JAN interactive contribution

Commentary on Drevin et. Al. (2017) Measuring pregnancy planning: A psychometric evaluation comparison of two scales

Dear Editor-in-Chief,

As those who have developed and evaluated a variety of language versions of the London Measure of Unplanned Pregnancy (LMUP), we welcome Drevin et al.'s (2017) new evaluation of the LMUP. Drevin et al. compare a translated Swedish version of the LMUP with a single question named the "Swedish Pregnancy Planning Scale" (SPPS). This asks (in Swedish) "How planned was your current pregnancy?" with the response options "highly planned", "quite planned", "neither planned nor unplanned", "quite unplanned", and "highly unplanned". They make the surprising admission that, without cognitive interviews, they do not know how women interpreted the question so a key aspect of validity is unknown. Given previous work about the variability of understanding of terms such as "planned" (Barrett and Wellings, 2002) this seems a high risk measurement strategy.

Drevin et al. state that "pregnancy planning is a concept that is difficult to measure due to the complexity of the concept" (2017, p.2). They continue this argument throughout their background section, thus seeming to suggest the need for a latent-trait model of measurement, i.e. that the concept is not easily observable and is hard to measure with a single question. Yet this is exactly what they propose. A single question of a latent construct is inherently prone to greater measurement error than a multi-item validated measure. Many of the tests of reliability and validity which the authors applied to the LMUP simply cannot be applied to the SPPS.

We have some concerns with how the evaluation of Swedish LMUP was conducted. The steps in the translation/cultural adaptation and evaluation of psychometric measures are well established. The authors report the translation and back translation of the LMUP, but no cognitive testing was carried out. Furthermore, the sample was based on women recruited via antenatal clinics, which (by omitting those with pregnancies ending in abortion) means that a portion of the construct (the less planned end of the pregnancy planning continuum) was poorly represented. This may be significant given that analyses based on Classical Test Theory (as these are) may be affected by the range of the construct contained within the sample. Certainly, the authors report a strong left skew to their LMUP scores (towards the more planned end of the spectrum). Unusually, the authors reported the

split-half reliability of the LMUP items (items 1-3 vs items 4-6); Cronbach's alpha is normally reported as it is the average of all possible split-half coefficients. It would also have been useful if the authors had reported the item-rest correlations and the range of the inter-item correlations, as this would have given more detail on the internal consistency of the Swedish LMUP. The authors reported Spearman's correlation coefficient for test-retest reliability; weighted Kappa should have been used given it is a measure of agreement rather than correlation (i.e. if all scores had risen by one point in the re-test the correlation using Spearman's coefficient would have been excellent, though the agreement would not have been).

Drevin et al. are disingenuous when they say that the LMUP "has previously not been psychometrically evaluated using a method that tests the fit of the pre-specified London Measure of Unplanned Pregnancy model" (2017, p2). In fact, the LMUP has been psychometrically validated, including using methods that test the fit of the pre-specified LMUP model, in ten language versions across eight countries (LMUP publications, 2018) with further studies underway. While confirmatory factor analysis may not have been done previously, the unidimensionality of the LMUP items has been assessed in all psychometric evaluations except one by means of Principal Components Analysis or Principal Axis Factoring. These are methods in the exploratory factor analysis family, often used in a hypothesis testing role and used appropriately with new translations. Running a confirmatory factor analysis on the second field test of the original UK development and evaluation study produces the following standardized factor loadings: item 1, 0.62; item 2, 0.88; item3 – 0.93; item 4 – 0.90, item 5 – 0.86; and item 6, 0.68; with good model fit (CFI, 0.99; SRMR, 0.01; RMSEA, 0.07, 90% CI 0.04 to 0.09). Unsurprisingly, the factor loadings are extremely similar to those produced by the principal component analyses in the development study and subsequent evaluations, confirming what we already know about the fit of the LMUP. The authors also make much of their finding of "item reliability", including it in their key findings. Again, this is unusual. The "item reliability" is the square of the standardized factor loading in the confirmatory factor analysis, rarely reported because it is implied by the factor loading (which the authors present in table 2). The authors did find that all six LMUP items were measuring one construct (i.e. fitting the pre-specified unidimensional LMUP model) but they did not include this in their key findings.

On the basis of their confirmatory factor analysis, Drevin et al. recommend removing one, and possibly two, LMUP items, both of which measure behaviour. Whilst revision of established measures does happen, one has to consider how these changes relate to the underpinning qualitative work/conceptual model and, in this case, the contribution of the behaviour items to content validity, despite their lower statistical coherence. Indeed, the authors could have carried out sensitivity analyses relating to these items, as has been done in previous studies. These analyses have supported retaining these items given that they do not affect the performance of the scale overall and because there are good reasons for the performance of these items, such as reflecting unmet need for contraception or low awareness of preconception care, which may change over time and can be detected using the LMUP.

Drevin et al. conclude that researchers should use the SPPS rather than the LMUP. We believe that researchers, however, should be aware of the limitations of this single question, some of which we have detailed here, and, in contrast, the body of work that underpins the LMUP, particularly that the LMUP meets internationally accepted standards of psychometric validation (U.S. Department of

Health and Human Services Food and Drug Administration, 2009; Mokkink et al, 2010a, 2010b; Reeve et al, 2013) whereas the SPPS does not.

References

Barrett, G., Wellings, K. (2002) What is a "planned" pregnancy? Empirical data from a British study. Social Science and Medicine 55:545-557. https://doi.org/10.1016/S0277-9536(01)00187-3

Drevin, J., Kristiansson, P., Stern, J., Rosenblad, A. (2017) Measuring pregnancy planning: A psychometric evaluation comparison of two scales. *Journal of Advanced Nursing,* 00:1–11. https://doi.org/10.1111/jan.13364

LMUP publications. (2018) www.lmup.org.uk/pubs.htm

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. (2010a) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes*, Journal of Clinical Epidemiology*, 2010, 63:737-745. doi:10.1016/j.jclinepi.2010.02.006

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. (2010b) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study, *Quality of Life Research*, 2010, 19:539-549 doi:10.1007/s11136-010-9606-8

Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Schwartz C, Revicki DA, Moinpour CM, McLeod LD, Lyons JC, Lenderking WR, Hinds PS, Hays RD, Greenhalgh J, Gershon R, Feeny D, Fayers PM, Cella D, Brundage M, Ahmed S, Aaronson NK, Butt Z. (2013) ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research* 22:1189-1905 doi:10.1007/s11136-012-0344-y

U.S. Department of Health and Human Services Food and Drug Administration.(2009) *Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims.* *http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM 193282.pdf*

Dr Geraldine Barrett, PhD,
Principal Research Associate, Institute for Women's Health, University College London, London WC1E 6AU
g.barrett@ucl.ac.uk

Dr Jennifer A. Hall, PhD
Principal Clinical Researcher, Institute for Women's Health, University College London, London WC1E 6AU
Jennifer.hall@ucl.ac.uk

Dr Ana Luiza Vilela Borges, PhD

Associate Professor, Department of Public Health Nursing, University of São Paulo School of Nursing, São Paulo, Brazil

alvilela@usp.br

Dr Corinne Rocca, PhD

Associate Professor, Bixby Center for Global Reproductive Health, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, U.S.A

Corinne.rocca@ucsf.edu

Dr Eman Almaghaslah, MPH

Health Promotion Officer and Medical Resident, Primary Health Care Administration and Preventive Health Department in Qatif, Saudi Arabian Ministry of Health, Qatif, Eastern Province, Saudi Arabia

e.almaghaslah@gmail.com

Professor Judith Stephenson, FFPH

Professor of Sexual and Reproductive Health, Institute for Women's Health, University College London, London WC1E 6AU

Judith.stephenson@ucl.ac.uk