# Utilising Computerised Adaptive Testing to Alleviate Respondent Burden in Maternal Stress Assessment

**Authors:**

Chloe Wei En Teo[1], Chua Hng Teng Rachelle[2], Rishabh Garg[3], Malorie Rui Yi Yoong[2], Ethel Siew Ee Tan[4], Nikita Rane[4], Nawal Binti Mohamed Hashim[4], Helen Chen[5], Kok Hian Tan[6], Shiao-Yng Chan[7], Jonathan Huang[8], Michelle Kee[9], Cornelia Chee[10], Tan Hong Ming[11], Jussi Keppo[12], Michael Meaney[13], Geoffrey Chern-Yee Tan[4,*]

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, 636921, Singapore
[2]National University of Singapore, NUS Faculty of Arts and Social Sciences, Singapore, 117570, Singapore
[3]National University of Singapore, NUS Faculty of Science, Singapore, 117546, Singapore
[4]Institute of Mental Health Singapore, Department of Mood and Anxiety, Singapore, 539747, Singapore
[5]KK Women's and Children's Hospital, Department of Psychological Medicine, Singapore, 229899, Singapore
[6]KK Women's and Children's Hospital, Perinatal Audit & Epidemiology, Singapore, 229899, Singapore
[7]National University Hospital, Department of Obstetrics & Gynaecology, Singapore, 119074, Singapore
[8]A*STAR Singapore Institute for Clinical Sciences (SICS), Department of Biostatistics, 117609, Singapore
[9]A*STAR Singapore Institute for Clinical Sciences (SICS), Department of Translational Neuroscience, 117609, Singapore
[10]Yong Loo Lin School of Medicine NUS, Department of Psychological Medicine, Singapore, 119228, Singapore
[11]National University of Singapore, Institute of Operations Research and Analytics NUS, Singapore, 117602, Singapore
[12]National University of Singapore, NUS Business School, Singapore, 119245, Singapore
[13]A*STAR Singapore Institute for Clinical Sciences (SICS); Brenner Centre for Molecular Medicine, Singapore, 117609, Singapore
*Corresponding: geoffrey_tan@imh.com.sg

## ABSTRACT

Maternal mental health plays a pivotal role in perinatal care, with far-reaching implications for both maternal well-being and child development. Amidst global challenges like the COVID-19 pandemic, the demand for effective mental health assessment has surged. In response, this study investigates the utility of Computerised Adaptive Testing (CAT) for profiling maternal stress. Using the Growing Up in Singapore Towards Healthy Outcomes (GUSTO) dataset, we focused on the Maternal Stress module, comprising the Perceived Stress Scale (PSS) and Parental Stress Index (PSI). We generated an item pool from both scales and employed the Graded Response Model (GRM) for calibration. A final item bank of 105 items was consolidated. Using the Concerto software, we devised a CAT questionnaire for test administration. We then used the 'Firestar' R package to simulate the CAT with various stopping criteria, revealing substantial question reduction (up to 84.8%) while still maintaining high correlations with true theta scores (up to 99.9%). Nonetheless, limitations include assumptions in simulations, item calibration, and a specific focus on maternal stress. This study underscores CAT's potential to streamline assessments, enhancing perinatal mental health evaluations, while signaling the need to further explore CAT's potential in this domain.


Keywords: Maternal health, Stress, Computerized adaptive testing, Healthcare


## INTRODUCTION

Maternal mental health constitutes a pivotal facet of perinatal care, having considerable implications on both mother and child. Against the backdrop of recent global developments, most notably the COVID-19 pandemic, optimising maternal mental health has become more imperative than ever[1]. Notably, the negative impacts of maternal stress can manifest itself both directly and indirectly. For instance, prenatal stress has been found to result in increased inflammatory and maternal hypothalamo-pituitatry-adrenal (HPA) axis activity, directly affecting fetal HPA development[2]. In addition, this increased HPA axis activation resulting from maternal stress has also been linked to other fetal outcomes such as low birth weight and pre-term birth[3]. Besides these detrimental effects on one's child, prenatal stress is also linked to prenatal complications such as pre-eclampsia[4].

Yet in the busy clinician's psychiatric practice, conducting comprehensive and extensive psychiatric assessments can be challenging; this study emerges as a response to this challenge, aiming to explore novel strategies for efficient empirical assessment of the psychiatric patient while still upholding standards of accuracy.

Prior research underscores the potential of measures aimed at reducing assessment burden to enhance outcomes of measurement-based clinical decision making[5].

Consequently, this study delves into the feasibility of leveraging Computerised Adaptive Testing (CAT) to alleviate respondent burden while still maintaining accuracy in the administration of psychometric questionnaires, specifically within the realm of maternal mental health.

CAT is an innovative assessment methodology that optimises the efficiency of psychological evaluations, dynamically tailoring each assessment to the respondents' unique ability levels by selecting the next question to administer based on previous responses. CAT begins with an initial question, typically of moderate difficulty, and based on the response, the individual's level of ability in the trait measured is assessed. The algorithm then selects the next question that will be the most informative given the previous response[6], continuing this process until the pre-determined stopping rule is fulfilled[7]. Finally, the participant's ability level is estimated and represented by the value 'theta'.

The ability of CAT to adapt to individual ability levels in real time hence not only minimises the number of questions administered, but also hones in on specific areas of concern[7]. That precision at the individual level is ensured despite a reduction in the number of questions administered is paramount, especially within the demographic of pregnant mothers. Given the mounting responsibilities and pressures faced by the modern mother, being able to monitor maternal well-being during this vulnerable period in a concise, yet precise manner is more important than ever. This also enables CAT to be administered repeatedly, as opposed to long, tedious questionnaires, which is especially relevant when it comes to monitoring and observing changes in a mother's mood over time, without compromising on engagement level.

In longitudinal cohort studies like the Growing Up in Singapore Towards Healthy Outcomes (GUSTO) study, an array of factors such as sleep, depressive symptoms and familial structure have been demonstrated to be pertinent in maternal mental health[8]. This study focuses on one such factor, namely maternal stress. Current studies show that maternal stress has severe implications for both mother and child; for example, stressful life events in the prenatal period are negatively associated with mental development and positively associated with fearfulness in infancy[9]. Additionally, maternal stress has also shown to have lasting impacts on children, with studies suggesting that early postnatal stress affects the child's future reactivity to stress, possibly by altering developing neural circuits responsible for the neuroendocrine response[10].

In response, this study investigates CAT's ability to accurately assess respondents' abilities while reducing the number of questions administered in the sphere of maternal stress, using the Parental Stress Index (PSI) and the Perceived Stress Scale (PSS). Both the PSS and PSI have been shown to be useful and valid instruments to

measure mental well being in mothers during the perinatal period[11,12]. By examining CAT's efficacy in gauging maternal stress levels, this study aims to further enhance the methodological toolkit available to the clinician for perinatal mental health assessment. This study hence not only aligns with modern healthcare demands, but is also a step forward in the pursuit of personalised medicine.

# METHODOLOGY

### Study design and data preparation

Questionnaires from the GUSTO dataset were categorised into distinct modules based on the trait examined. Participants from GUSTO comprised Singaporean pregnant women at 7-11 weeks of pregnancy between June 2009 to October 2010[13]. Responses from a total of 736 participants were used.

In this study, we focused on the Maternal Stress module, which comprised the Perceived Stress Scale (PSS) and Parental Stress Index (PSI). Each questionnaire is a 'scale', and individual questions within the scale are termed 'items'. The full-length PSS and PSI scales comprised 10 and 120 questions, respectively. The exclusion criteria for items are explained below.

### Item pool generation and data analysis

First, we generated an item pool by consolidating all 130 questions from both scales. We obtained and analysed item-level data containing response breakdowns for each respondent to identify and eliminate unsuitable items. To achieve this, we divided the dataset into 7 subgroups of approximately equal sizes and removed items with insufficient responses across response categories[14]; this is done to ensure sufficient responses for cross-validation to be done as explained later. Based on this criterion, 2 items were excluded.

### Item calibration using Graded Response Model (GRM)

Given the ordered, categorical nature of the response categories in the items, the Graded Response Model (GRM) was chosen, to which the items would be calibrated. The GRM is a variant of the Item Response Theory (IRT) model that handles ordered, polytomous categories, such as 'strongly agree', 'agree', and so forth[15], as seen in the PSS and PSI scales.

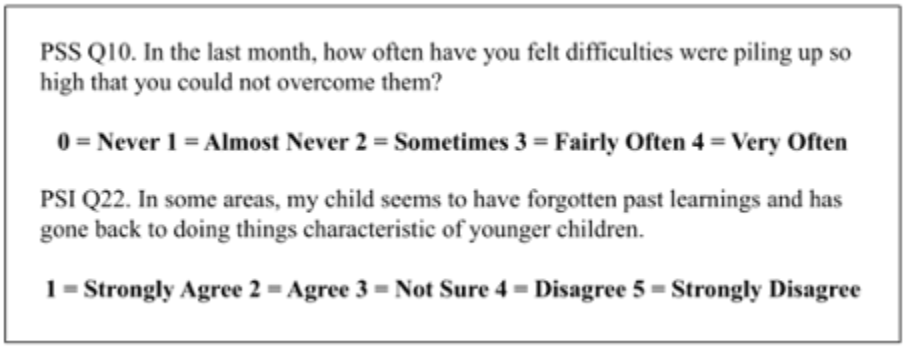


*Figure 1: Example of items in PSS and PSI scales*

Subsequently, using the 'ltm' package in RStudio, items were fitted to the GRM to calculate their discrimination and extremity parameters. Items featuring negative discrimination parameters were reviewed and removed[16]; 23 items were hence removed and the remaining 105 items were compiled into a final item bank.

**Utilisation of CAT platform**

Using the Concerto software, we created a CAT test to measure maternal stress. . The test included an introductory page containing respondent instructions, the test segment, and a results page displaying the respondent's score. The item bank curated earlier was then integrated into the CAT test.

**CAT simulation using Firestar R package**

To evaluate CAT's accuracy in predicting respondents' ability levels, we employed the 'Firestar' R package to simulate the CAT. This simulation entailed a 7-fold cross-validation approach[17], wherein the GUSTO dataset was divided into 7 approximately equal subsets (folds). Of these, 6 folds were used for model training using item discrimination and extremity parameters.

The remaining fold was designated as the test set to evaluate CAT's accuracy, with each respondent's response pattern used to simulate the test. This simulation yielded the Pearson correlation coefficient (r), comparing CAT-based theta estimates to the full-length true theta scores. The root mean squared error (RMSE) was also calculated based on the theta estimate and true theta scores. This process was repeated 6 additional times for the remaining folds, with each iteration simulating the CAT under 3 different stopping criteria (SEM $<0.2$, $<0.3$, $<0.4$).

**Ethical Considerations**

In utilising the GUSTO dataset, we prioritise the principles of informed consent, ensuring that participants' rights and autonomy are respected. To safeguard anonymity and privacy, we rigorously anonymise data and refrain from disclosing personally identifiable information. Additionally, we maintain transparency

regarding our objectives and methods, aligning them with the dataset's original purpose. This study was conducted aligned with the guidelines of the Declaration of Helsinki and was approved by the National Healthcare Group Domain Specific Review Board (reference D/09/021) and the SingHealth Centralised Institutional Review Board.

## RESULTS

Simulating the CAT using Firestar under a stopping rule with a maximum standard error (max.SEM) of 0.2, the number of questions administered was reduced to a mean of 49.3. This was achieved while maintaining a 99.9% correlation with the true theta scores obtained from a full-length test (r = 0.999), and an RMSE of 0.0470.

Upon increasing the threshold of the stopping criteria to max.SEM<0.3, the number of questions administered decreased to a mean of 16.0, with a lower correlation of 97.7% (r=0.977) and a higher RMSE of 0.220. Likewise, when max.SEM was set to <0.4, a mean of 9.38 questions were administered, with the resulting theta estimates having a 95.2% correlation with true theta scores (r = 0.952), and an RMSE of 0.318.

| Test fold | max.SEM | Number of items used | | Mean standard error | Pearson correlation coefficient | Root mean-squared error |
|---|---|---|---|---|---|---|
| | | Mean | Standard deviation | | | |
| Fold 1 | <0.2 | 55.0 | 40.7 | 0.447 | 1.00 | 0.0340 |
| | <0.3 | 17.2 | 10.7 | 0.499 | 0.988 | 0.186 |
| | <0.4 | 9.94 | 3.26 | 0.554 | 0.967 | 0.307 |
| Fold 2 | <0.2 | 50.0 | 37.4 | 0.423 | 0.999 | 0.0465 |
| | <0.3 | 16.0 | 8.16 | 0.475 | 0.973 | 0.232 |
| | <0.4 | 9.39 | 2.71 | 0.531 | 0.942 | 0.340 |
| Fold 3 | <0.2 | 42.4 | 30.4 | 0.397 | 0.998 | 0.0669 |
| | <0.3 | 14.2 | 5.56 | 0.456 | 0.971 | 0.232 |
| | <0.4 | 8.58 | 2.56 | 0.517 | 0.948 | 0.310 |
| Fold 4 | <0.2 | 45.5 | 35.1 | 0.415 | 0.998 | 0.0656 |
| | <0.3 | 15.6 | 9.47 | 0.468 | 0.980 | 0.235 |
| | <0.4 | 9.36 | 3.11 | 0.524 | 0.958 | 0.345 |
| Fold 5 | <0.2 | 55.2 | 43.2 | 0.446 | 1.00 | 0.0225 |
| | <0.3 | 16.4 | 10.6 | 0.495 | 0.976 | 0.223 |
| | <0.4 | 9.54 | 1.88 | 0.546 | 0.954 | 0.306 |
| Fold 6 | <0.2 | 56.9 | 40.9 | 0.426 | 0.999 | 0.0331 |
| | <0.3 | 17.8 | 10.6 | 0.477 | 0.973 | 0.230 |
| | <0.4 | 9.70 | 2.13 | 0.534 | 0.949 | 0.319 |
| Fold 7 | <0.2 | 39.8 | 31.9 | 0.465 | 0.998 | 0.0604 |
| | <0.3 | 14.6 | 8.40 | 0.514 | 0.975 | 0.202 |
| | <0.4 | 9.11 | 2.40 | 0.565 | 0.946 | 0.298 |

*Figure 2: Results of simulation in Firestar*

| max.SEM | Average number of items used | Average correlation coefficient | Average RMSE |
|---|---|---|---|
| <0.2 | 49.3 | 0.999 | 0.0470 |
| <0.3 | 16.0 | 0.977 | 0.220 |
| <0.4 | 9.38 | 0.952 | 0.318 |

*Figure 3: Average number of items used for each maximum SEM threshold, with their corresponding correlation and RMSE*

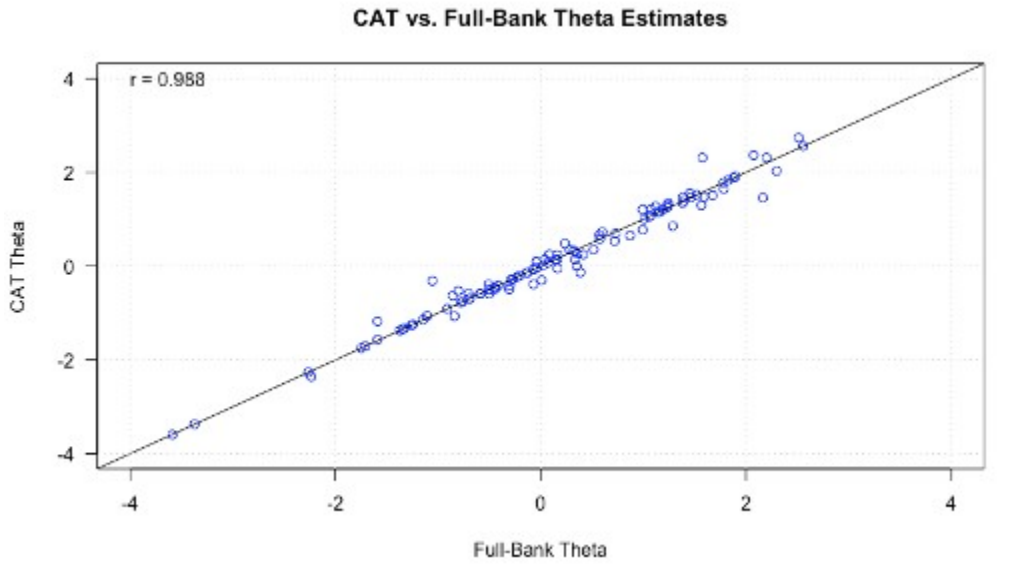


*Figure 4: Graph of CAT vs Full-Bank Theta Estimates in Fold 1, max.SEM <0.3, as shown in Firestar*

## DISCUSSION

The outcomes of the CAT simulation offer significant insights into the potential efficacy of the adaptive approach for maternal mental health assessment. The utilisation of CAT aimed to strike a balance between reducing the number of administered questions and maintaining a high degree of precision in estimating maternal mental health profiles.

One of the key findings of this study was the trade-off between assessment length and precision. When the max.SEM threshold was set to 0.2, question load was reduced to a mean of 49.3 questions. This assessment yielded a precise assessment of the participants' theta scores, exemplified by a 99.9% correlation with true theta scores and an RMSE of 0.0470. Consequently, when the max.SEM threshold was increased to 0.3, the question load was further decreased, but at the cost of

precision, with a lower correlation and an increased RMSE value. It is important to note, however, that even with the relaxed max.SEM threshold of 0.4, the observed correlation of 95.2%, while decreased, is still indicative of CAT's potential to deliver remarkably accurate estimates, which is significant in the face of traditional concerns over losing important information when shortening questionnaires[6].

These findings illuminate the essential consideration faced by clinicians and researchers in the use of CAT. The ability to customise assessments by adjusting stopping criteria presents an intriguing opportunity to balance respondent burden and precision. While existing literature has illustrated that the direct link between questionnaire length and response quality may not necessarily be clear[18], many authors still recommend keeping questionnaires as concise as possible while striving to maintain accuracy. This is especially so in healthcare, where patients' health status and functional disability may affect response quality[19,20].

As such, we present this trade-off not as a disadvantage of CAT, but as a tool that clinicians may strategically leverage. Clinicians may fine-tune the stopping criteria to align with the unique demands of specific assessments. In situations where more comprehensive understanding of an individual's mental health profile is paramount, a lower max.SEM threshold may be warranted. Conversely, when expediency is crucial, a higher threshold can reduce respondent fatigue without severely compromising the utility of the assessment.

The significance of this study is highlighted when reviewing existing literature; while CAT tests have been successfully constructed for specific questionnaires[21] and in other populations[22], few studies venture into implementing CAT in maternal mental health. Given the benefits of CAT in alleviating respondent burden as illustrated above, CAT can be utilised as a powerful tool to optimise maternal mental health today. For example, administering CAT via online platforms may render mental health services more accessible to mothers in the comfort of their homes. Presently, there is promising research suggesting the benefits of Internet-based mental health services for the perinatal population[23]; the integration of CAT into these services can provide healthcare providers with information about the individual's specific vulnerabilities and psychological profile. Simultaneously, the patient is spared from multiple long questionnaires before they can benefit from therapeutic measures.

While this study presents valuable insights into the application of CAT for maternal mental health assessment, certain limitations warrant consideration. Firstly, the simulation framework, while offering a controlled environment, relies on assumptions that may not entirely mirror real-world complexities3. Variability in respondent characteristics and responses, which may not be fully captured within the simulation, could impact the generalisability of findings. Moreover, the calibration of items using GRM assumes that the model accurately captures the

relationships between items and latent traits. Deviations from these assumptions could affect the item bank's quality and subsequently the CAT's performance.

Lastly, the study's focus on maternal stress, while important, represents a specific facet of maternal mental health, potentially limiting the generalisability of findings to a broader spectrum of mental health indicators. Overall, these limitations underline the need for caution in extrapolating findings and highlight areas for future refinement.

## CONCLUSION

The results of this study underscore the potential advantages of adopting a CAT approach in maternal mental health assessment. The ability to drastically reduce the number of questions while retaining remarkable correlation with true theta scores speaks to the efficiency and accuracy of CAT. This approach not only addresses participant burden but also opens avenues for wider implementation of maternal mental health assessments, particularly in settings where resources are limited or participant engagement is a concern.

In conclusion, this study presents compelling evidence for the viability of adaptive testing techniques in profiling maternal mental health. The findings provide a foundation for future research and application of CAT in maternal mental health assessment, potentially revolutionising the way mental health evaluations are conducted for the care of expectant and new mothers.

## DATA AVAILABILITY

The data used is available upon request and approval by the GUSTO team.

## REFERENCES


1. Hessami, K., Romanelli, C., Chiurazzi, M. & Cozzolino, M. COVID-19 pandemic and maternal mental health: a systematic review and meta-analysis. *J Matern Fetal Neonatal Med* **35**, 4014–4021 (2022).
2. Ruffaner-Hanson, C. *et al.* The maternal-placental-fetal interface: Adaptations of the HPA axis and immune mediators following maternal stress and prenatal alcohol exposure. *Experimental Neurology* **355**, 114121 (2022).
3. Lobel, M. *et al.* Pregnancy-specific stress, prenatal health behaviors, and birth

outcomes. *Health Psychology* **27**, 604–615 (2008).

4. Schneider, S. *et al.* Risk groups and maternal-neonatal complications of preeclampsia – Current results from the national German Perinatal Quality Registry. *Journal of Perinatal Medicine* **39**, (2011).
5. Gibbons, R. D. *et al.* Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr Serv* **59**, 361–368 (2008).
6. Walker, J., Böhnke, J. R., Cerny, T. & Strasser, F. Development of symptom assessments utilising item response theory and computer-adaptive testing--a practical method based on a systematic review. *Crit Rev Oncol Hematol* **73**, 47–67 (2010).
7. Harrison, C. J. *et al.* Computerised adaptive testing accurately predicts CLEFT-Q scores by selecting fewer, more patient-focused questions. *J Plast Reconstr Aesthet Surg* **72**, 1819–1824 (2019).
8. Growing up in Singapore towards Healthy Outcomes. GUSTO Data Vault.
9. Bergman, K., Sarkar, P., O'Connor, T. G., Modi, N. & Glover, V. Maternal stress during pregnancy predicts cognitive ability and fearfulness in infancy. *J Am Acad Child Adolesc Psychiatry* **46**, 1454–1463 (2007).
10. Shonkoff, J. P., Garner, A. S., Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, & Section on Developmental and Behavioral Pediatrics. The lifelong effects of early childhood adversity and toxic stress. *Pediatrics* **129**, e232-246 (2012).
11. Katus, L. *et al.* Perceived stress during the prenatal period: assessing measurement invariance of the Perceived Stress Scale (PSS-10) across cultures and birth parity. *Arch Womens Ment Health* **25**, 633–640 (2022).
12. A measure of parental stress in mothers with small children: dimensionality,

stability and validity - Östberg - 1997 - Scandinavian Journal of Psychology - Wiley Online Library. https://onlinelibrary-wiley-com.libproxy1.nus.edu.sg/doi/10.1111/1467-9450.00028.

13. About the Cohort. (2023).
14. Sijtsma, K. & Molenaar, I. *Introduction to Nonparametric Item Response Theory*. (SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America, 2002). doi:10.4135/9781412984676.
15. Samejima, F. Graded Response Model. in *Handbook of Modern Item Response Theory* (eds. Van Der Linden, W. J. & Hambleton, R. K.) 85–100 (Springer New York, New York, NY, 1997). doi:10.1007/978-1-4757-2691-6_5.
16. DeMars, C. *Item Response Theory*. (Oxford University Press, 2010). doi:10.1093/acprof:oso/9780195377033.001.0001.
17. Smits, N., Cuijpers, P. & van Straten, A. Applying computerized adaptive testing to the CES-D scale: a simulation study. *Psychiatry Res* **188**, 147–155 (2011).
18. Rolstad, S., Adler, J. & Rydén, A. Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health* **14**, 1101–1108 (2011).
19. Aiyegbusi, O. L. *et al.* Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nat Commun* **13**, 6026 (2022).
20. Turner, R. R. *et al.* Patient-reported outcomes: instrument development and selection issues. *Value Health* **10 Suppl 2**, S86-93 (2007).

21. Petersen, M. A. *et al.* Development of computerised adaptive testing (CAT) for the EORTC QLQ-C30 dimensions - general approach and initial results for physical functioning. *Eur J Cancer* **46**, 1352–1358 (2010).
22. Gibbons, R. D. *et al.* Development of the CAT-ANX: a computerized adaptive test for anxiety. *Am J Psychiatry* **171**, 187–194 (2014).
23. Mao, F., Sun, Y., Li, Y., Cui, N. & Cao, F. Internet-delivered mindfulness-based interventions for mental health outcomes among perinatal women: A systematic review. *Asian J Psychiatr* **80**, 103321 (2023).

## ACKNOWLEDGEMENTS

This research is supported by A*STAR Brain-Body Initiative (BBI) (#21718) and the FY21 Prenatal / Early Childhood Grant: H22P0M0005. The GUSTO study is supported by the National Research Foundation (NRF) under the Open Fund-Large Collaborative Grant (OF-LCG; MOH-000504) administered by the Singapore Ministry of Health's National Medical Research Council (NMRC) and the Agency for Science, Technology and Research (A*STAR). In RIE2025, GUSTO is supported by funding from the NRF's Human Health and Potential (HHP) Domain, under the Human Potential Programme.

## ADDITIONAL INFORMATION

**Competing interests**

The author(s) delcare no competing interests.