

The Tokophobia Severity Scale (TSS): Measurement model, power and sample size considerations

Colin R. Martin¹

Catriona Jones¹

Chao Huang¹

Julie Jomeen²

Laura Boubert³

and

Claire A. Marshall⁴

¹*Institute for Clinical and Applied Health Research (ICAHR), University of Hull, Hull, UK*

²*Health & Human Sciences, Southern Cross University, Gold Coast, Queensland, Australia*

³*Department of Psychology, University of Westminster, London, UK*

⁴*Perinatal Mental Health Team, Humber Teaching NHS Foundation Trust, Hull, UK*

Correspondence: Professor Colin R. Martin

Professor of Perinatal Mental Health | Faculty of Health Sciences

Institute for Clinical and Applied Health Research (ICAHR)

Rm 329, Allam Medical Building

University of Hull

Hull, HU6 7RX, UK

Email: C.R.Martin@hull.ac.uk

The Tokophobia Severity Scale (TSS): Measurement model, power and sample size considerations

Abstract

Objective: To examine the measurement model of the Tokophobia Severity Scale and consider issues of statistical power and sample size from the original instrument development study.

Background: Fear of childbirth (FoC) and tokophobia represents an area of increasing concern within perinatal mental health research and clinical practice. Existing measures of the FoC have been criticised due to either measurement inconsistencies, difficulties in scoring or practical clinical application. Attempting to address these limitations, Wootton and colleagues (2020) recently developed the Tokophobia Severity Scale (TSS), a 13-item self-report measure. A fundamental assumption underpinning the use of the TSS is unidimensionality, however this assertion may have been based on sub-optimal sample size and approach taken to factor structure determination.

Method: Parallel analysis (PA), principal components analysis (PCA), exploratory factor analysis (EFA), power analysis and sample size calculation using a reconstruction of the original dataset from published summary data.

Results: Following replication of the original PCA, a three-factor model was found to offer a significantly better fit to data than a unidimensional model. Power analysis for all factor models evaluated suggested the original study was underpowered.

Conclusion: The TSS remains a promising tool but assumptions regarding its measurement model are based on an inadequate sample size. An alternative

three-factor model is suggested and sample sizes for a sufficiently powered study indicated.

Introduction

An area garnering much clinical and research interest within perinatal mental health is fear of childbirth (FoC)/tokophobia (Dencker et al., 2019; Nilsson et al., 2018). This is perhaps unsurprising given the potential impact of FoC on both mother and baby (Pazzagli et al., 2015). The clinical importance and relevance of the area is underscored by the spectrum of activity related to effective screening (Richens, Smith, & Lavender, 2018), intervention innovation (Klabbers, Wijma, Paarlberg, Emons, & Vingerhoets, 2019), epidemiology (O'Connell, Leahy-Warren, Khashan, Kenny, & O'Neill, 2017) and risk factor identification (O'Connell, Leahy-Warren, Kenny, O'Neill, & Khashan, 2019).

The advent of specialist clinical pathways for FoC (Jones, Marshall, Martin, & Jomeen, 2020) has focused the spotlight on service access and in particular identification and screening. Given that identification for FoC may be accomplished or at least facilitated by validated screening questionnaires (Jones et al., 2020), an important focus of recent research activity has been to identify the most appropriate and best-performing measures (Richens et al., 2018). Several measures have been suggested as appropriate for use in assessing FoC, one of the most widely used being the Wijma Delivery Expectancy/Experience Questionnaire (WDEQ; Wijma, Wijma, & Zar, 1998), however, this particular measure is lengthy even within a shortened format, making practical routine clinical use of the measure potentially problematic. A recent study by Martin et al. (2020) examined the clinician-assessed content overlap between the WDEQ and three other measures suggested as appropriate for the assessment of FoC, the Fear of Birth Scale (FOBS; Haines, Pallant, Karlstrom, & Hildingsson, 2011), the Slade-Pais Expectations of Childbirth Scale (SPECS; Slade, Pais, Fairlie, Simpson, & Sheen, 2016) and the Oxford Worries about Labour Scale (OWLS; Redshaw, Martin, Rowe, & Hockley, 2009). The findings from this study (Martin et al., 2020) were particularly illuminating in that there was relatively little overlap in content

between measures suggesting each tool may be assessing conceptually distinct aspect of FoC and highlighting at the very least that the instruments were by no means interchangeable assessment-wise.

Attempting to address the problem that established measures are often lengthy, sometimes difficult to score within a clinical context and occasionally conceptually incongruent, Wootton and colleagues (2020) recently developed a short (13-item) unidimensional measure circumscribed at the outset as a tool for screening and case identification. Following the selection of an initial pool of items for scale development, Wootton et al. (2020) followed a stringent protocol (Clark & Watson, 1995) to select the best performing items for final inclusion within the tokophobia severity scale (TSS). The TSS demonstrated exceptional internal validity, good convergent validity and measurement unidimensionality, commensurate with the use of the measure as an easy to use and interpret scale with a single unitary score. However, there were some more challenging findings, for example divergent validity could not be adequately established (using the Patient Health Questionnaire-9; Kroenke, Spitzer, & Williams, 2001) and moreover there was a potential tension in the selection of factor analysis method. Wootton et al. (2020) selected principal components analysis (PCA) as the item-reduction method of choice, an approach which maximises the amount of variance (Conway & Huffcutt, 2003) in contrast to factor analysis (FA) which seeks to identify interpretable constructs within a set of items, the relationship between them (constructs) in terms of covariances and consequently establish dimensionally the construct validity of a measure (Byrne, 2005). Interestingly, in terms of the item generation phase of the study, the researchers focused on behavioural and cognitive items related to FoC thus, there would seem a prima face case to justify the evaluation of a two-factor model, though it is conceded that the PCA approach used seems primarily applied to item reduction with a second PCA undertaken to establish unidimensionality. Indeed, with a driver of the study

being to establish a simple to use measure, it is understandable that an underpinning unidimensional model may be preferred, though such an assumption might be more rigorously evaluated given that the dimensionality of a measure circumscribes both application and construct validity (Kline, 1994). A potential ambiguity in the study then was the use of two PCA's one for item-reduction and a second to confirm unidimensionality. It is therefore arguable that the second analysis might have been undertaken using FA. Since PCA and FA often give similar findings this may seem a trivial point, however the factor identification approach undertaken, parallel analysis, identifies both components and factors which then leads to the selection of number of components/factors to extract either by PCA or FA and these (N components/N factors) can be different. Therefore the interpretation of the underlying measurement model of the scale can potentially be radically different when circumscribed by parallel analysis either by components (for PCA) or factors (for FA). Finally, the sample size of the study was very modest (N=122) for the type of analysis (PCA) undertaken and again the veracity of the findings and conclusions drawn may have been impacted by the study being potentially underpowered (Costello & Osborne, 2005; MacCallum, Browne, & Sugawara, 1996).

A particularly impressive aspect of Wootton et al.'s (2020) study however was the level of detail with which the data was described, including a correlation matrix of the original 24 items, the means and standard deviations of the final 13-item scale and the item-component loadings of the 13-item scale. This detail of summary data offers the possibility of recreating a dataset faithful to the original, re-running the original analysis PCA and running a FA without requiring access to the original dataset. **The aim of the current investigation was four-fold:**

1. **Reconstruct the original dataset from the published data and check the accuracy of the data matrix by rerunning the PCA of Wootton et al. (2020).**

2. Rerun the parallel analysis to determine number of components and factors.
3. Run an FA using maximum-likelihood estimation based on number of factors identified in the parallel analysis
4. Undertake a power analysis to determine the sample size for an adequately powered replication study based on aim 3.

Methods

The correlation matrix of the initial 24-item pool of the TSS (Wootton et al., 2020) was duplicated from that published in Table 2. It was noted that one essential correlation was missing (item 9 and item 13) and that the final 13 items of the TSS could not be readily identified by reference to Table 2. and the summary item statistics of the 13-item TSS shown in Table 3 of the paper. The corresponding author was contacted and kindly gave the missing correlation value (0.33) and details to identify the items from the final scale from the initial pool. The corresponding TSS individual item standard deviations were then incorporated into a bespoke program file. The correlation matrix was then converted matrix with the use of the item standard deviations to a full covariance matrix. All requirements for statistical analysis were satisfied with the correlation and covariance matrices.

Statistical analysis

To evaluate the accuracy of the correlation matrix the PCA of Wootton et al. (2020) was replicated (PCA with direct oblimin rotation specifying one factor) on the 13-item TSS. Factor loadings were compared against those reported by Wootton and colleagues (2020) in Table 3. Following confirmation of replicability of item-factor loadings a parallel analysis (Horn, 1965) was undertaken to determine the number of underlying components and factors. A maximum-likelihood exploratory factor analysis (EFA) was then run, with oblimin rotation specifying the number of *factors* identified by the parallel analysis. Two SEM-based approaches to power analysis and sample size calculation were taken based on the model fit index, the Root Mean Squared Error of Approximation (RMSEA), that of MacCallum et al. (1996) whereby a null specified RMSEA value (0.5) is evaluated against an alternate model RMSEA value (0.08) and thus power and sample size are based on a criteria to adequately reject the alternate model in favour of the null model which represents a close-fit of data to

the model. A second related approach advocated by MacCallum et al. (1996) is that of determination of non-close fit whereby the **hypothesis is evaluated that the model** is not a close fit to the data and therefore rejection of this specific hypothesis would indicate support for an alternative hypothesis of a close fit to data, the suggestion being that evaluating a null hypothesis of non-close fit prevents an issue in which support for the null hypothesis of good fit is unsupportable. This also presents the opportunity to reject specific hypotheses of close fit and non-close fit in relation to a measurement model. To evaluate non-close fit a RMSEA value of 0.05 is evaluated against an RMSEA value of 0.01, 0.01 representing an extremely good model fit. Statistical analysis was undertaken using the computer program R (R Core Team, 2019) and the following R packages, lavaan (Rosseel, 2012), semTools (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2020), psych (Revelle, 2019) and GPArotation (Bernaards & Jennrich, 2005).

Results

Replication of the PCA of Wootton et al. (2020) revealed identical item-factor loadings to those shown in Table 3. of their paper and identical proportion of the variance explained (55%) thus confirming the accuracy of the correlation matrix to the original dataset. Parallel analysis revealed one component (for PCA) and three factors (for factor analysis). EFA with maximum-likelihoods estimation, oblimin rotation and specifying a single factor (chi-square = 212.18, $df = 65$, $p < 0.05$) revealed a poor fit to data, RMSEA = 0.136, 90% confidence intervals (CI) 0.117-0.158, CFI = 0.85 and explaining 52% of the variance. Rerunning the EFA specifying a three-factor solution (chi-square = 82.55, $df = 42$, $p < 0.05$) based on the parallel analysis revealed an improved fit to data, RMSEA = 0.089, 90% CI 0.060-0.118, CFI = 0.96 and explaining (63%) of the variance. A chi-square differences test (chi-square diff. = 129.64, df diff. = 23, $p < 0.001$) indicated superiority of the three-factor model in model fit.

Scrutiny of item-factor loadings revealed however that two of the items (items 1. and 2.) cross-loaded between factors 1. and 2. and item 10. did not load significantly on any of the extracted factors (Table 1.). Consequently, a 10-item 3-factor model with these items removed was also considered for power analysis and sample size calculation.

TABLE 1. ABOUT HERE

Power analysis was undertaken on the single-factor model as specified by Wootton et al. (2020). Under the RMSEA close-fit method of MacCallum et al. (1996), it was observed that the study was underpowered (power = 0.605), a finding also observed for the three-factor 13-item model (power = 0.462) and the three-factor 10-item model (power = 0.271). Using the non-close fit method of MacCallum et al. (1996) confirmed that the original study was underpowered (power = 0.403), again a finding also observed for the three-factor 13-item model (power = 0.294) and the three-factor 10-item model (power = 0.168). Using the close fit sample size calculation method of MacCallum et al. (1996) suggested that the required sample size to evaluate the single-factor model was N=177, the three-factor model N=243 and the three-factor 10-item model N=473. Using the non-close fit sample size calculation method of MacCallum et al. (1996) suggested that the required sample size to evaluate the single-factor model was N=229, the three-factor model N=298 and the three-factor 10-item model N=508.

Discussion

The current investigation has yielded a number of useful insights into the measurement characteristics of the TSS and in addition has highlighted key statistical considerations that might be incorporated into a follow-on study/further research with the measure. Firstly, the

study represents an efficient approach to assessing further key measurement characteristics of a measure by the recreation of the pertinent aspects of the dataset through the construction (reconstruction) of correlation and covariance matrices. Clearly, such an approach is contingent on the original authors supplying sufficient information in the paper for such matrices to be constructed, but where this is possible, key aspects of replicability can be established and ambiguities in the original analysis identified or queried where these are observed. Additionally, as was found in the current study, reproduction of the original analysis confirms the rigour of the original findings statistically speaking as well as confidence in the constructed dataset for reanalysis.

The parallel analysis was of particular relevance in the current analysis because of the divergence observed between components and factors and the relationship of these dimensions to the type of analysis to be undertaken in terms of measurement model. Thus, it would seem to be entirely appropriate for Wootton et al. (2020) to focus on a unidimensional model within the context of the observed single component and preference for a PCA. However, the observation of three factors from the parallel analysis does raise the possibility of multi-dimensionality within the measure and also indicates the appropriateness of exploring such a possibility using EFA. Indeed, there is an implicit indication of potential multi-dimensionality within the measure by the authors' item generation procedure which had a focus on both *behavioural* and *cognitive* aspects of FoC.

The current analysis found preferential support for the three-factor model following EFA generally which demonstrated better model fit statistics than the single-factor model and specifically in terms of the statistically significant better fit to data of the three-factor model when compared to the single factor model using the chi-square differences test. This

evidence is both convincing, in relation to multi-dimensionality and problematic in terms of the split-loading items *'I worry about medical complications during pregnancy and/or childbirth'* and *'I worry about the type of delivery that I will have when I have a baby'* and the single item which did not load on any of the three factors *'I check excessively to determine if I am pregnant'*. Scrutiny of item content would suggest the factors within the tri-dimensional structure represent domains of fear (factor 1.), coping (factor 2.) and intrusive thoughts (factor 3.). This presents a conundrum in that a statistically preferential interpretation of the measure as a multi-dimensional tool also highlights limitations in that factor 3. (as a sub-scale) contains just two items which may be too few for effective screening of a clinically relevant domain of tokophobia in its own right. Further, this may be a similar issue for factor 2. comprising three items.

There is a reflective discussion to be had about the merits of scoring a measure as a uni-dimensional tool with a single summary score which may be practically easier but conceptually inconsistent with the underlying measurement model. This is of particular relevance since a thrust of the original tool development rationale was to avoid the length and complex scoring used in other FoC measures. Whether the TSS should be used as is, or further evaluated or developed, is a question that can only be addressed by further research, however one valuable element of the current study is that evidence-informed suggestions can be made regarding sample size for such future research enquiry, specified within each measurement model examined.

Power estimates of each model though markedly different between the two approaches (close fit/ non-close fit) used based on RMSEA. However, it is clear that for all the models evaluated, including the original TSS single-factor model the study was underpowered. The

sample sizes estimated for each model varied by model and method, quite dramatically in the case of the 10-item three-factor model, however between methods (close fit/non close fit) there was little difference between sample size calculations specific to a particular model. A parsimonious approach may therefore be to take the largest N (N=508) since these are based on the power to detect both good fitting and bad fitting models based on RMSEA. Therefore, with a minimum sample size of N=508 all three models detailed in the current paper could be evaluated within the context of an appropriately powered study design.

Wootton et al. (2020) reported against prediction poor divergent validity with the TSS showing a highly significant correlation with the PHQ-9 (Kroenke et al., 2001). This raises a question as to whether the TSS measures a construct of FoC or a related domain, for example and extrapolating from the PHQ-9 observations, depression. Examination of the item content of the TSS would in fact clearly indicate the content not to be depression-specific. However, an interesting aspect of the TSS is that most of the questions in the measure contain the word ‘worry’.

An investigation examining women’s preferences for measures to assess FoC by Sheen, Slade, Balling, and Houghton (2018) found that the Oxford Worries about Labour Scale (OWLS; Redshaw et al., 2009) was a preferred measure to assess their experiences of FoC, even though the OWLS was developed as a measure of worry rather than FoC *per se*. Worry, as distinct from FoC is nonetheless a significant component of FoC and importantly whether a measure like the OWLS assesses FoC, a possibility beyond its original operational parameters, is not currently clear, although women’s preferences for this tool in this context is clear. Therefore women’s experiences, in the case of the OWLS have shown a new perspective on use and application of the tool to FoC. Contrasting this with the TSS, which

followed a similar and 'classical' instrument development route as the OWLS, namely item generation by experts in the field and psychometric evaluation of key properties would also indicate that the TSS may benefit from the approach taken by Sheen et al. (2018) and be scrutinised by women themselves both in terms of preferences for use and whether it represents their experiences of FoC. Conversely and taking on board the observations of Sheen et al. (2018) regarding the OWLS, it is possible that the TSS may potentially be more circumscribed by worry in contrast to tokophobia. Clearly, a study with a methodology such as that of Sheen and colleagues (2018) is required to address this particular question.

The divergent validity findings also potentially highlights another issue. The high correlation between depression and FoC, as measured by the PHQ-9 may be explained by the literature that suggests the PHQ-9 might also be tapping into the construct of anxiety as co-morbid to depression (Shin, Ko, An, Yoon, & Han, 2020). Worry and anxiety are different in predictive content but highly correlated (Jomeen & Martin, 2005), perhaps suggesting that when women endorse worry questions they are endorsing something akin to anxiety and greater than just the construct of worry.

A replication study with sufficient sample size might also benefit from considering some modifications to the original study design, particularly related to sampling. Wootton et al.'s (2020) study would appear to have an over representation of well-educated participants either with university/college experience and/or academic success. In order to increase the relevance and accessibility of childbirth related fear pathways of care for less educated and more diverse groups of women, further research may therefore benefit from exploring the TSS as a tool for eliciting the most useful clinical information that is sensitive to underserved

populations, as perinatal mental illness disproportionately affects disadvantaged women and children (Ban et al., 2012; Raymond et al., 2014).

In summary, the investigation reanalysed the data of Wootton et al. (2020) by reconstruction of discrete aspects of the original dataset through summary data published in their paper.

Driven by the clinical need to develop a short, easily administered and easily scored measure of tokophobia, the TSS demonstrated excellent internal consistency and adequate convergent validity. However, the current investigation raises some questions regarding assumptions that can currently be made with confidence about the TSS, in particular the assumption of uni-dimensionality which is fundamental to the measurement model of the tool, and thus its use, scoring and interpretation. Wootton et al. (2020) highlighted that their study sample size was adequate for preliminary testing, supporting this assertion with a key reference (Clark & Watson, 1995) and also identified the need for further research with TSS including larger sample to assess psychometric properties more comprehensively. The current study thus gives a robust and evidence-informed indication of the required sample size for such a study where the focus is on the underlying dimensionality of the tool. However, it is important to reflect that the current study findings suggest that the original investigation was underpowered and this must thus be an important caveat to conclusions made regarding the measurement characteristics of the tool in the original paper. Further, the item content in relation to the three-factor solution identified by EFA suggests differentiated domains of fear, coping and intrusive thoughts all of which are relevant to tokophobia and moreover, the accruing evidence, even from the analysis of assumed unidimensional measures of FoC/tokophobia is that it is conceptually multi-dimensional (Konig, 2019).

References

- Ban, L., Gibson, J. E., West, J., Fiaschi, L., Oates, M. R., & Tata, L. J. (2012). Impact of socioeconomic deprivation on maternal perinatal mental illnesses presenting to UK general practice. *British Journal of General Practice*, *62*(603), e671-678.
doi:10.3399/bjgp12X656801
- Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*, 676-696. Retrieved from <http://www.stat.ucla.edu/research/gpa>
- Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment*, *85*, 17-32.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319.
- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, *6*, 147-168. doi:10.1177/1094428103251541
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*. Retrieved from <http://pareonline.net/getvn.asp?v=10&n=7>
- Dencker, A., Nilsson, C., Begley, C., Jangsten, E., Mollberg, M., Patel, H., . . . Sparud-Lundin, C. (2019). Causes and outcomes in studies of fear of childbirth: A systematic review. *Women and Birth*, *32*(2), 99-111. doi:10.1016/j.wombi.2018.07.004
- Haines, H., Pallant, J. F., Karlstrom, A., & Hildingsson, I. (2011). Cross-cultural comparison of levels of childbirth-related fear in an Australian and Swedish sample. *Midwifery*, *27*(4), 560-567. doi:10.1016/j.midw.2010.05.004

- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Jomeen, J., & Martin, C. R. (2005). The factor structure of the Cambridge Worry scale in early pregnancy. *Journal of Prenatal and Perinatal Psychology and Health*, 20.
- Jones, C., Marshall, C., Martin, C. R., & Jomeen, J. (2020). Pregnant with fear. *Community Practitioner*, 2. Retrieved from <https://www.communitypractitioner.co.uk/features/2020/02/pregnant-fear>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). semTools: Useful tools for structural equation modeling. R package version 0.5-3. Retrieved from <https://CRAN.R-project.org/package=semTools>
- Klabbers, G. A., Wijma, K., Paarlberg, K. M., Emons, W. H. M., & Vingerhoets, A. (2019). Haptotherapy as a new intervention for treating fear of childbirth: a randomized controlled trial. *Journal of Psychosomatic Obstetrics and Gynaecology*, 40(1), 38-47. doi:10.1080/0167482X.2017.1398230
- Kline, P. (1994). *An Easy Guide to Factor Analysis*. Abingdon: Routledge.
- Konig, J. (2019). The German W-DEQ version B-Factor structure and prediction of posttraumatic stress symptoms six weeks and one year after childbirth. *Health Care for Women International*, 40(5), 581-596. doi:10.1080/07399332.2019.1583230
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606-613. doi:doi.org/10.1046/j
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149.

- Martin, C., Jones, C., Marshall, C. A., Huang, C., Reeve, J., Fleming, M. P., . . . Jomeen, J. (2020). Fear of childbirth measurement: appraisal of the content overlap of four instruments. *Journal of Reproductive and Infant Psychology*, 1-13.
doi:10.1080/02646838.2020.1861226
- Nilsson, C., Hessman, E., Sjoblom, H., Dencker, A., Jangsten, E., Mollberg, M., . . . Begley, C. (2018). Definitions, measurements and prevalence of fear of childbirth: a systematic review. *BMC Pregnancy and Childbirth*, 18(1), 28. doi:10.1186/s12884-018-1659-7
- O'Connell, M. A., Leahy-Warren, P., Kenny, L. C., O'Neill, S. M., & Khashan, A. S. (2019). The prevalence and risk factors of fear of childbirth among pregnant women: A cross-sectional study in Ireland. *Acta Obstetrica et Gynecologica Scandinavica*, 98(8), 1014-1023. doi:10.1111/aogs.13599
- O'Connell, M. A., Leahy-Warren, P., Khashan, A. S., Kenny, L. C., & O'Neill, S. M. (2017). Worldwide prevalence of tocophobia in pregnant women: systematic review and meta-analysis. *Acta Obstetrica et Gynecologica Scandinavica*, 96(8), 907-920.
doi:10.1111/aogs.13138
- Pazzagli, C., Laghezza, L., Capurso, M., Sommella, C., Lelli, F., & Mazzeschi, C. (2015). Antecedents and consequences of fear of childbirth in nulliparous and parous women. *Infant Mental Health Journal*, 36(1), 62-74. doi:10.1002/imhj.21483
- R Core Team. (2019). R: A language and environment for statistical computing. . Vienna, Austria.: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Raymond, N. C., Pratt, R. J., Godecker, A., Harrison, P. A., Kim, H., Kuendig, J., & O'Brien, J. M. (2014). Addressing perinatal depression in a group of underserved urban

women: a focus group study. *BMC Pregnancy and Childbirth*, 14, 336.

doi:10.1186/1471-2393-14-336

Redshaw, M., Martin, C., Rowe, R., & Hockley, C. (2009). The Oxford Worries about Labour Scale: women's experience and measurement characteristics of a measure of maternal concern about labour and birth. *Psychology, Health and Medicine*, 14(3), 354-366. doi:10.1080/13548500802707159

Revelle, W. (2019). psych: Procedures for Personality and Psychological Research (Version 1.9.12). Evanston, Illinois: Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>

Richens, Y., Smith, D. M., & Lavender, D. T. (2018). Fear of birth in clinical practice: A structured review of current measurement tools. *Sexual and Reproductive HealthCare*, 16, 98-112. doi:10.1016/j.srhc.2018.02.010

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1-36. Retrieved from <http://www.jstatsoft.org/v48/i02/>.

Sheen, K., Slade, P., Balling, K., & Houghton, G. (2018). *An examination of clarity, acceptability and content validity of existing questionnaires for the measurement of fear of childbirth in a UK population*. Paper presented at the Society for Reproductive and Infant Psychology Annual conference, Lodz, Poland.

Shin, C., Ko, Y. H., An, H., Yoon, H. K., & Han, C. (2020). Normative data and psychometric properties of the Patient Health Questionnaire-9 in a nationally representative Korean population. *BMC Psychiatry*, 20(1), 194. doi:10.1186/s12888-020-02613-0

Slade, P., Pais, T., Fairlie, F., Simpson, A., & Sheen, K. (2016). The development of the Slade–Pais Expectations of Childbirth Scale (SPECS). *Journal of Reproductive and Infant Psychology*, 34(5), 495-510. doi:10.1080/02646838.2016.1209300

Wijma, K., Wijma, B., & Zar, M. (1998). Psychometric aspects of the W-DEQ; a new questionnaire for the measurement of fear of childbirth. *Journal of Psychosomatic Obstetrics and Gynaecology*, *19*(2), 84-97. doi:10.3109/01674829809048501

Wootton, B. M., Davis, E., Moses, K., Moody, A., & Maguire, P. (2020). The development and initial validation of the Tokophobia Severity Scale. *Clinical Psychologist*. doi:10.1111/cp.12220

Table 1. TSS Item-factor loadings following maximum-likelihood factor extraction and oblimin rotation. Significant (>.30) and single factor item-factor loadings are shown in bold.

| Item | Factor 1. | Factor 2. | Factor 3. |
|---|-------------|-------------|-------------|
| TSS1 I worry about medical complications during pregnancy and/or childbirth | 0.33 | 0.47 | -0.17 |
| TSS2 I worry about the type of delivery that I will have when I have a baby | 0.35 | 0.49 | 0.04 |
| TSS3 I worry that something terrible will happen to me during my pregnancy and/or childbirth | 0.90 | -0.04 | -0.02 |
| TSS4 I worry that something terrible will happen to my baby during my pregnancy and/or childbirth | 0.63 | 0.28 | -0.18 |
| TSS5 I worry that I will not be able to cope with the pain of pregnancy and/or childbirth | -0.12 | 0.98 | 0.06 |
| TSS6 I worry about the medical procedures required during pregnancy and/or childbirth | 0.17 | 0.72 | 0.07 |
| TSS7 I avoid talking about pregnancy, children and childbirth because of my fears | 0.26 | 0.10 | 0.51 |
| TSS8 I worry that I will not be in control of the medical procedures during my pregnancy and/or delivery | 0.51 | 0.26 | -0.17 |
| TSS9 I worry that pregnancy and/or childbirth will be too painful | 0.11 | 0.78 | 0.00 |
| TSS10 I check excessively to determine if I am pregnant | 0.11 | 0.24 | 0.28 |
| TSS11 I have nightmares about being pregnant and/or delivering a child | 0.07 | 0.20 | 0.58 |
| TSS12 I worry that something dangerous will happen to me during pregnancy or the delivery (e.g., a ruptured uterus, preeclampsia, emergency interventions, death) | 0.87 | -0.04 | 0.17 |
| TSS13 I worry that something dangerous will happen to my child during pregnancy or the delivery (e.g., injury or death) | 0.77 | 0.08 | 0.10 |