International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 20(4): 224–234 (2011)
Published online in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/mpr.356

Quality assessment of observational studies in psychiatry: an example from perinatal psychiatric research

L. E. ROSS,[1,2,3] S. GRIGORIADIS,[2,3,4,5] L. MAMISASHVILI,[4,5] G. KOREN,[6,7,8] M. STEINER,[9,10,11] C.-L. DENNIS,[2,12] A. CHEUNG[4] & P. MOUSMANIS[13]

1 Social and Epidemiological Research Department, Centre for Addiction & Mental Health, Toronto, ON, Canada
2 Women's College Research Institute, Toronto, ON, Canada
3 Department of Psychiatry, University of Toronto, Toronto, ON, Canada
4 Department of Psychiatry, Sunnybrook Health Sciences Centre, Toronto, ON, Canada
5 Department of Psychiatry, University Health Network, Toronto, ON, Canada
6 Pharmacology, Pharmacy and Medical Genetics, University of Toronto, Toronto, ON, Canada
7 Pediatrics and Physiology/Pharmacology, and the Ivey Chair in Molecular Toxicology, University of Western Ontario, London, ON, Canada
8 The Motherisk Program, The Hospital for Sick Children, Toronto, ON, Canada
9 Psychiatry & Behavioral Neurosciences and Obstetrics & Gynecology, McMaster University, Hamilton, ON, Canada
10 Women's Health Concerns Clinic, St Joseph's Healthcare, Hamilton, ON, Canada
11 Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada
12 Lawrence S. Bloomberg Faculty of Nursing, Toronto, ON, Canada
13 Ontario College of Family Physicians, Toronto, ON, Canada

Key words
quality assessment, depression in pregnancy, antidepressant medication

Correspondence
L. E. Ross, Social and Epidemiological Research Department, Centre for Addiction & Mental Health, 455 Spadina Ave. Suite 300, Toronto, ON, M5S 2G8, Canada.
Email: l.ross@utoronto.ca

Received 21 September 2010;
revised 24 January 2011;
accepted 4 April 2011

Abstract

In perinatal psychiatry, randomized controlled trials are often not feasible on ethical grounds. Many studies are observational in nature, while others employ large databases not designed primarily for research purposes. Quality assessment of the resulting research is complicated by a lack of standardized tools specifically for this purpose. The aim of this paper is to describe the Systematic Assessment of Quality in Observational Research (SAQOR), a quality assessment tool our team devised for a series of systematic reviews and meta-analyses of evidence-based literature regarding risks and benefits of antidepressant medication during pregnancy. *Copyright © 2011 John Wiley & Sons, Ltd.*

Introduction

There is increasing pressure to base health care decisions upon research evidence (Straus *et al.*, 2005), and the field of psychiatry is no exception (Paris, 2000). Although the application of evidence-based medicine (EBM) to psychiatry has been criticized (Gupta, 2009), its advocates argue that practicing EBM will result in the most effective treatment interventions, will lead to discovery of the optimal treatment methods for specific illnesses, and will resolve some of the ethical dilemmas encountered in psychiatric treatment (Evidence-based Medicine Working Group, 1992; Paris, 2000). However, whether these ideals can be satisfactorily achieved depends upon the quality of the evidence applied in decision-making.

In perinatal psychiatry, randomized controlled trials (RCTs) of therapeutic interventions are often not ethical, and so treatment decisions must be made without the benefit of this "gold standard" research evidence (Meldrum, 2000). Though there are limitations of observational studies (including cohort and case-control designs), they are an important method of inquiry to address research questions when RCTs cannot. However, interpretation of these studies must be carried out with caution, taking into account a multitude of confounding factors and design limitations that can affect the results. As a result, rigorous quality assessment processes are required to evaluate the quality of psychiatric research studies with observational designs.

There are presently widely accepted guidelines relevant to observational studies; in particular, the STROBE and the MOOSE. The STROBE statement is a 22-item checklist of issues that should be addressed in reporting of observational studies (Von Elm *et al.*, 2007), while the MOOSE is a consensus statement on conducting meta-analyses of observational studies (Stroup *et al.*, 2000). Both the STROBE and the MOOSE guidelines offer important recommendations to ensure adequate and transparent reporting, which facilitates the assessment of study quality. However, despite their commonplace use for this purpose (Mallen *et al.*, 2006), neither the STROBE nor the MOOSE provide any assessment of study quality; rather, these tools simply quantify the extent to which the details required for quality assessment are adequately reported.

There is no single widely accepted tool for quality assessment of observational studies. In fact, in a recent review of systematic reviews of observational studies published during 2003–2004, only 50% of the studies included any quality assessment, and of those that did, 31% included a quality assessment tool developed for the study and 25% used a quality grading system based solely on study design. In total, 10 different quality assessment tools were used in the 39 systematic reviews of observational studies reported in that two year period (Mallen *et al.*, 2006). Of these 10, six were unstandardized instruments developed specifically for the study, one cited the National Health Service Centre for Reviews and Dissemination report (Khan *et al.*, 2001), and one cited the MOOSE, which, as described earlier, is not intended for the purpose of quality assessment. Only two of the 10 instruments were published, standardized quality assessment instruments. In the section that follows, we briefly review these two quality assessment tools: the Downs and Black checklist (1998) and the Newcastle–Ottawa quality assessment scale (NOS) (Wells *et al.*, 2003).

The Downs and Black checklist was developed to assess the quality of both randomized and non-randomized intervention studies (Downs and Black, 1998), although it has been adapted for application to non-intervention studies, including observational designs (Gaynes *et al.*, 2005). This 27-item checklist includes assessment of both the quality of reporting, as well as the internal and external validity of the study. Specifically, the Downs and Black checklist produces an overall Quality Index in addition to the following subscales: reporting, confounding, bias, external validity and power. This tool is reported to provide an efficient and effective assessment of the quality of both randomized and non-randomized intervention studies.

The NOS was designed to evaluate the quality of observational studies (Wells *et al.*, 2003). The assessment consists of two separate scales for case control and cohort studies. The scale for evaluating case control studies assesses selection, adequacy of case definition of the groups, representativeness of the sample, source and definition of controls, comparability and exposure. The scale for evaluation of cohort studies assesses selection, representativeness of the exposed group, selection of unexposed group, exposure status, and the absence of the outcome of interest at the outset of a study, comparability and outcome. Some psychometric data have been reported for the NOS (Wells *et al.*, 2003).

Although the two instruments described here have had some utility in psychiatric observational research, both have limitations. In particular, the Downs and Black checklist (1998) was developed for intervention studies, and so some items (e.g. statistical power) are not applicable in their published format. The NOS does not control for confounding variables, a critical indicator of quality in observational studies. In this context, there is a need for further development of quality assessment tools that will be applicable to observational studies in the field of psychiatry.

In this report, we describe the development of the "Systematic Appraisal of Quality in Observational Research

(SAQOR)", a quality assessment procedure for observational studies in a particular area of psychiatry for which evidence-based decision-making is particularly complex: the use of antidepressant medication during pregnancy.

Observational studies in the area of reproductive psychiatry often have to rely on data from various registries (e.g. Swedish Birth Registry), data obtained from Teratology Information Services around the world (such as Motherisk in Canada), and data retrieved from existing health databases for which research was not the primary purpose when they were created. For this reason, available data are often incomplete (e.g. with respect to demographic and clinical characteristics), not sufficiently detailed (e.g. regarding exposure to medication), and/or based upon maternal self-report (e.g. regarding outcomes in the neonate). As a result of these methodological challenges, we sought to create a tool that would help evaluate the most important criteria presented in the selected studies. We thought this course of action imperative to the process of interpreting the available evidence as accurately as possible within the context of methodological limitations. As the STROBE checklist was designed to improve quality of reporting, we tried to ensure that we covered all STROBE criteria relevant to our purpose.

Background

Major Depressive Disorder is a serious and life-threatening condition. It is the second leading cause of disability in women (Michaud *et al.*, 2001) and can be chronic and recurrent (Judd *et al.*, 1998). Contrary to popular beliefs, the prevalence of depression increases during the second and third trimesters of pregnancy, exceeding rates in the general population (Bennett *et al.*, 2004). Postpartum depression (PPD) occurs in 10% to 15% of women (O'Hara and Swain, 1996) and is more likely when a woman has had an episode of major depression during pregnancy. Depressed women during pregnancy (a) exhibit poor self-care, (b) are less compliant with prenatal care, (c) experience poorer weight gain due to decreased appetite, (d) are more likely to smoke and use alcohol or illicit drugs, and (e) are at risk for self-injurious behaviour and suicide (Zuckerman *et al.*, 1989). Maternal depression is also associated with factors that predict poor neonatal outcome (Orr and Miller, 1995; Steer *et al.*, 1992; Zuckerman *et al.*, 1990). With treatment, the prognosis is good and the outcome may be better in those women who receive treatment early (Nonacs and Cohen, 2000).

Both antidepressant medication and psychotherapy have been shown to be effective treatment options for perinatal depression (Dennis, 2004a, 2004b, 2004c; Dennis and Stewart, 2004; Gjerdingen, 2003; Sharma, 2002). However, despite the benefits of treatment for mother and infant, depression is seriously under-treated during pregnancy and the postpartum period. For example, a recent US chart review study found that only 6.9% of women had a documented diagnosis of depression during pregnancy, compared to rates of 8.7% and 10.4% prior to and following pregnancy, respectively. Similarly, use of antidepressants was less frequent during pregnancy than either prior to or following pregnancy (Dietz *et al.*, 2007). Research suggests that the low rates of antidepressant medication use by pregnant and breastfeeding women may be related to their concerns about the safety of these therapies for the fetus/infant (Chabrol *et al.*, 2004; Ramos *et al.*, 2007).

Prescribing physicians may also have concerns about the safety of antidepressant exposure for the fetus/neonate. Transient short-term adverse neonatal effects (Boucher *et al.*, 2008; Costei *et al.*, 2002; Ferreira *et al.*, 2007; Galbally *et al.*, 2009; Laine *et al.*, 2003; Levinson-Castiel *et al.*, 2006; Nordeng *et al.*, 2001; Oberlander *et al.*, 2002; Oberlander *et al.*, 2007; Oberlander *et al.*, 2008; Rampono *et al.*, 2009; Stiskal *et al.*, 2001) have been associated with late third trimester exposure to selective serotonin reuptake inhibitors (SSRIs). Risk for neonatal cardiovascular and other congenital malformations (Bakker *et al.*, 2010; Berard *et al.*, 2007; Davis *et al.*, 2007; Kallen and Otterblad Olausson, 2007; Merlob *et al.*, 2009; Pedersen *et al.*, 2009) and increased risk for various poor pregnancy and delivery outcomes (Einarson *et al.*, 2010; Kornum *et al.*, 2010; Lewis *et al.*, 2010; Pastuszak *et al.*, 1993; Reis and Kallen, 2010; Suri *et al.*, 2007; Wen *et al.*, 2006) have also been reported. In response to emerging data regarding potential risk for cardiovascular malformations among neonates exposed to antidepressant medications *in utero*, both the Food and Drug Administration (FDA) and Health Canada issued advisories during the summer of 2004 warning about potential risks associated with the use of antidepressants during pregnancy. A study examining media reports of public health advisories about use of antidepressants during pregnancy underscored the impact these reports have on women who may not even be an intended audience for a specific advisory (i.e. in late pregnancy) (Einarson *et al.*, 2005).

Potential risks of antidepressant treatment must be considered in light of other recent evidence indicating that as many as 68% of women who discontinue antidepressant use during pregnancy will relapse (Cohen *et al.*, 2006), leaving them and their infants vulnerable to the potential effects of untreated perinatal depression. Treatment decisions must therefore weigh the effects of untreated maternal depression (both in the immediate and long term) for a

mother and fetus/infant against the potential adverse effects of antidepressant exposure on the fetus or neonate.

In order to address this clinical need, the University of Toronto Perinatal Antidepressant Treatment Project sought to synthesize the existing research evidence regarding areas of knowledge that inform risk-benefit decisions regarding use of antidepressants during pregnancy; specifically: (a) risks and benefits of antidepressant treatment during pregnancy; and (b) impact of untreated antenatal depression on mother/infant. Owing to the large variability in study quality in this field of research, we have operationally decided to exclude studies of very low quality from our analysis. We hypothesized that analysis of the higher quality studies was more likely to accurately represent the evidence by avoiding distortion of the results by the inclusion of poorly designed and executed studies. In order to accomplish this, we required a comprehensive quality assessment procedure that would be tailored to the specific methodological concerns of importance to this type of psychiatric observational research. In this paper, we describe the resulting quality assessment tool SAQOR and offer some discussion regarding its potential contribution to the field of psychiatric epidemiology.

Methods

Literature search

A systematic literature review was conducted to identify evidence relevant to risk-benefit decision-making regarding the use of antidepressant medications during pregnancy, as described earlier. Databases including MEDLINE (1966–2010), EMBASE (1980–2010), CINAHL (1982–2010), PsycInfo (1887–2010), and the Cochrane Library (2010), were searched independently by two professional librarians with expertise in the areas of psychiatry and psychopharmacology. Studies were considered for inclusion if they were written in English and reported original data regarding outcomes associated with fetal exposure to antidepressant medications or outcomes associated with fetal exposure to maternal depression. Selected studies were required to include a comparison group; database, cross-sectional, case-control and cohort designs including a comparison group of mothers without the exposure of interest (e.g. antidepressant medications, depression during pregnancy) were all eligible for inclusion. Due to the volume of potentially eligible studies, abstracts and unpublished data were not eligible for inclusion in this review. The systematic literature review was guided by an Advisory Committee of key stakeholders, comprised of representatives from psychiatry, primary care/family medicine, pharmacology, obstetrics, neonatology, public health, and patient advocacy.

Data extraction

We developed a data extraction form based on the STROBE checklist. The data extraction form contains five sections, including the following information: (1) general study characteristics (e.g. study design, location, and main purpose); (2) participants (e.g. sample size, demographic and clinical characteristics, and inclusion/exclusion criteria); (3) intervention/exposure (e.g. antidepressant name, dosage, and duration); (4) data collection (e.g. baseline outcome measure, primary and secondary outcomes, and length of follow-up); (5) loss to follow-up, if applicable (e.g. initial sample recruited and number of exposed/unexposed groups completed). We also developed a companion document which includes specific instructions for each of the criteria on the data extraction form, to ensure a consistent protocol for the data extraction procedure. The full data extraction form and companion document are available from the authors upon request. Two research assistants independently extracted data from the selected studies, and a third research assistant, not otherwise involved in this project, compared the completed forms to ensure consistency. This comparison identified rater agreement on $>$ 90% of items extracted from a total of 82 articles.

Quality assessment

The same research assistants who performed data extraction independently assessed quality using a tool developed by our team specifically for this project, provided in Appendix A. The SAQOR was adapted from existing quality assessment instruments (Downs and Black, 1998; Wells *et al.*, 2003) in order to assess the specific criteria necessary for evaluation of data presented in studies relevant to risk-benefit decision-making regarding use of antidepressant medication during pregnancy.

The specific adaptations were made in consultation with our Advisory Committee members and experts in epidemiology, and informed by literature both on methodological issues in observational studies in general (Mann, 2003) and observational studies in reproductive psychiatry specifically, to ensure important confounders and other criteria relevant to this field of study were addressed. Later, we describe the tool we developed for assessing quality of observational studies examining outcomes associated with antidepressant exposure during pregnancy. Subsequently, we describe how we revised this original tool for the assessment of studies examining the impact of untreated antenatal depression on the mother/infant. A detailed companion guide providing instructions for use of the SAQOR is available from the authors upon request.

Quality assessment tool criteria for assessing safety of antidepressant medication(s) during pregnancy

The tool went through several revisions and adjustments based on feasibility testing with several studies selected at random. The final quality assessment tool included the following categories: sample, control/comparison group, quality of measurement(s) and outcome(s), follow-up, and distorting influences. Each of the five categories was further broken down into 3–5 criteria each.

Sample. For both cohort and case-control studies, five criteria related to the sample were evaluated: (a) the sample had to be representative of the source population, i.e. recruited using consecutive or random sampling with 60% of eligible women consenting; (b) the source had to be clear, i.e. the study had to include a clear description of where the sample was drawn from; (c) the sampling method had to be described, i.e. the method of participant recruitment/selection had to be explicitly stated; (d) the sample size had to be adequate to identify statistical differences between groups for Primary Outcomes, as determined based on a power calculation provided in the report; and (e) inclusion/exclusion criteria had to be clear and justified, i.e. explicitly described and applied consistently to all groups. Each of the five criteria could be marked as "yes", "no", or "unclear". In order to receive a rating of "adequate" sample, a minimum of three of these five criteria had to be met. The category was marked "inadequate" if only two or one criteria out of five were met; and "unclear" only if three out of five criteria were marked unclear.

Control/comparison group. The following five criteria were assessed in relation to control/comparison groups: (a) control group had to be included; (b) control/comparison group had to be easily identifiable, i.e. there had to be a clear distinction between the groups in the study, and the same variables considered in the control group had to have been considered in the exposed group(s); (c) the source of the controls had to be clear, i.e. control group had to be drawn from the same population as the exposed group(s); (d) controls had to be matched or randomized (for matched studies, matching criteria had to be given); and (e) statistical differences between cases and controls had to be controlled for, i.e. groups selected for comparison had to be as similar as possible in all characteristics except for their exposure status. The rating procedure was the same as for the Sample category described earlier.

Quality of exposure/outcome measures. Two criteria were assessed with respect to the quality of exposure/outcome measures. (a) Adequate assessment of exposure, i.e. the paper clearly stated how the authors ascertained that the cases/exposed group had indeed been exposed to the variable of interest (antidepressant medication or maternal depression). Exposure was required to be assessed either by a gold standard method (e.g. blood levels, for exposure to antidepressant medication), or confirmed through two independent sources (e.g. maternal report confirmed through chart review). Maternal report alone was considered inadequate in all cases. (b) Adequate measure of outcome(s), i.e. the paper clearly stated what measures were used to assess outcomes proposed to be associated with the exposure, and these assessment procedures were considered methodologically sound. For observer-rated assessment methods, the outcome assessor was required to be blind to the group exposure status. For data acquired through medical chart reviews, independent review of an appropriate specialist was required to confirm the outcome (i.e. infant examined by a pediatrician). Both assessment of exposure and assessment of outcome were required to be marked "adequate" in order to receive an overall rating of "adequate" in this category.

Follow-up (applicable for longitudinal studies only). The following two criteria were assessed with respect to follow-up: (a) the number of participants lost to follow-up was stated; (b) explanations as to why participants could not or would not complete the study were provided. Longitudinal studies were required to meet at least one of these two criteria to be marked "adequate" in this category; other study designs were marked "not applicable".

Distorting influences. We considered this category to be especially important as confounding is one of the most problematic challenges in observational research. Specifically, when trying to parse out effects of depression on fetus from effects of antidepressant medication exposure in utero, it is desirable to have a depressed control/comparison group (i.e. group comprised of women who refused to take antidepressants or who were exposed to known nonteratogens). As such, we assessed the following three criteria with respect to distorting influences: (a) depression was controlled for, specifically, the authors included a disease-matched comparison group or adjusted for maternal depression in multivariate analyses; (b) other psychotropic drugs were controlled for, specifically, the authors excluded women using other psychotropic drugs, matched for this variable in their selection of the comparison group, or adjusted for exposure to other psychotropic medication in multivariate analyses. This criterion was important to consider as many

Int. J. Methods Psychiatr. Res. 20(4): 224–234 (2011). DOI: 10.1002/mpr

psychotropic medications (e.g. benzodiazepines) can potentially have teratogenic effects. (c) Other confounders, especially smoking and exposure to alcohol, were controlled in the analysis; specifically, the report was required to indicate which potential confounders were considered, and how they were assessed or controlled for in the analysis. Simply stating that groups were similar was not interpreted to mean that there were no statistical differences between groups. When the authors stated that the confounders (such as diagnoses other than depression, alcohol use and/or smoking) were taken into account but the data were not given, we marked "yes" only if it was explicitly stated that there were no significant differences between the groups after adjusting for the confounders; we marked "no" if the confounders were not mentioned; and we marked "unclear" if the study mentioned confounders but did not explain how they were dealt with. Two of these three criteria were required to be met in order for a study to be marked "adequate" for distorting influences.

Reporting of data. We assessed the following two criteria with respect to reporting of data: (a) missing data – the authors explained how missing data were addressed and/or dealt with in the analysis. Specifically, the authors indicated the number of participants with missing data for each variable of interest and an appropriate analysis plan (e.g. imputation, list-wise deletion) was provided. b) Data were clearly and accurately presented including confidence intervals where appropriate. Specifically, we noted whether sample sizes were included in figures or tables, and if so, whether the numbers in tables and figures added up as expected.

Impact of untreated antenatal depression on mother/infant

We made only minor adjustments to the original SAQOR, as described earlier, to apply it to this area of research. Specifically, the two criteria assessing the quality of exposure/outcome measures were adapted in the following ways: the measure of exposure was considered to be adequate if the authors employed a standardized, clinician rated measure of depressive symptomatology or diagnosis, e.g. Structural Clinical Interview for DSM (SCID) and the Hamilton Depression Rating Scale (Ham-D). The measure of outcome(s) was considered to be adequate if an independent blind assessment was carried out, or a parental report was corroborated by a health professional.

We also revised the "Distorting Influences" section to reflect the research area. Specifically, we replaced "Depression controlled for in the design or analysis" with "Depression treatment controlled for in the design or analysis"; and within the category of "Other potential confounders" we considered whether the authors controlled for "Smoking, alcohol, and illicit drugs".

A research team member not involved in the tool development assessed inter-rater reliability of the two quality assessment tools. Over 80% agreement was achieved for both quality assessment forms.

Final grade. Results of the quality assessment were compared between raters and differences were discussed with the Principal Investigators until consensus was achieved. A final quality level assignment was then identified based on a modified version of the GRADE quality assessment criteria. Our modification was adapted from the system for grading the quality of evidence and the strength of recommendations developed by the GRADE Working Group, a partnership of scientists concerned with consistency of judgments to aid in making better informed choices in health care (Grade Working Group; Guyatt *et al.*, 2008). The World Health Organization (WHO) has been involved with the GRADE quality assessment development since the project's inception and uses the GRADE method for its own endeavors (Grade Working Group, 2008).

A final quality level (High, Moderate, Low, Very low) was assigned to each study on the basis of study design (e.g. with population-based studies being considered at higher quality levels than cohort studies) and the number of categories marked "adequate" (e.g. a cohort study with three of five categories marked "inadequate" would receive a quality rating of "Very low").

Discussion

In this paper, we have provided a template that can be used in the assessment of study quality for observational studies in the field of perinatal psychiatry. Furthermore, we have shown how this template can be adapted to address specific potential confounds of importance to the area of study. To our knowledge, the SAQOR is the first quality assessment tool that has been adapted to address issues of particular importance in the field of reproductive psychiatry and that can have applications to quality assessment of observational studies in other fields. As research studies in this area are often picked up by the media and can substantially influence provider and patient attitudes towards antidepressant treatment (Einarson *et al.*, 2005), adequate assessment of study quality is essential.

Of particular concern for the quality assessment of the studies included in our systematic review was adequate

control for important confounding variables. In general, few studies reviewed by our team included adequate controls for key variables that could potentially confound relationships between exposure to antidepressants or maternal depression, and various outcomes in the fetus/neonate/child. Lack of adequate control for potential confounders is a key limitation of many observational studies in the field of psychiatry that must be considered in the interpretation of this research (Egger *et al.*, 1998). As such, we believe that these criteria are essential for inclusion in any quality assessment tool. However, this complicates the use of standardized quality assessment tools, in that the key confounders will vary somewhat depending on the specific research question at hand (for example, when our research question examined the impact of antidepressant exposure, maternal depression was a key potential confounder; when our research question examined the impact of maternal depression, exposure to antidepressant treatment was a key potential confounder). With the use of any quality assessment tool, then, the onus will be on the user to identify the relevant potential confounders, and to operationalize adequate control for these variables.

While this work offers an important contribution to psychiatric research methods, important limitations of the SAQOR should be noted. Future research by independent research teams will be required to confirm the reliability and to establish validity of our quality assessment tool. Further, although some items from our tool were derived from other published instruments, some items have not been included in previously published checklists (e.g. Downs and Black checklist (1998), NOS (Wells *et al.*, 2003)). Although each of these items represent issues that are commonly noted as important considerations in study design and execution (von Elm *et al.*, 2007), additional research is needed to determine if these items contribute significantly to the determination of study quality, above and beyond those items more typically included in assessments of study quality. Finally, as in other quality assessment tools (e.g. Downs and Black, 1998), we have not investigated the potential impact of differential weighting of the component domains of our tool. In our use of this instrument, each of the five component domains (sample, control, outcome/exposure, follow-up, distorting influences, reporting of data) is given equal weighting in the overall quality level assigned. Variations in weighting may be appropriate depending upon the research question; future research to address this issue is warranted.

In conclusion, there is a need for standardized quality assessment procedures to describe and ultimately improve methodological rigor in observational studies in psychiatry. Our study provides one template that may be useful for other investigators in this field. For example, a recent critical review highlighted the limitations in available studies assessing the prevalence and incidence of postnatal depression. In particular, few studies in this area have utilized appropriate assessments of study quality (Mann *et al.*, 2010). Our tool could prove useful in research of this nature, as well as in psychiatric research outside of the perinatal period (e.g. in assessing the quality of studies examining the putative association between a given risk factor and psychopathology).

We encourage other scientists and clinicians to further refine and develop the SAQOR for their purposes, and more broadly, encourage greater attention to the assessment of quality in psychiatric observational studies.

Acknowledgments

This research was funded by a Research Syntheses grant from the Canadian Institutes of Health Research (CIHR), KRS-83127, and the Ontario Ministry of Health and Long-Term Care through the Drug Innovation Fund, 2008–005. Dr L. Ross is supported by a New Investigator Award from CIHR and the Ontario Women's Health Council, Award NOW-84656. Dr S. Grigoriadis also holds a New Investigator Award in Women's Health Research from CIHR in partnership with the Ontario Women's Health Council, Award 182838. In addition, support to the Centre for Addiction & Mental Health for salary of scientists and infrastructure has been provided by the Ontario Ministry of Health and Long Term Care. The views expressed here do not necessarily reflect those of the Ministry of Health and Long Term Care.

Declaration of interest statement

The authors have no competing interests.

References

Bakker M.K., Kerstjens-Frederikse W.S., Buys C.H. C.M., De Walle H.E.K., De Jong-van Den Berg L.T.W. (2010) First-trimester use of paroxetine and congenital heart defects: A population-based case-control study. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, **88**(2), 94–100. DOI: 10.1002/bdra.20641

Bennett H.A., Einarson A., Taddio A., Koren G., Einarson T.R. (2004) Prevalence of depression during pregnancy: Systematic review. *Obstetrics and Gynecology*, **103**(4), 698–709. DOI: 10.1097/01.AOG.0000116689.75396.5f

Berard A., Ramos E., Rey E., Blais L., St-Andre M., Oraichi D. (2007) First trimester exposure to paroxetine and risk of cardiac malformations

Int. J. Methods Psychiatr. Res. 20(4): 224–234 (2011). DOI: 10.1002/mpr

in infants: The importance of dosage. *Birth Defects Research. Part B, Developmental and Reproductive Toxicology*, **80**(1), 18–27. DOI: 10.1002/bdrb.20099

Boucher N., Bairam A., BeaulacBaillargeon L. (2008) A new look at the neonate's clinical presentation after in utero exposure to antidepressants in late pregnancy. *Journal of Clinical Psychopharmacology*, **28**(3), 334–339. DOI: 10.1097/JCP.0b013e318173aa2e

Chabrol H., Teissedre F., Armitage J., Danel M., Walburg V. (2004) Acceptability of psychotherapy and antidepressants for postnatal depression among newly delivered mothers. *Journal of Reproductive and Infant Psychology*, **22**(1), 5–12. DOI: 10.1080/02646830310001643094

Cohen L.S., Altshuler L.L., Harlow B.L., Nonacs R., Newport D.J., Viguera A.C., *et al.* (2006) Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment [corrected] [published erratum appears in *JAMA* **296**(2): 170]. *JAMA: The Journal of the American Medical Association*, **295**(5), 499–507. DOI: 10.1001/jama.295.5.499

Costei A.M., Kozer E., Ho T., Ito S., Koren G. (2002) Perinatal outcome following third trimester exposure to paroxetine. *Archives of Pediatrics & Adolescent Medicine*, **156**(11), 1129–1132.

Davis R.L., Rubanowice D., McPhillips H., Raebel M.A., Andrade S.E., Smith D., Yood M.U., Platt R., H.M.O. Research Network Center for Education, Research in Therapeutics. (2007) Risks of congenital malformations and perinatal events among infants exposed to antidepressant medications during pregnancy [see comment]. *Pharmacoepidemiology and Drug Safety*, **16**(10), 1086–1094. DOI: 10.1002/pds.1462

Dennis C.L. (2004a) Preventing postpartum depression part I: A review of biological interventions. *Canadian Journal of Psychiatry [Revue Canadienne De Psychiatrie]*, **49**(7), 467–475.

Dennis C.L. (2004b) Preventing postpartum depression part II: A critical review of non-biological interventions. *Canadian Journal of Psychiatry [Revue Canadienne De Psychiatrie]*, **49**(8), 526–538.

Dennis C.L. (2004c) Treatment of postpartum depression, part 2: A critical review of non-biological interventions. *The Journal of Clinical Psychiatry*, **65**(9), 1252–1265.

Dennis C.L., Stewart D.E. (2004) Treatment of postpartum depression, part 1: A critical review of biological interventions. *The Journal of Clinical Psychiatry*, **65**(9), 1242–1251.

Dietz P.M., Williams S.B., Callaghan W.M., Bachman D.J., Whitlock E.P., Hornbrook M.C. (2007) Clinically identified maternal depression before, during, and after pregnancies ending in live births. *The American Journal of Psychiatry*, **164**(10), 1515–1520. DOI: 10.1176/appi.ajp.2007.06111893

Downs S.H., Black N. (1998) The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, **52**(6), 377–384. DOI: 10.1136/jech.52.6.377

Egger M., Schneider M., Smith G.D. (1998) Meta-analysis: Suprious precision? Meta-analysis of observational studies. *British Medical Journal*, **316**(7125), 140.

Einarson A., Choi J., Einarson T.R., Koren G. (2010) Adverse effects of antidepressant use in pregnancy: An evaluation of fetal growth and preterm birth. *Depression and Anxiety*, **27**(1), 35–38. DOI: 10.1002/da.20598

Einarson A., Schachtschneider A.K., Halil R., Bollano E., Koren G. (2005) SSRI'S and other antidepressant use during pregnancy and potential neonatal adverse effects: Impact of a public health advisory and subsequent reports in the news media. *BMC Pregnancy and Childbirth*, **5**, 11. DOI: 10.1186/1471-2393-5-11

Evidence-based Medicine Working Group. (1992) Evidence-based medicine: A new approach to teaching the practice of medicine. *JAMA: The Journal of the American Medical Association*, **268**(17), 2420–2425. DOI: 10.1001/jama.1992.03490170092032

Ferreira E., Carceller A.M., Agogue C., Martin B.Z., St-Andre M., Francoeur D., et al. (2007) Effects of selective serotonin reuptake inhibitors and venlafaxine during pregnancy in term and preterm neonates. *Pediatrics*, **119**(1), 52–59. DOI: 10.1542/peds.2006-2133

Galbally M., Lewis A.J., Lum J., Buist A. (2009) Serotonin discontinuation syndrome following in utero exposure to antidepressant medication: Prospective controlled study. *The Australian and New Zealand Journal of Psychiatry*, **43**(9), 846–854.

Gaynes B.N., Gavin N., Meltzer-Brody S., Lohr K.N., Swinson T., Gartlehner G., et al. (2005) *Perinatal Depression: Prevalence, Screening Accuracy, and Screening Outcomes.* Evidence Report/Technology Assessment No. 119 (prepared by the RTI-University of North Carolina Evidence-based Practice Center, under contract no. 290-02-0016), No. AHRQ Publication No. 05-E006-2. Rockville, MD, Agency for Healthcare Research and Quality.

Gjerdingen D. (2003) The effectiveness of various postpartum depression treatments and the impact of antidepressant drugs on nursing infants. *The Journal of the American Board of Family Practice/American Board of Family Practice*, **16**(5), 372–382. DOI: 10.3122/jabfm.16.5.372

Gupta M. (2009) Ethics and evidence in psychiatric practice. *Perspectives in Biology and Medicine*, **52**(2), 276–288. DOI: 10.1353/pbm.0.0081

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., *et al.* (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, **336**, 924–926.

Judd L.L., Akiskal H.S., Maser J.D., Zeller P.J., Endicott J., Coryell W., *et al.* (1998) A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. *Archives of General Psychiatry*, **55**(8), 694–700.

Kallen B.A., Otterblad O.P. (2007) Maternal use of selective serotonin re-uptake inhibitors in early pregnancy and infant congenital malformations. *Birth Defects Research*, **79**(4), 301–308. DOI: 10.1002/bdra.20327

Khan K.S., ter Riet G., Glanville J., Sowden A.J., Kleijnen J. (2001) *Undertaking Systematic Reviews of Research on Effectiveness CRD's Guidance for those Carrying Out or Commissioning Reviews*, CRD Report Number 4 (2nd ed.). York, NHS Centre for Reviews and Dissemination, University of York.

Kornum J.B., Nielsen R.B., Pedersen L., Mortensen P.B., Norgaard M. (2010) Use of selective serotonin-reuptake inhibitors during early pregnancy and risk of congenital malformations: Updated analysis. *Clinical Epidemiology*, **2**(1), 29–36. DOI: 10.2147/CLEP.S9256

Laine K., Heikkinen T., Ekblad U., Kero P. (2003) Effects of exposure to selective serotonin reuptake inhibitors during pregnancy on serotonergic symptoms in newborns and cord blood monoamine and prolactin concentrations. *Archives of General Psychiatry*, **60**(7), 720–726.

Levinson-Castiel R., Merlob P., Linder N., Sirota L., Klinger G. (2006) Neonatal abstinence syndrome after in utero exposure to selective serotonin reuptake inhibitors in term infants. *Archives of Pediatrics & Adolescent Medicine*, **160**(2), 173–176.

Lewis A.J., Galbally M., Opie G., Buist A. (2010) Neonatal growth outcomes at birth and

one month postpartum following in utero exposure to antidepressant medication. *The Australian and New Zealand Journal of Psychiatry*, **44**(5), 482–487. DOI: 10.3109/00048670903559593

Mallen C., Peat G., Croft P. (2006) Quality assessment of observational studies is not commonplace in systematic reviews. *Journal of Clinical Epidemiology*, **59**(8), 765–769. DOI: 10.1016/j.jclinepi.2005.12.010

Mann C.J. (2003) Observational research methods. research design II: Cohort, cross sectional, and case-control studies. *EMJ: Emergency Medicine Journal*, **20**(1), 54–60. DOI: 10.1136/emj.20.1.54

Mann R., Gilbody S., Adamson J. (2010) Prevalence and incidence of postnatal depression: What can systematic reviews tell us? *Archives of Women's Mental Health*, **13**(4):295–305.

Meldrum M.L. (2000) A brief history of the randomized controlled trial. From oranges and lemons to the gold standard. *Hematology/Oncology Clinics of North America*, **14**(4), 745–760, vii.

Merlob P., Birk E., Sirota L., Linder N., Berant M., Stahl B., et al. (2009) Are selective serotonin reuptake inhibitors cardiac teratogens? Echocardiographic screening of newborns with persistent heart murmur. *Birth Defects Research*, **85**(10), 837–841. DOI: 10.1002/bdra.20615

Michaud C.M., Murray C.J., Bloom B.R. (2001) Burden of disease – implications for future research. *JAMA: The Journal of the American Medical Association*, **285**(5), 535–539. DOI: 10.1001/jama.285.5.535

Nonacs R., Cohen L.S. (2000) Postpartum psychiatric syndromes. In Sadock B., Sadock V. (eds) *Comprehensive Textbook of Psychiatry* (7th ed.), pp. 1276–1283. Philadelphia, PA, Lippincott Williams and Wilkins.

Nordeng H., Lindemann R., Perminov K.V., Reikvam A. (2001) Neonatal withdrawal syndrome after in utero exposure to selective serotonin reuptake inhibitors. *Acta Paediatrica*, **90**(3), 288–291. DOI: 10.1111/j.1651-2227.2001.tb00306.x

O'Hara M.W., Swain A.M. (1996) Rates and risk of postpartum depression – a meta–analysis. *International Review of Psychiatry*, **8**(1), 37–54.

Oberlander T.F., Eckstein G.R., Fitzgerald C., Ellwood A.L., Misri S., Rurak D., et al. (2002) Prolonged prenatal psychotropic medication exposure alters neonatal acute pain response. *Pediatric Research*, **51**(4), 443–453.

Oberlander T.F., Reebye P., Misri S., Papsdorf M., Kim J., Grunau R.E. (2007) Externalizing and attentional behaviors in children of depressed mothers treated with a selective serotonin reuptake inhibitor antidepressant during pregnancy. *Archives of Pediatrics & Adolescent Medicine*, **161**(1), 22–29.

Oberlander T.F., Warburton W., Misri S., Aghajanian J., Hertzman C. (2008) Effects of timing and duration of gestational exposure to serotonin reuptake inhibitor antidepressants: Population-based study. *The British Journal of Psychiatry*, **192**(5), 338–343. DOI: 10.1192/bjp.bp.107.037101

Orr S.T., Miller C.A. (1995) Maternal depressive symptoms and the risk of poor pregnancy outcome. Review of the literature and preliminary findings. *Epidemiologic Reviews*, **17**(1), 165–171.

Paris J. (2000) Canadian psychiatry across five decades: From clinical inference to evidence-based practice. *Canadian Journal of Psychiatry*, **45**, 34–39.

Pastuszak A., SchickBoschetto B., Zuber C., Feldkamp M., Pinelli M., Sihn S., *et al.* (1993) Pregnancy outcome following first-trimester exposure to fluoxetine (prozac). *JAMA: The Journal of the American Medical Association*, **269**(17), 2246–2248. DOI: 10.1001/jama.1993.03500170076037

Pedersen L.H., Henriksen T.B., Vestergaard M., Olsen J., Bech B.H. (2009) Selective serotonin reuptake inhibitors in pregnancy and congenital malformations: Population based cohort study. *British Medical Journal*, **339**(7723), 735.

Ramos E., Oraichi D., Rey E., Blais L., Berard A. (2007) Prevalence and predictors of antidepressant use in a cohort of pregnant women. *BJOG: An International Journal of Obstetrics and Gynaecology*, **114**(9), 1055–1064. DOI: 10.1111/j.1471-0528.2007.01387.x

Rampono J., Simmer K., Ilett K.F., Hackett L.P., Doherty D.A., Elliot R., *et al.* (2009) Placental transfer of SSRI and SNRI antidepressants and effects on the neonate. *Pharmacopsychiatry*, **42**(3), 95–100.

Reis M., Kallen B. (2010) Delivery outcome after maternal use of antidepressant drugs in pregnancy: An update using Swedish data. *Psychological Medicine*, **40**(10), 1723–1733. DOI: 10.1017/S0033291709992194

Sharma V. (2002) Pharmacotherapy of postpartum depression. *Expert Opinion on Pharmacotherapy*, **3**(10), 1421–1431. DOI: 10.1517/14656566.3.10.1421

Steer R.A., Scholl T.O., Hediger M.L., Fischer R.L. (1992) Self-reported depression and negative pregnancy outcomes. *Journal of Clinical Epidemiology*, **45**(10), 1093–1099.

Stiskal J.A., Kulin N., Koren G., Ho T., Ito S. (2001) Neonatal paroxetine withdrawal syndrome. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, **84**(2), F134–F135.

Straus S.E., Ball C., Balcombe N., Sheldon J., McAlister F.A. (2005) Teaching evidence-based medicine skills can change practice in a community hospital. *Journal of General Internal Medicine*, **20**(4), 340–343. DOI: 10.1111/j.1525-1497.2005.04045.x

Stroup D.F., Berlin J.A., Morton S.C., Olkin I., Williamson G.D., Rennie D., *et al.* (2000) Meta-analysis of observational studies in epidemiology: A proposal for reporting. meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA: The Journal of the American Medical Association*, **283**(15), 2008–2012. DOI: 10.1001/jama.283.15.2008

Suri R., Altshuler L., Hellemann G., Burt V.K., Aquino A., Mintz J. (2007) Effects of antenatal depression and antidepressant treatment on gestational age at birth and risk of preterm birth. *The American Journal of Psychiatry*, **164**(8), 1206–1213. DOI: 10.1176/appi.ajp.2007.06071172

Von Elm E., Altman D.G., Egger M., Pocock S.J., Gotzsche P.C., Vandenbroucke J.P., et al. (2007) The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet*, **370**(9596), 1453–1457. DOI: 10.1016/S0140-6736(07)61602-X

Wells G.A., Shea B., O'Connell D., Petersen J., Welch V., Losos M., et al. (2003) *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-analyses*. Ottawa, Department of Epidemiology and Community Medicine, University of Ottawa.

Wen S.W., Yang Q., Garner P., Fraser W., Olatunbosun O., Nimrod C., *et al.* (2006) Selective serotonin reuptake inhibitors and adverse pregnancy outcomes. *American Journal of Obstetrics and Gynecology*, **194**(4), 961–966. DOI: 10.1016/j.ajog.2006.02.019

Zuckerman B., Amaro H., Bauchner H., Cabral H. (1989) Depressive symptoms during pregnancy: Relationship to poor health behaviors. *American Journal of Obstetrics and Gynecology*, **160**(5 Pt 1), 1107–1111.

Zuckerman B., Bauchner H., Parker S., Cabral H. (1990) Maternal depressive symptoms during pregnancy, and newborn irritability. *Journal of Developmental and Behavioral Pediatrics*, **11**(4), 190–194.

Appendix A. Systematic Appraisal of Quality for Observational Research (SAQOR)

Sample

The sample is representative of the population from which it was drawn.
O Yes O No O Unclear O N/A

The source of the sample is clearly stated.
O Yes O No O Unclear O N/A

The sampling method is described (e.g., consecutive, clinical, community, convenience).
O Yes O No O Unclear O N/A

The sample size is appropriate to determine statistical significance for primary outcomes.
O Yes O No O Unclear O N/A

Entry criteria and exclusions are stated and justified.
O Yes O No O Unclear O N/A

Summary. Sample is:
O Adequate O Unclear O Inadequate
Comments: _____

Control/comparison group

Control group is included.
O Yes O No O Unclear O N/A

The control group is easily identifiable.
O Yes O No O Unclear O N/A

The source of the controls is explained and is appropriate.
O Yes O No O Unclear O N/A

Controls are matched or randomized.
O Yes O No O Unclear O N/A

Statistical differences between cases and controls have been controlled for.
O Yes O No O Unclear O N/A

Summary. Control is:
O Adequate O Unclear O Inadequate
Comments: _____

Quality of exposure/outcome measurements

Adequate assessment of exposure.
O Yes O No O Unclear O N/A

Adequate measure of outcome(s).
O Yes O No O Unclear O N/A

Summary. Quality of exposure/outcome measurements is:
O Adequate O Unclear O Inadequate
Comments: _____

Follow-up

The number of participants lost to follow-up is stated.
O Yes O No O Unclear O N/A

Explanations for loss to follow-up are given.
O Yes O No O Unclear O N/A

Summary. Follow-up is:
O Adequate O Unclear O Inadequate O N/A
Comments: _____

Distorting influences

Key confounder 1 (MDD) is controlled for/taken into account in the design or analysis.
O Yes O No O Unclear O N/A

Key confounder 2 (other psychotropic medications) are controlled for/taken into account in the design or analysis.
O Yes O No O Unclear O N/A

Other potential confounders are controlled for/taken into account in the design or analysis.
O Yes O No O Unclear O N/A

Summary. Description of influences is:
O Adequate O Unclear O Inadequate
Comments: _____

Reporting of data

Explanation for missing data is given.
O Yes O No O Unclear O N/A

Data are clearly and accurately presented including CI where appropriate.
O Yes O No O Unclear O N/A

Summary. Reporting of data is:
O Adequate O Unclear O Inadequate

Comments: _____
_____
_____
_____

Conclusion

Quality of study: O High
O Moderate
O Low
O Very Low

Comments: _____
_____
_____
_____
_____
_____

_____
_____
_____
_____

Int. J. Methods Psychiatr. Res. 20(4): 224–234 (2011). DOI: 10.1002/mpr