Tetzlaff, & Altman, 2009) statement was used as a methodological framework.

Data Search

A systematic search of the following electronic databases was undertaken from inception to 01 April 2019: Scopus (Elsevier), Medline (Ovid), PsycINFO (Ovid), Embase (Elsevier), ProQuest Dissertations and Theses, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). The following terms were used to conduct all searches: “eating disorder* OR disordered eating OR inappropriate eating OR maladaptive eating OR problematic eating OR eating disturbance” AND “pregnan* OR antenatal OR perinatal OR intrapartum OR maternity*” AND “screen* OR questionnaire OR scale OR instrument OR measure OR assessment OR tool”. The reference lists of the included studies were also crossed checked and relevant citations were manually searched and entered. The primary researcher sought regular expert support from the Faculty Liaison Librarian regarding the search terms, search strategy, and relevant databases.

Eligibility Criteria

Studies were included if they: (1) were published in English or an English translation was available; (2) the sample or subsample was pregnant at the time of data collection; (3) the reliability and/or validity of a disordered eating measure was examined (regardless of whether this was or was not the main aim of the study). In instances where different psychometric properties of a measure were assessed by more than one study using the same sample, both studies were included with a note indicating multiple use of a sample. Where

the same psychometric property of a measure was reported by more than one publication using the same sample, only the first citation was included. Studies were excluded if the methodological design was inappropriate such as review articles, retrospective studies in the postnatal period, or longitudinal designs that evaluated the psychometric performance of a test from prepartum to postpartum, without any clear distinction between time points.

Study Selection

Two reviewers (AB and EM) independently screened the titles and abstracts of all studies identified via the search strategy, followed by the full texts of relevant articles, using the eligibility criteria. Disagreements were resolved by consensus or by consulting with a third author (PS).

Data Extraction

Two reviewers (AB and EM) independently extracted data from included studies. Information extracted included: author; year of publication; country where study occurred; study aim; study methodology (design & setting); sample characteristics; screening instrument used; psychometric properties reported from the sample (i.e., validity or reliability estimates), in addition to data quality and performance adequacy domains.

Data Quality

The quality of included studies was assessed using a combined checklist based on the quality assessment of diagnostic accuracy studies (QUADAS; Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003) and a checklist developed by Mirza and Jenkins (2004). This modified checklist has been used in a previous systematic review examining the psychometric properties of anxiety measures in perinatal populations (Meades & Ayers, 2011). A modified checklist was required due to discipline-specific limitations associated with the QUADAS and standards for reporting diagnostic accuracy (STARD) statements (see Streiner, Sass, Meijer, & Furr, 2016). Based on Meades and Ayers (2011), quality criteria in

the current study were assessed as present (score of 1) or absent (score of 0) on 11 dimensions: 1) explicit study aims, 2) adequate sample size and/or justification, 3) sample described in sufficient detail, 4) population representative of sample receiving test/measure in practice, 5) inclusion and exclusion criteria clearly stated, 6) use of appropriate reference standard, 7) reliability of measure reported, 8) validity of measure investigated and reported, 9) participant withdrawals and dropouts clearly explained, 10) adequate description of data, and 11) discussion of generalisability. The total number of points received for each study was summed, with quality scores for each study ranging from zero to 11. A score of eight or above was considered reasonable quality. A high level of inter-rater reliability was found for the quality coding (ICC = .95), and remaining disagreements were resolved by consensus.

Performance adequacy was evaluated using standardised criteria developed by Terwee et al. (2007), which considers the quality domains of validity, reliability, and responsiveness. The checklist consists of nine domains (see Table 1 for a description). Across each domain, studies receive either a rating of “+” if positively evaluated, a “?” for an intermediate evaluation, or a “-” for negative evaluation. A “0” is assigned when no information in that domain is available or reported in a study. To assist with the current review (i.e., determine whether general measures of disordered eating, which were developed and standardised in a non-pregnant population, are suitable for use in a pregnancy), the content validity domain was adjusted to include explicit mention of a process whereby an instrument’s items were evaluated for appropriateness in a pregnancy context and any modifications were clearly detailed and explained. Similar to Burton et al. (2016), the reproducibility domain was adjusted to include test-retest correlations greater than .70, with means and standard deviations for both time points reported. This change was adopted to be consistent with the methods more frequently reported in ED literature. Unlike other quality appraisal tools, scores on the Terwee et al. (2007) performance adequacy criteria are not

summed into an overall quality score. Terwee et al. (2007) argue that an overall quality score would inaccurately suggest all measurement properties are equally weighted.

Table 1

Performance Adequacy/Quality Criteria from Terwee et al. (2007)

Property	Definition	Criteria of Adequacy ^{a,b}
1. Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured in a particular context.	(+) A clear description is provided of the measurement aim, the process for evaluating the suitability of items in a pregnancy context, and any item modifications made are explicitly detailed (?) A clear description of the above mentioned aspects is lacking OR doubtful design or method (-) Description of the above-mentioned elements is lacking and there appears to be no consideration of whether items are appropriate for pregnancy.
2. Internal consistency	The degree which items are intercorrelated, thus measuring the same construct.	(+) Factor analyses are performed on adequate sample (seven times the number of items) AND Cronbach's α (s) between .70 and .95 for each subscale and/or total scale (?) Cronbach's α (s) presented without factor analysis considered OR doubtful design or method (-) Cronbach's α (s) < .70 or > .95 for each subscale and/or total scale (0) No information found on internal consistency
3. Criterion validity	The degree to which the scores of an instrument are an adequate reflection of a "gold standard"	(+) Convincing argument that gold standard is "gold" AND correlation with gold standard is $\geq .70$ OR AUC is > .70 (?) $\geq .70$ correlation with gold standard OR AUC > .70 is presented without convincing argument that gold standard is "gold" OR doubtful design or method (-) Correlation with gold standard < .70 (0) No information found on criterion validity
4. Construct validity	The degree to which scores on a questionnaire relate to other measures in a manner consistent with theoretically derived hypotheses concerning the concepts being measured	(+) Explicitly tested for and at least 75% of the results are in the expected direction and size (?) Doubtful design or method (e.g., not explicitly tested) (-) Less than 75% of results as expected (0) No information found on construct validity

Notes. MIC = minimal important changes; SDC = smallest detectable change; LOA = limits of agreement; ICC = intraclass correlation; AUC = area under the receiver operating characteristic curve; RR = responsiveness ratio; SD = standard deviation

^a (+) = positive rating; (?) = intermediate or indeterminate rating; (-) = negative rating; (0) = no information available.

^b Doubtful design or method = lacking clear description of the design or methods of study, sample size smaller than 50 participants (should be at least 50 in every [subgroup] analysis), or any important methodological weaknesses in the design or execution of the study.

Table 1 (continued)

Performance Adequacy/Quality Criteria from Terwee et al. (2007)

Property	Definition	Criteria of Adequacy
5. Reproducibility 5.1 Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	(+) $r > .70$ and means and SD for both time points reported (?) $r > .70$; however, means and SD for both time points not reported (-) $r < .70$ OR doubtful design or method (e.g., time interval not mentioned) (0) No information found on agreement
5.2 Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	(+) t -tests, ICC, or weighted $\kappa \geq .70$ (?) Doubtful design or method (e.g., time interval not mentioned, or less valid measures used) (-) t -tests, ICC, or weighted $\kappa < .70$ (0) No information found on reliability
6. Responsiveness	The ability of an instrument to detect clinically important changes over time in the construct to be measured	(+) Treatment program outlined and longitudinal expected changes presented and 75% of results are as expected OR $SDC < MIC$ OR MIC outside of LOA or $RR > 1.96$ or $AUC > .70$ (?) Doubtful design or method (-) SDC , $SDC > MIC$, or MIC equals or inside LOA or $RR < 1.96$ or $AUC < .70$ (0) No information found on responsiveness
7. Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	(+) $< 15\%$ of respondents achieved the highest or lowest possible scores (?) Doubtful design or method (-) $> 15\%$ of respondents achieved the highest or lowest possible scores (0) No information found on floor and ceiling effects
8. Interpretability	Degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores.	(+) Mean and SD scores presented for at least four relevant subgroups of patients and MIC defined (?) Doubtful design or method (e.g., data provided on less than four subgroups or MIC not defined) (0) No information found on interpretation

Notes. MIC = minimal important changes; SDC = smallest detectable change; LOA = limits of agreement; ICC = intraclass correlation; AUC = area under the receiver operating characteristic curve; RR = responsiveness ratio; SD = standard deviation

^a (+) = positive rating; (?) = intermediate or indeterminate rating; (-) = negative rating; (0) = no information available.

^b Doubtful design or method = lacking clear description of the design or methods of study, sample size smaller than 50 participants (should be at least 50 in every [subgroup] analysis), or any important methodological weaknesses in the design or execution of the study.

Results

Results of the Search Strategy and Study Selection

The literature search yielded 1724 potentially relevant citations. A total of 1047 citations remained after removal of all duplicates. These 1047 citations were title and abstract screened, with 151 full text articles assessed for eligibility. After assessment of the full-text articles, 8 citations were included and 142 were excluded. The main reason for exclusion of full-text articles was psychometric properties of utilised measures not being explored or reported. In one case (e.g., Crow et al., 2008), inter-rater reliability was reported; however, the longitudinal nature of the study meant reliability estimates were inclusive of pre-partum, intra-partum, and postpartum, rather than pregnancy alone. Other common reasons were ineligible study designs (e.g., review articles or retrospective studies in the postnatal period) or studies where standardised ED measures had not been employed. The PRISMA flowchart of the article selection process can be seen in Figure 1. The data were managed and stored using Covidence (Veritas Health Innovation, 2018), an electronic systematic review platform.

Characteristics of Included Studies

Eight publications based on seven studies were included in this systematic review, with 1642 participants. The date range was 2005 to 2017. Country breakdown of the studies was as follows: United States ($n = 3$), Hong Kong ($n = 1$), Sweden ($n = 1$), Pakistan ($n = 1$), Portugal ($n = 1$), and Iran ($n = 1$). All studies utilised cross-sectional designs. Sample sizes ranged from 39 to 426. Ages ranged from 15 to 42 years. Most were of reasonable quality with six (75%) of the eight publications obtaining a score of eight or more on the quality assessment.

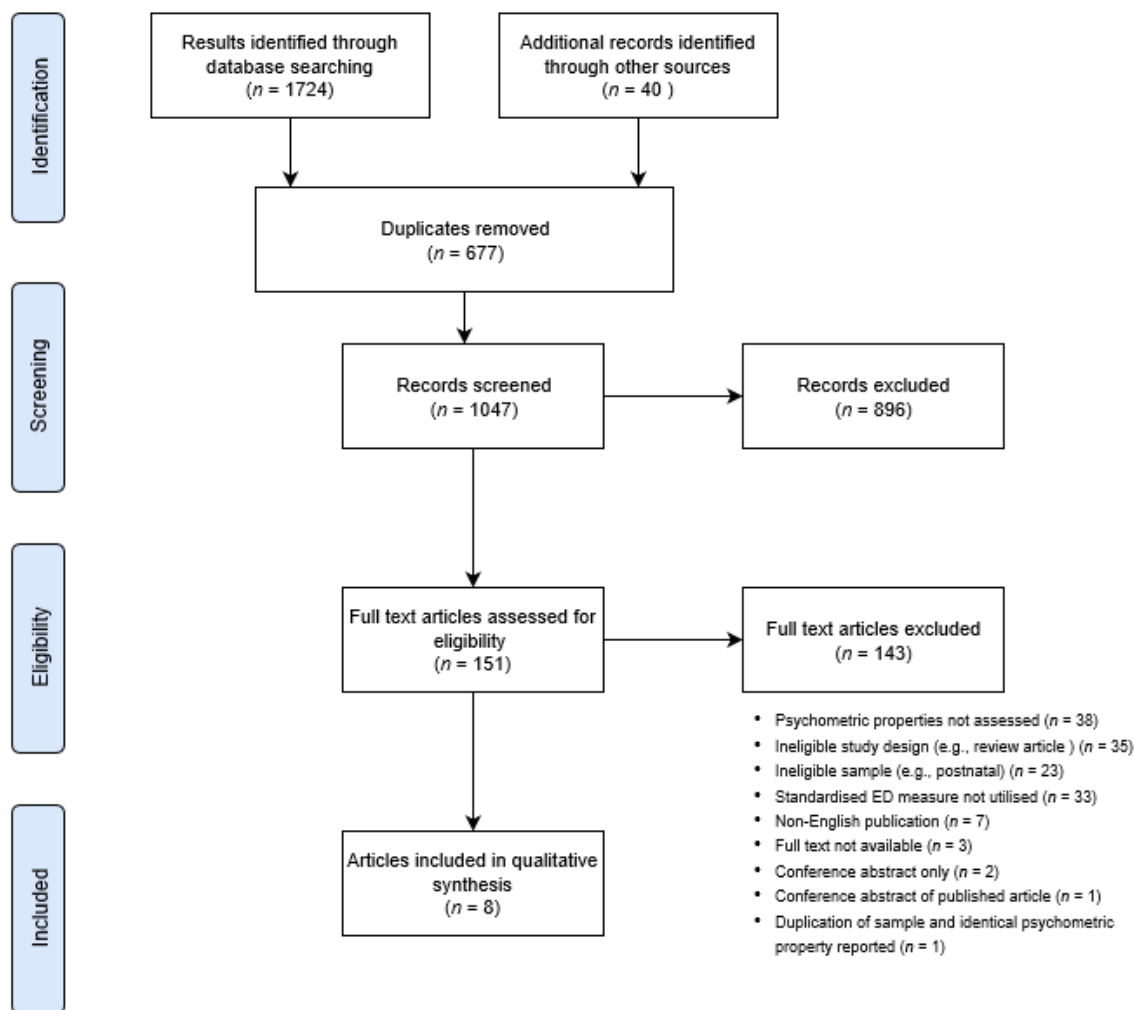


Figure 1. PRISMA article selection flowchart.

Of the eight publications included, only two studies had the aim of assessing the psychometric properties of the employed instruments/measures in a pregnancy sample (Emery et al., 2017; Pettersson et al., 2016). The other six studies reported psychometric information when providing a methodological description of an instrument, as a routine descriptive statistic, and/or indirectly in a correlation coefficient matrix with no explicit explanation. Only two studies assessed validity. One study assessed construct validity (Pettersson et al., 2016), another provided details of discriminant (divergent) validity (Mohamadirizi et al., 2015). No studies assessed criterion-related validity. Seven studies reported reliability, with six reporting internal consistency (Emery et al., 2017; Gonçalves et al., 2015; Lai et al., 2005; Mohamadirizi et al., 2015; Sohail & Muazzam, 2012; Tremblay,

2015). One study reported inter-rater reliability (Kolko et al., 2017). The heterogeneity of the studies and inconsistent psychometric data reported precluded any scope for meta-analysis in this review. Table 2 presents an overview of the included studies.

Psychometric Instruments Identified

Although 16 psychometric instruments that had been used in pregnancy samples were identified during the full-text review process, only four had psychometric information available, including three different self-report instruments: the Eating Disorder Examination Questionnaire (EDE-Q; Fairburn & Beglin, 1994, 2008), the Eating Disorder Inventory-2 (EDI-2; Garner, 1991), and the Disordered Eating Behaviour Scale (DEBS; Muazzam & Khalid, 2011); and one semi-structured clinical interview known as the Eating Disorder Examination (EDE; Cooper & Fairburn, 1987).

Assessment of Psychometric Performance

For each instrument, the psychometric properties reported in the nine publications were assessed using the Terwee et al. (2007) performance appraisal tool. The cumulated results of this evaluation for each instrument are presented in Table 3. Of the four instruments, two did not receive any positive ratings, while two instruments received a positive rating in only one of the nine domains. A description of each measure is detailed following the table, including a summary of the reported psychometric properties in pregnancy samples.

Table 2

Overview of Included Studies

Article citation	(#)	Country	<i>N</i>	Setting	Sample description (<i>M</i> ± <i>SD</i>)	Instrument	Validity reported	Reliability reported	Quality score
Lai et al. (2005)	1	Hong Kong	359	Primary care	Age (29.00 ± 4.78) Gestation (28.66 ± 9.20) Trimester 10% 1 st trimester 20% 2 nd trimester 70% 3 rd trimester BMI – not reported	EDI-2 (self report)	—	Internal consistency	10
Sohail & Muazzam (2012)	2	Pakistan	300	Primary care	Age (25.78 ± 2.55) Trimester 33.3% 1 st trimester 33.3% 2 nd trimester 33.3% 3 rd trimester BMI – not reported	DEBS (self report)	—	Internal consistency	8

Note. ± = standard deviation; EDI-2 = Eating Disorder Inventory-2; DEBS = Disordered Eating Behaviour Scale; EDE-Q = Eating Disorder Examination Questionnaire; EDE = Eating Disorder Examination.

Table 2 (continued)

Overview of Included Studies

Article citation	(#)	Country	<i>N</i>	Setting	Sample description (<i>M</i> ± <i>SD</i>)	Instrument	Validity reported	Reliability reported	Quality score
Mohamadirizi et al. (2015)	3	Iran	213	Primary care	Age (24.12 ± 4.40) Gestation (33.84 ± 3.90) BMI (23.22 ± 4.60)	EDE-Q (self-report)	Discriminant / Divergent	Internal consistency	7
Tremblay (2015)	4	United States	39	Community	Age (26.90 ± 5.12) Gestation (28.10 ± 6.51) Trimester 53.8% 2 nd trimester 46.2% 3 rd trimester BMI – not reported	EDE-Q (self-report)	—	Internal consistency	8
Gonçalves et al. (2015)	5	Portugal	105	Primary care	Age (<i>M</i> = 29.95 years) Trimester 100% 3 rd trimester BMI (<i>M</i> = 28.77)	EDE-Q (self-report)	—	Internal consistency	6

Note. EDI-2 = Eating Disorder Inventory-2; DEBS = Disordered Eating Behaviour Scale; EDE-Q = Eating Disorder Examination Questionnaire; EDE = Eating Disorder Examination.

Table 2 (continued)

Overview of Included Studies

Article citation	(#)	Country	<i>N</i>	Setting	Sample description (<i>M</i> ± <i>SD</i>)	Instrument	Validity reported	Reliability reported	Quality score
Pettersson et al. (2016)	6	Sweden	426	Primary care	Age (32.50 ± 4.60) Trimester 100% 1 st trimester (10-12 weeks)	EDE-Q (self-report)	Construct validity	Internal consistency	8
Emery et al. (2017)	7	United States	129	Community	Age (27.25 ± 5.48) Gestation (14.18 ± 2.15) Trimester 100% 1 st trimester (12-20 weeks) BMI (34.38 ± 6.98)	EDE Pregnancy Version (interview)	—	Internal consistency	8
Kolko et al. (2017) (note – Kolko et al., 2018 was not included as it reported the same psychometric property from the sample same)	8	United States	200	Same sample as Emery et al. (2017) w 71 additional cases	Age (27.67 ± 5.53) Gestation (15.32 ± 2.40) BMI (34.14 ± 7.23)	EDE Pregnancy Version (interview)	—	Inter-rater reliability	8

Note. ± = standard deviation; EDI-2 = Eating Disorder Inventory-2; DEBS = Disordered Eating Behaviour Scale; EDE-Q = Eating Disorder Examination Questionnaire; EDE = Eating Disorder Examination.

Table 3

Assessment of Psychometric Performance Using the Terwee et al. (2007) Criteria

	EDE	EDE-Q	EDI-2	DEBS
Content validity	–	–	–	–
Internal consistency	0	+	–	?
Criterion validity	0	0	0	0
Construct validity	0	0	0	0
Reproducibility (agreement)	0	0	0	0
Reproducibility (reliability)	+	0	0	0
Responsiveness	0	0	0	0
Interpretability	0	0	0	0

Note. EDE = Eating Disorder Examination (clinical interview). EDE-Q = Eating Disorder Examination Questionnaire (self report). EDI-2 = Eating Disorder Inventory-2 (self report). DEBS = Disordered Eating Behaviour Scale (self report).

EDE. The EDE is a semi-structured interview that provides a comprehensive assessment of core ED psychopathology. The instrument was developed and validated for use with non-pregnant adults. The EDE consists of 28 items, of which 22 assess core ED symptomatology. These 22 items assess four main areas/subscales: dietary restraint (5 items), eating concern (5 items), weight concern (5 items), and shape concern (8 items) over the previous 28 days. Each item has a number of prompts for the clinician to elicit greater information. As such, the number of questions asked to obtain sufficient information for item scoring is often much higher. A clinician rates the frequency or intensity of each item on 7-point Likert scales (0 = *feature was absent* to 6 = *feature was present every day or to an extreme degree*). Items within each subscale are summed and averaged to provide subscale scores. Summing and averaging the four subscale scores creates a global score. Higher scores are indicative of greater ED-related symptomatology. An additional six items provide information on key behavioural features of EDs in terms of number of episodes of the behaviour and in some instances number of days on which the behaviour has occurred.

Responses to the EDE items are commonly mapped to the DSM criteria to determine whether a diagnosis of an ED is present or not. Administration of the EDE takes between 45 and 90 minutes (Fairburn, Cooper, & O'Connor, 2008). Clinicians must be trained in administration of the EDE to ensure a comprehensive understanding of the concepts being addressed and rules governing scoring.

The EDE has strong psychometric properties in non-pregnant adult populations and is widely regarded as the “gold standard” instrument in the assessment and diagnosis of EDs (Berg et al., 2012). Two studies included in this review (Emery et al., 2017; Kolko et al., 2017) explored the reliability of a pregnancy-modified EDE in a sample of pregnant women who were overweight or obese. Both studies were derived from the same sample. Three major modifications were made to create the EDE pregnancy version (EDE-PV), including a change in the time periods assessed and removal of two items due to a lack of relevance in the context of pregnancy (e.g., loss of menstruation, desire for flat stomach). Emery et al. (2017) revealed the EDE-PV global scale had a less than adequate internal consistency in a pregnancy sample ($\alpha = .65$). Questionable internal consistency estimates were also revealed for three of the four EDE-PV subscales: dietary restraint ($\alpha = .67$), shape concern ($\alpha = .65$), and weight concern ($\alpha = .59$). Due to excessive skewness on the eating concern subscale, Cronbach's alpha was not calculated. Inter-rater reliability of the EDE-PV was found to be high (ICC = .89) when assessing the intensity and frequency of loss of control over eating (Kolko et al., 2017). The validity of the EDE in pregnancy was not assessed in any available studies.

EDE-Q. The EDE-Q is a self-report derivative of the EDE interview (Fairburn & Beglin, 1994), which provides a brief and comprehensive assessment of core ED psychopathology. The instrument was developed in a non-pregnant population. The EDE-Q consists of 28 items, of which 22 assess core ED symptomatology. These 22 items assess four

main areas/subscales: dietary restraint (5 items), eating concern (5 items), weight concern (5 items), and shape concern (8 items) over the previous 28 days. The frequency or intensity of each item is rated on 7-point Likert scales (0 = *feature was absent* to 6 = *feature was present every day or to an extreme degree*). Items within each subscale are summed and averaged to provide subscale scores. Summing and averaging the four subscale scores creates a global score. Higher scores are indicative of greater ED symptomatology. Various clinical cut-offs have been used in non-pregnant samples, ranging from ≥ 2.30 (Mond et al., 2004) to ≥ 4.00 (Giovazolias et al., 2013; Kelly, Cotter, & Mazzeo, 2012; Penelo et al., 2013). A cut-off of ≥ 2.80 has demonstrated optimal sensitivity and specificity in non-pregnant samples (Mond et al., 2008). The EDE-Q is considered a psychometrically sound instrument in a non-pregnant context, demonstrating good reliability and validity across a range of non-pregnant samples (see Berg et al., 2012, for a review).

Three studies in the current systematic review suggested the EDE-Q global scale had excellent internal consistency in pregnancy samples with Cronbach coefficient alphas ranging from .91 (Gonçalves et al., 2015; Mohamadirizi et al., 2015) to .95 (Tremblay, 2015). Internal consistency estimates for the four subscales scores were also strong based on Gonçalves et al. (2015) and Mohamadirizi et al. (2015): weight concern ($\alpha = .91$), shape concern ($\alpha = .89$), eating concern ($\alpha = .68$ to $.90$), and restraint ($\alpha = .82$ to $.91$). Pettersson et al. (2016) examined the factorial validity of the EDE-Q in a pregnancy sample. Results did not support the theorised four-factor structure of the EDE-Q, favouring a three-factor structure instead.

Results of Pettersson et al. (2016), which examined the EDE-Q using exploratory factor analysis (EFA), revealed a three-factor (as opposed to the original four-factor) model was the most appropriate fit for the pregnancy sample. The three factors were labelled as: (1) dissatisfaction with shape and weight, (2) eating concern and restraint, and (3) importance of

shape and weight. At least three of the EDE-Q items were also found to have low factor loadings (i.e., below .40) on all three factors, indicating these items were not suited to pregnancy. As a result, Pettersson et al. (2016) recommended the use of a 14-item ‘pregnancy optimised’ EDE-Q, derived from the results of the EFA (i.e., the 3-factor model fit). This optimised EDE-Q reportedly provided a more accurate and reliable measurement of disordered eating symptomatology during pregnancy, compared to the traditional EDE-Q; however, in-depth validation analyses (e.g., criterion-related validity) were not undertaken to determine sensitivity and specificity of the optimised version. No other validity evidence for use of the EDE-Q in pregnancy was found in literature.

EDI-2. The EDI-2 is a revised version of the original EDI, a self-report questionnaire designed to measure psychological and behavioural traits pertaining to EDs, particularly anorexia nervosa and bulimia nervosa. The instrument was developed and validated in a non-pregnant population. The EDI-2 consists of 91-items comprised of three main subscales: drive for thinness (7 items), bulimia (7 items), and body dissatisfaction (9 items). The remaining items contribute to eight additional subscales: ineffectiveness, perfectionism, interpersonal distrust, interoceptive awareness, maturity fears, asceticism, impulse regulation, and social insecurity. Items are rated on a 6-point Likert scale (1 = *never* to 6 = *always*). Ratings in the “non-ED range” (e.g., *never*, *rarely*, and *sometimes*) are collapsed and given a score of zero. Ratings in the “ED range” (e.g., *often*, *usually*, and *always*) are given scores of one, two, and three, respectively (Garner, 1991). Higher scores are indicative of a greater tendency to endorse attitudinal and behavioural dimensions pertaining to EDs.

In non-pregnant samples, the EDI-2 is considered a psychometrically sound instrument, with subsequent revisions (e.g., EDI-3) used widely in research and clinical settings as part of a comprehensive diagnostic assessment (see Clausen, Rokkedal, & Rosenvinge, 2009, for a review). Only one study (Lai et al., 2005) in the current systematic

review provided estimates of test score reliability for the EDI-2 in a pregnant sample. Results of Lai et al. (2005) revealed good internal consistency for the body dissatisfaction subscale ($\alpha = .84$), adequate internal consistency for the drive for thinness subscale ($\alpha = .72$), and poor internal consistency for the bulimia subscale ($\alpha = .50$). The validity of the EDI-2 in pregnancy was not explored in any available studies.

DEBS. The DEBS is a self-report instrument developed to assess disordered eating behaviour in non-pregnant Pakistani adolescents and adults. According to the test developers, the instrument captures culture-specific disordered eating practices (Muazzam & Khalid, 2011); however, detail of these unique culture specific items is limited. The DEBS consists of 26-items comprising four subscales: social pressures (6 items), eating choices and habits (5 items), eating withdrawal (8 items), and overeating (7 items). Items are scored on a 5-point Likert scale (0 = *never* to 4 = *always*). Higher scores indicate an individual is more prone to engaging in disordered eating behaviours. Comprehensive exploration of the psychometric properties of the DEBS in non-pregnant populations is limited. Preliminary research by the test developers revealed sound psychometric properties in clinical and community samples of young Pakistani adults (Muazzam & Khalid, 2011). Sohail and Muazzam (2012) recently administered the DEBS to a sample of pregnant women in Pakistan, reporting a Cronbach coefficient alpha of .92 for the full scale. Subscale coefficient alphas were not reported. The validity of the DEBS in pregnancy was not explored in any available studies.

Discussion

This systematic review highlighted the paucity of research validating measures of disordered eating in pregnancy populations. Of the sixteen instruments identified during the full-text review process, only three self-report inventories and one semi-structured clinical interview had some form of psychometric information available. Most studies reported reliability, with only two reporting validity. No studies assessed screening accuracy (i.e.,

sensitivity and specificity). When the Terwee et al. (2007) criteria were applied, no instrument demonstrated adequate properties in each of the nine domains evaluated. Only two measures obtained a positive rating (one domain each), while the other two measures did not obtain any positive ratings. Recommendations regarding the suitability of each evaluated measure are provided, followed by key issues identified during this review.

Evidence for use of the EDE-Q, EDI-2, DEBS, and EDE in Pregnancy

Of the four instruments assessed, the EDE-Q had the most psychometric information available; however, there was still insufficient evidence to confer any decision about the appropriateness of the EDE-Q in pregnancy. The EDE-Q demonstrated good to excellent internal consistency at global and subscale levels, but poor factorial validity based on the hypothesised four-factor structure (Pettersson et al., 2016). Notably, research in non-pregnant samples has also questioned the four-factor structure; suggesting use of the global score is more reliable (Becker et al., 2010; Fairburn et al., 2009). As such, there was preliminary evidence to suggest the global EDE-Q score might be appropriate in pregnancy. No studies assessed criterion-related validity; therefore, the accuracy of the EDE-Q in pregnancy is unknown. Further exploration of the EDE-Q, particularly criterion-related validity, is warranted to determine whether the instrument can be validated for pregnancy and, if so, at which clinical cut-off.

Insufficient evidence precluded a thorough psychometric evaluation of the EDI-2 and DEBS, thereby limiting any recommendations regarding the suitability/appropriateness of these instruments in measuring disordered eating symptomatology in pregnancy. The reported culture specific items of the DEBS are, however, likely to limit generalisability in samples outside Pakistan. Further research exploring the validity of the EDI-2 and the DEBS in pregnancy samples is required to determine utility in pregnancy.

A similar outcome was revealed for the EDE interview. While the EDE interview is considered the preeminent instrument in the field of EDs and the standard by which all other EDs instruments are validated (Berg et al., 2012), there was no empirical evidence to confirm the EDE is suitable for use in pregnancy. From the two studies reviewed (Emery et al., 2017; Kolko et al., 2017), the EDE was found to have poor internal consistency; suggesting the intended construct may be compromised when used with pregnancy samples. The EDE did demonstrate good inter-rater reliability (Kolko et al., 2017) when assessing loss of control over eating; however, these results should be interpreted cautiously as the sample was comprised of pregnant women in the overweight and obese BMI range only and the modified EDE employed had never been administered prior to the study. Overall, there is insufficient evidence at the current time to suggest the EDE interview can serve as the gold standard instrument for identifying disordered eating in pregnancy. Further research investigating the psychometric properties of the EDE in pregnancy samples is urgently required, particularly exploration of validity if the EDE serves as the gold standard to which existing and new instruments are validated. Without an appropriate gold standard, the development and validation of new self-report instruments is significantly hindered (Greenhalgh, 1997; Wacholder, Armstrong, & Hartage, 1993).

Evidence for use of the SCOFF in Pregnancy

Despite empirical literature and various antenatal guidelines encouraging clinicians to screen for disordered eating using the SCOFF (Andersen & Ryan, 2009; Harris, 2010; Hawkins & Gottlieb, 2013; Lowes et al., 2012; Micali, 2010; Mitchell & Bulik, 2006; NEDC, 2015), no published studies were found to support this recommendation. Only one published study (Hubin-Gayte & Squires, 2012) had administered the SCOFF in a pregnant sample, with no psychometric data reported. While discussion of the results of Hubin-Gayte and Squires (2012) is beyond the scope of this review, it is important to note the researchers did

not verify whether disordered eating was present or not using an appropriate reference standard, therefore the sensitivity and specificity (i.e., accuracy) of the SCOFF in pregnancy remains unclear. Other psychometric data (e.g., reliability and other forms of validity) were also unreported by Hubin-Gayte and Squires (2012).

A Need for Pregnancy-Specific Measures of Disordered Eating

A key issue noted in existing literature (Bannatyne et al., 2018; Easter et al., 2013; Pettersson et al., 2016), and confirmed in this review, is the absence of pregnancy-specific measures of disordered eating. This is in contrast to postnatal depression where several instruments specific to the perinatal period have been developed (e.g., the Edinburgh [Postnatal] Depression Scale and the Postpartum Depression Screen) after researchers acknowledged the poor content validity of general depression measures in pregnancy (Meades & Ayers, 2011). Similar to the measurement of postnatal depression, the present review revealed there is insufficient evidence to support the use of general measures of disordered eating with women who are pregnant, including the SCOFF. Furthermore, given the overlap between disordered eating and normative pregnancy symptoms, a pregnancy-specific screening instrument that is sensitive to the eating and weight-related changes that occur during pregnancy is warranted. Several researchers have noted the need for such instrument in recent years (e.g., Easter et al., 2013; Pettersson et al., 2016). Terwee et al. (2007) also consider content validity to be one of the most important measurement properties.

Limitations

The results of this review were limited by the use of stringent performance adequacy criteria defined by Terwee et al. (2007), which has been criticised in literature for being overly conservative in the allocation of positive ratings (Burton et al., 2016; Reneman, Dijkstra, Geertzen, & Dijkstra, 2010). Despite this, the Terwee et al. (2007) criteria continues to be the most widely utilised tool when evaluating the psychometric performance of

instruments (Bird et al., 2012; Cavelti, Kvrjic, Beck, Kossowsky, & Vauth, 2012). The small number of studies available also limited the current review; therefore, the low performance appraisal scores in the current study were mostly due to a lack of available data, rather than poor psychometric performance of the instruments. This issue highlights the dearth of research investigating accurate and reliable screening/measurement of disordered eating symptomatology in pregnancy. It is also acknowledged that including studies where an ED instrument was administered in a language other than English is a limitation given language alone can impact psychometric performance (Beaton et al., 2000). Removal of these studies would have reduced the already small number of available articles that could be included and again highlights the paucity of research in this area.

Lastly, it is important to note the current study systematically identified and reviewed the psychometric properties of general measures of disordered eating in pregnancy samples only; however, research has also indicated the postpartum period carries substantial vulnerability for relapse, exacerbation, or onset of disordered eating (Astrachan-Fletcher et al., 2008; Larsson & Andersson-Ellstrom, 2003; Micali et al., 2007; Pettersson et al., 2016) and there may be significant short- and long- term risks for women and their children (see Astrachan-Fletcher et al., 2008, Micali et al., 2011, Watson et al., 2014). As such, future research should also systematically identify and review the psychometric performance of general measures of disordered eating in the immediate postpartum context (i.e., 6 to 12 months post-birth).

Conclusion

Other than the EDE-Q, which had some preliminary evidence to suggest possible utility, findings of this review revealed little to no evidence to support the use of traditional measures of disordered eating in pregnancy and strong need for research exploring the validity of existing self-report inventories in a pregnancy context, including the EDE-Q.

Comprehensive validation of these instruments requires validation against an accepted reference standard such as a clinical interview; however, there was insufficient evidence to support the utility of the traditional gold standard instrument (the EDE interview) in pregnancy. Furthermore, despite widespread endorsement, there was also no empirical evidence to support the SCOFF questionnaire as an appropriate screening instrument in antenatal settings. Without reliable and valid measures of disordered eating in pregnancy, researchers and clinicians will have difficulty identifying, measuring, and monitoring disordered eating symptoms in pregnancy. As such, development of pregnancy-specific instruments is required.

References

- Abebe, D.S., Lien, L., & von Soest, T. (2012). The development of bulimic symptoms from adolescence to young adulthood in females and males: A population-based longitudinal cohort study. *International Journal of Eating Disorders, 45*, 737–745. doi:10.1002/eat.20950
- Andersen, A. E., & Ryan, G. L. (2009). Eating disorders in the obstetric and gynecologic patient population. *Obstetrics and Gynecology, 114*, 1353-1367. doi:10.1097/AOG.0b013e3181c070f9
- Astrachan-Fletcher, E., Veldhuis, C., Lively, N., Fowler, C., & Marcks, B. (2008). The reciprocal effects of eating disorders and the postpartum period: A review of the literature and recommendations for clinical care. *Journal of Women's Health, 17*, 227-239. doi:10.1089/jwh.2007.0550
- Bannatyne, A. J., Hughes, R., Stapleton, P., Watt, B., & MacKenzie-Shalders, K. (2018). Signs and symptoms of disordered eating in pregnancy: A Delphi consensus study. *BMC Pregnancy and Childbirth, 18*, 262-278. doi:10.1186/s12884-018-1849-3
- Bauer, A., Parsonage, M., Knapp, M., Iemmi, V., & Adelaja, B. (2014). *The costs of perinatal mental health problems*. Centre for Mental Health and London School of Economics.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*, 3186-3191. doi:10.1097/00007632-200012150-00014
- Becker, A. E., Thomas, J. J., Bainivualiku, A., Richards, L., Navara, K., Roberts, A. L., Gilman, S. E., Striegel-Moore, R. H., & Healthy Fiji Study Group. (2010). Validity and reliability of a Fijian translation and adaptation of the Eating Disorder Examination Questionnaire. *International Journal of Eating Disorders, 43*, 171–178. doi:10.1002/eat.20675

- Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2011). Psychometric evaluation of the eating disorder examination and Eating Disorder Examination-Questionnaire: A systematic review of the literature. *International Journal of Eating Disorders, 45*, 428–438. doi:10.1002/eat.20931
- beyondblue. (2008). *Perinatal mental health national action plan 2008-2010: Full report*. Melbourne, VIC: beyondblue.
- Bird, V. J., Le Boutillier, C., Leamy, M., Larden, J., Oades, G., Williams, L. J., & Slade, M. (2012). Assessing the strengths of mental health consumers: A systematic review. *Psychological Assessment, 24*, 1024-1033. doi:10.1037/a0028983
- Broussard, B. (2012). Psychological and behavioural traits associated with eating disorders and pregnancy: A pilot study. *Journal of Midwifery & Women's Health, 57*, 61-66. doi:10.1111/j.1542-2011.2011.00089
- Bulik, C. M., Von Holle, A., Hamer, R., Knoph Berg, C., Torgersen, L., Magnus, P., Stoltenberg, C., . . . & Reichborn-Kjennerud, T. (2007). Patterns of remission, continuation and incidence of broadly defined eating disorders during early pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Psychological Medicine, 37*, 1109–1118. doi:10.1017/S0033291707000724
- Burton, A. L., Abbott, M. J., Modini, M., & Touyz, S. (2016). Psychometric evaluation of self-report measures of binge eating symptoms and related psychopathology: A systematic review of the literature. *International Journal of Eating Disorders, 49*, 123-140. doi:10.1002/eat.22453
- Cavelti, M., Kvrjic, S., Beck, E-M., Kossowsky, J., & Vauth, R. (2012). Assessing recovery from schizophrenia as an individual process: A review of self-report instruments. *European Psychiatry, 27*, 19-32. doi:10.1016/j.eurpsy.2011.01.007

- Chizawsky, L. L., & Newton, M. S. (2006). Eating disorders: Identification and treatment in obstetrical patients. *Nursing for Women's Health, 10*, 482-488. doi:10.1111/j.1552-6356.2006.0097.x
- Clausen, L., Rokkedal, K., & Rosenvinge, J. H. (2009). Validating the Eating Disorder Inventory (EDI-2) in two Danish samples: A comparison between female eating disorder patients and females from the general population. *European Eating Disorders Review, 17*, 462-467. doi:10.1002/erv.945
- Cooper, Z., & Fairburn, C. G. (1987). The Eating Disorder Examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders, 6*, 1-8. doi:10.1002/1098-108x(198701)6:1<1::aid-eat2260060102>3.0.co;2-9
- Darvill, R., Skirton, H., & Farrand, P. (2010). Psychological factors that impact on women's experiences of first-time motherhood: A qualitative study of the transition. *Midwifery, 26*, 357-366. doi:10.1016/j.midw.2008.07.006
- Davies, S. (2015). *Annual report of the Chief Medical Officer 2014: The health of the 51% women*. London: Department of Health. Retrieved from: www.gov.uk/government/publications/chief-medical-officer-annual-report-2014-womens-health
- Easter, A. (2015). Understanding eating disorders in the antenatal and postnatal periods. *Perspective, 26*, 14-15.
- Easter, A., Bye, A., Taborelli, E., Corfield, F., Schmidt, U., Treasure, J., & Micali, N. (2013). Recognising the symptoms: How common are eating disorders in pregnancy? *European Eating Disorders Review, 21*, 340-344. doi:10.1002/erc.2229
- Emery, R. L., Grace, J. L., Kolko, R. P., & Levine, M. S. (2017). Adapting the Eating Disorder Examination for use during pregnancy: Preliminary results from a

- community sample of women with overweight and obesity. *International Journal of Eating Disorders*, *50*, 597-601. doi:10.1002/eat.22646
- Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders*, *16*, 363–370. doi:10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-#
- Fairburn, C. G., & Beglin, S. J. (2008). Eating Disorder Examination Questionnaire (EDE-Q 6.0). In C. G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders* (pp. 309–313). New York: Guilford Press.
- Fairburn, C., Cooper, Z., Doll, H. A., O'Connor, M. E., Bohn, K., Hawker, D. M., . . . & Palmer, R. L. (2009). Transdiagnostic cognitive behavioral therapy for patients with eating disorders: A two-site trial with 60-week follow-up. *American Journal of Psychiatry*, *166*, 311–319. doi:10.1176/appi.ajp.2008.08040608
- Fairburn, C. G., & Cooper, Z. (1993). The Eating Disorder Examination. In C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: Nature, assessment, and treatment (12th edition)*. New York: Guilford Press (pp. 317–360).
- Fairburn, C. G., Cooper, Z., & O'Connor, M. (2008). Eating Disorder Examination (edition 16.0D). In C. G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders*. New York: Guilford Press (pp. 265–308).
- Garner, D. M. (1991). *Eating Disorder Inventory-2: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*, 871-878. doi:10.1017/S0033291700049163

- Gavin, N.I., Gaynes, B.N., Lohr, K.N., Meltzer-Brody, S., Gartlehner, G., & Swinson, T. (2005). Perinatal depression: A systematic review of prevalence and incidence. *Obstetrics and Gynecology, 106*, 1071-1083. doi:10.1097/01.AOG.0000183597.31630.db
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*, 304-312. doi:10.1037/1040F
- Giovazolias, T., Tsaousis, I., & Vallianatou, C. (2013). The factor structure and psychometric properties of the Greek version of the Eating Disorders Examination Questionnaire (EDE-Q). *European Journal of Psychological Assessment, 29*, 189-196. doi:10.1027/1015-5759/a000138
- Gonçalves, S., Freitas, F., Freitas-Rosa, M. A., & Machado, B. C. (2015). Dysfunctional eating behaviour, psychological well-being and adaptation to pregnancy: A study with women in the third trimester of pregnancy. *Journal of Health Psychology, 20*, 535-542. doi:10.1177/1359105315573432
- Gray, R. (2013). *Life stage: Pre-conception and pregnancy. Annual report of the chief medical officer 2012. Our children deserve better: Prevention pays*. London: Department of Health. Retrieved from: http://socialwelfare.bl.uk/subjectareas/services-client-groups/children-youngpeople/departmentofhealth/15526533571_2901304_CMO_All.pdf
- Greenhalgh, T. (1997). How to read a paper: Papers that report diagnostic or screening tests. *British Medical Journal, 315*, 540. doi:10.1136/bmj.315.7107.540
- Harris, A. A. (2010). Prenatal advice for caring for women with eating disorders during the perinatal period. *Journal of Midwifery and Womens Health, 55*, 579-586. doi:10.1016/j.mwh.2010.07.008

- Hawkins, L. K., & Gottlieb, B. (2013). Screening for eating disorders in pregnancy: How uniform screening during a high risk period could minimise under-recognition. *Journal of Women's Health, 22*, 390-392. doi:10.1089/jwh.2013.4313
- Henderson, M., & Freeman, C. P. (1987). A self-rating scale for bulimia: The "BITE." *The British Journal of Psychiatry, 150*, 18-24. doi:10.1192/bjp.150.1.18
- Hogg, S. (2013). *Prevention in mind: all babies count: Spotlight on perinatal mental illness*. London: NSPCC. Retrieved from: <https://www.nspcc.org.uk/services-and-resources/research-and-resources/2013/all-babies-count-spotlight-perinatal-mental-health/>
- Howard, L. M., Molyneaux, E., Dennis, C-L., Rochat, T., Stein, A., & Milgrom, J. (2014). Non-psychotic mental disorders in the perinatal period. *Lancet, 384*, 1775–1788. doi:10.1016/S0140-6736(14)61276-9
- Hsu, L. G. (1989). The gender gap in eating disorders: Why are the eating disorders more common among women? *Clinical Psychological Review, 9*, 393–407. doi:10.1016/0272-7358(89)90063-9
- Hubin-Gayte, M., & Squires, C. (2012). Étude de l'impact de la grossesse sur les comportements alimentaires à travers l'utilisation du questionnaire SCOFF. *L'Evolution Psychiatrique, 77*, 201-212. doi:10.1016/j.evopsy.2012.01.002
- Jacobi, C., Abascal, L., & Taylor, C. (2004). Screening for eating disorders and high-risk behavior: Caution. *International Journal of Eating Disorders, 36*, 280-295. doi:10.1002/eat.20048
- Kelly, N. R., Cotter, E. W., & Mazzeo, S. E. (2012). Eating Disorder Examination Questionnaire (EDE-Q): Norms for black women. *Eating Behaviors, 13*, 429–432. doi:10.1016/j.eatbeh.2012.09.001.

- Kolko, R. P., Emery, R. L., Marcus, M. D., & Levine, M. D. (2017). Loss of control over eating before and during early pregnancy among community women with overweight and obesity. *International Journal of Eating Disorders, 50*, 582-586. doi:10.1002/eat.22630
- Kolko, R. P., Salk, R. H., Sweeny, G. M., Marcus, M. D., & Levine, M. D. (2018). Mothers' loss of control over eating during pregnancy in relation to their infants' appetitive traits. *Appetite, 120*, 1-5. doi: 10.1016/j.appet.2017.08.007
- Lai, B. P.-Y., Tang, C. S.-K., & Tse, W. K.-L. (2005). Prevalence and psychosocial correlates of disordered eating among Chinese pregnant women in Hong Kong. *Eating Disorders, 13*, 171-186. doi:10.1080/10640260590918991
- Larsson, G., & Andersson-Ellström, A. (2003). Experiences of pregnancy-related body shape changes and of breastfeeding in women with a history of eating disorders. *European Eating Disorders Review, 11*, 116-124. doi:10.1002/erv.497
- Leddy, M. A., Jones, C., Morgan, M. A., & Schulkin, J. (2009). Eating disorders and obstetric-gynecologic care. *Journal of Women's Health, 18*, 1395-1401. doi:10.1089/jwh.2008.1183
- Linna, M. S., Raevuori, A., Haukka, J., Suvisaari, J. M., Suokas, J. T., & Gissler, M. (2014). Reproductive health outcomes in eating disorder. *International Journal of Eating Disorders, 46*, 826-833. doi:10.1002/eat.22179
- Lovestone, S., & Kumar, R., (1993). Postnatal psychiatric illness: The impact on partners. *British Journal of Psychiatry, 163*, 210-216. doi:10.1192/bjp.163.2.210
- Lowes, H., Kopeika, J., Micali, N., & Ash, A. (2012). Anorexia nervosa in pregnancy. *The Obstetrician and Gynaecologist, 14*, 179-187. doi:10.1111/j.1744-4667.2012.00110.x

- Marquer, C., Barry, C., Mouchenik, Y., Hustache, S., Djibo, D., & Manzo, M. (2012). A rapid screening tool for psychological distress in children 3–6 years old: Results of a validation study. *BMC Psychiatry, 12*, 170. doi:10.1186/1471-244x-12-170
- Meades, R., & Ayers, S. (2011). Anxiety measures validated in perinatal populations: A systematic review. *Journal of Affective Disorders, 133*, 1-15.
doi:10.1016/j.jad.2010.10.009
- Meltzer-Brody, S., & Stuebe, A. (2014). The long-term psychiatric and medical prognosis of perinatal mental illness. *Best Practice & Research: Clinical Obstetrics & Gynaecology, 28*, 49–60. doi:10.1016/j.bpobgyn.2013.08.009
- Micali, N. (2010). Management of eating disorders during pregnancy. *Progress in Neurology and Psychiatry, 23*, 635–641. doi:10.1002/pnp.158
- Micali, N., Simonoff, E., Stahl, D., & Treasure, J. (2011). Maternal eating disorders and infant feeding difficulties: Maternal and child mediators in a longitudinal general population study. *Journal of Child Psychology & Psychiatry, 52*, 800-807.
doi:10.1111/j.1469-7610.2010.02341.x
- Micali, N., Treasure, J., & Simonoff, E. (2007). Eating disorders symptoms in pregnancy: A longitudinal study of women with recent and past eating disorders and obesity. *Journal of Psychosomatic Research, 63*, 297–303. doi:10.1016/j.jpsychores.2007.05.003
- Mirza, I., & Jenkins, R. (2004). Risk factors, prevalence, and treatment of anxiety and depressive disorders in Pakistan: A systematic review. *British Medical Journal, 328*, 794-797. doi:10.1136/bmj.328.7443.794
- Mitchell, A. M., & Bulik, C. M. (2006). Eating disorders and women's health: An update. *Journal of Midwifery and Women's Health, 51*, 193-201. doi:10.1016/j.jmwh.2006.01.005

- Mohamadirizi, S., Kordi, M., Shakeri, M. T., & Modares-Gharavi, M. (2015). The relationship between eating disorder symptoms and obsessive compulsive disorder in primigravida women. *Iranian Journal of Nursing and Midwifery Research*, *20*, 642-646. doi:10.4103/1735-9066.170015
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*, *62*, 1006-1012. doi:10.1371/journal.pmed.1000097
- Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*, *42*, 551–567. doi:10.1016/S0005-7967(03)00161-X
- Mond, J. M., Myers, T. C., Crosby, R. D., Hay, P. J., Rodgers, B., Morgan, J. F., Lacey, J. H., & Mitchell, J. E. (2008). Screening for eating disorders in primary care: EDE-Q versus SCOFF. *Behaviour Research and Therapy*, *46*, 612–622. doi:10.1016/j.brat.2008.02.003
- Muazzam, A. B., & Khalid, R. (2011). Development and validation of Disordered Eating Behavior Scale: Identification, prevalence and difference with clinically diagnosed eating disorders. *Pakistan Journal of Psychological Research*, *26*, 127-148.
- Myers, K., & Winters, N. C. (2002). Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*, 114-122. doi:10.1097/00004583-200202000-00004
- National Eating Disorders Collaboration. (2015). *Pregnancy and eating disorders: A professional's guide to assessment and referral*. Crows Nest, NSW: NEDC.

- Naylor, C., Das, P., Ross, S., Honeyman, M., Thompson, J., & Gilbert, H. (2016). *Bringing together physical and mental health: A new frontier for integrated care*. London: The King's Fund. Retrieved from: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Bringing-together-Kings-Fund-March-2016_1.pdf
- Oates, M. (2015). *Perinatal mental health services. Recommendations for the provision of services for childbearing women*. London: Royal College of Psychiatrists. Retrieved from: www.rcpsych.ac.uk/usefulresources/publications/collegereports/cr/cr197.aspx
- Penelo, E., Negrete, A., Portell, M., & Raich, R. M. (2013). Psychometric properties of the Eating Disorder Examination Questionnaire (EDE-Q) and norms for rural and urban adolescent males and females in Mexico. *PLoS ONE*, 8, e83245.
doi:10.1371/journal.pone.0083245
- Pettersson, C. B., Zandian, M., & Clinton, D. (2016) Eating disorder symptoms pre- and postpartum. *Archives of Womens Mental Health*, 19, 675-680. doi:10/1007/s00737-016-0619-3
- Reneman, M. F., Dijkstra, A., Geertzen, J. H., & Dijkstra, P. U. (2010). Psychometric properties of Chronic Pain Acceptance Questionnaires: A systematic review. *European Journal of Pain*, 14, 457-465. doi:10.1016/j.ejpain.2009.08.003
- Soares, R. M., Nunes, M. A., Schmidt, M. A., Giacomelle, A., Manzolli, P., Camey, S., Buzz, C., . . . & Duncan, B. B. (2009). Inappropriate eating behaviours during pregnancy: Prevalence and associated factors among pregnant women attending primary care in southern Brazil. *International Journal of Eating Disorders*, 42, 387-393.
doi:10.1002/eat.20643
- Sohail, R., & Muazzam, A. (2012). Correlates of disordered eating behaviour among pregnant women. *Pakistan Journal of Psychological Research*, 27, 153-172.

- Stice, E., Fisher, M., & Martinez, E. (2004). Eating Disorder Diagnostic Scale: Additional evidence of reliability and validity. *Psychological Assessment, 16*, 60-71.
doi:10.1037/1040-3590.16.1.60
- Stice, E., Marti, C. N., & Rohde, P. (2013). Prevalence, incidence, impairment, and course of the proposed DSM-5 eating disorder diagnoses in an 8-year prospective community study of young women. *Journal of Abnormal Psychology, 122*, 445-457.
doi:10.1037/a0030679
- Streiner, D. L., Sass, D. A., Meijer, R. R., & Furr, R. M. (2016). STARDing again: Revised guidelines for reporting information in studies of diagnostic test accuracy. *Journal of Personality Assessment, 98*, 559-562. doi:10.1080/00223891.2016.1202708
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34-42. doi:10.1016/j.jclinepi.2006.03.012
- Tierney, S., Fox, J. R. E., Butterfield, C., Stringer, E. & Furber, C. (2011). Treading the tight-rope between motherhood and an eating disorder: A qualitative study. *International Journal of Nursing Studies, 48*, 1223–1233. doi:10.1016/j.ijnurstu.2010.11.007
- Tremblay, K. A. (2015). *Eating and psychological distress during pregnancy: The use of ecological momentary assessment* (doctoral dissertation). Retrieved from: ProQuest Dissertations and Theses database (UMI No. 3732671).
- Turton, P., Hughes, P., Bolton, H., & Sedgwick, P. (1999). Incidence and demographic correlates of eating disorder symptoms in a pregnant population. *International Journal of Eating Disorders, 26*, 448-452. doi:10.1002/(SICI)1098-108X(199912)26:4<3C448::AID-EAT10%3E3.0.CO;2-3

Veritas Health Innovation. (2019). *Covidence systematic review software*. Available at:

www.covidence.org

Wacholder, S., Armstrong, B., & Hartage, P. (1993). Validation studies using an alloyed gold standard. *American Journal of Epidemiology*, *137*, 1251-1258.

Ward, V. B. (2008). Eating disorders in pregnancy. *British Medical Journal*, *336*, 93-96.

doi:10.1136/bmj.39393.689595.BE

Watson, H. J., Torgersen, L., Zerwas, S., Reichborn-Kjennerud, T., Knoph, C., Stoltenberg, C., Siega-Riz, A. M., . . . & Bulik, C. M. (2014). Eating disorders, pregnancy, and the postpartum period: Findings from the Norwegian Mother and Child Cohort Study (MoBa). *Norsk Epidemiologi*, *24*, 51-62. doi:10.5324/nje.v24i1-2.1758

Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., & Kleijnen, J. (2003). The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, *3*, 25. doi: 10.1186/1471-2288-3-25