# Personalised Profiling in Mental Health:

# A CAT-based Approach for Maternal Well-being and Mood Disorders


Geoffrey Chern-Yee Tan
*Department of Mood and Anxiety*
*Institute of Mental Health*
Singapore, Singapore
Corresponding:
geoffrey_tan@imh.com.sg

Hong Ming Tan
*Institute of Operations Research and Analytics*
*National University of Singapore*
Singapore, Singapore
thm@nus.edu.sg

Nikita Rane
*Department of Mood and Anxiety*
*Institute of Mental Health*
Singapore, Singapore
Nikita_Rane@imh.com.sg

Ethel Siew Ee Tan
*Department of Mood and Anxiety*
*Institute of Mental Health*
Singapore, Singapore
Ethel_SE_TAN@imh.com.sg

Nawal Hashim
*Department of Psychology*
*Institute of Mental Health*
Singapore, Singapore
Nawal_MOHAMED_HASHI M@imh.com.sg

Chloe Wei En Teo
*Lee Kong Chian School of Medicine*
*Nanyang Technological University*
Singapore, Singapore
chlo0017@e.ntu.edu.sg

Kah Vui Fong
*Faculty of Science*
*National University of Singapore*
Singapore, Singapore
e0773790@u.nus.edu

Malorie Rui Yi Yoong
*Faculty of Arts and Social Sciences*
*National University of Singapore*
Singapore, Singapore
e0774843@u.nus.edu

Rishabh Garg
*Faculty of Science*
*National University of Singapore*
Singapore, Singapore
e0540222@u.nus.edu

Shayaan Sultan
*School of Computing*
*National University of Singapore*
Singapore, Singapore
sultan.shayaan@u.nus.edu

Jussi Keppo
*Business School*
*National University of Singapore*
Singapore, Singapore
keppo@u.nus.edu

Sharon Lu Huixian
*Department of Psychology*
*Institute of Mental Health* Singapore, Singapore
sharon_hx_lu@imh.com.sg



***Abstract*—Mood disorders are highly heterogeneous conditions that are a leading cause of disability around the world. This is particularly important in perinatal mental health, where there is an increased incidence during the perinatal period and significant impact on child outcomes. Psychological interventions such as internet-based Cognitive Behavioural Therapy (iCBT) have the potential to treat depression and anxiety and address underlying vulnerabilities, however it is limited in its ability to address individual vulnerabilities. We describe a framework where computerized adaptive testing (CAT), that has conventionally been applied in education, can be used to efficiently profile individual vulnerabilities. These responses as well as information from medical records and cognitive task information are incorporated into a recommender system for selecting iCBT modules that would be most likely to address individual vulnerabilities.***



***Keywords*— Healthcare, Personalised medicine, Mental health, Depression, Anxiety, Maternal health, Recommender systems, Computerised Adaptive Testing Introduction**

This research was supported by Translational Clinical Research (TCR) Flagship Program on Developmental Pathways to Metabolic Disease funded by the National Research Foundation (NRF) and administered by the National Medical Research Council (NMRC), Singapore- NMRC/TCR/004-NUS/2008 and A*STAR Brain–Body Initiative (BBI) (#21718), IMH Research Seed Funding (642–2018), LKCMed-NUSMed-NHG Collaborative Mental Health Research Pilot Grant Call 2020 (MHRPG/2003) FY21 PRENATAL / EARLY CHILDHOOD GRANT CALL (H22P0M0005). Additional funding is provided by the Singapore Institute for Clinical Sciences – A*STAR. K.M.G. is supported by the National Institute for Health Research through the NIHR Southampton Biomedical Research Centre.

## I. Introduction

Mood disorders are a leading cause of disability globally. This is particularly important during the perinatal period, where the incidence of postpartum depression is estimated at about 12% and overall prevalence of depression at about 17% among healthy mothers [1]. The antenatal period, preceding birth, presents a window of opportunity for early intervention to mitigate vulnerability to mental health issues and impact maternal well-being and subsequent child outcomes. Interventions such as internet-based Cognitive Behavioural Therapy (iCBT) may provide a means to address this issue at the population level and have been demonstrated to be as effective as face-to-face CBT in the Singapore context [2].

However, mood disorders such as postnatal depression are heterogenous in presentation, causes and response to treatment. For example, a study in women with postnatal depression (PND) identified that heterogeneity in symptoms could account for differences in response to CBT [3]. This variability necessitates the development of personalised interventions that are tailored to individual vulnerability factors. This is a process that tends to occur organically between therapist and client over the course of assessment, where the therapist attempts to formulate the various vulnerabilities and psychological factors that can be targeted in psychotherapy and to work collaboratively to identify appropriate interventions. In the context of self-help or internet-delivered therapies or where the service-provider or caregiver is not proficient in formulation and the delivery of such therapies, this is not usually possible. We seek in this paper to present a general framework for approaches to personalisation of psychological interventions in mental health using our work with iCBT for perinatal health as an example and how statistical and AI approaches can support this.

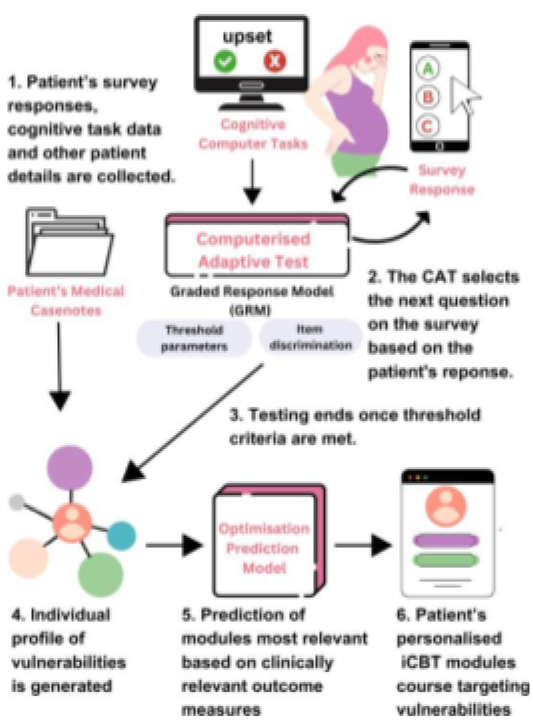


Fig. 1. Framework for personalisation in psychological intervention

## II. Framework for personalisation in psychological intervention

We propose a framework that incorporates the factors and decisions in Fig. 1, that are involved in a psychotherapy assessment including the vulnerability and psychological factors related to mental health and the factors related to engagement and response to specific interventions and psychotherapy in general. This information could be provided from various sources including the clinician, family, medical records, cognitive tasks as well as, self-report. A framework involving Computerised Adaptive Testing (CAT) will allow efficient collection of the information needed to optimise clinical decisions and choice of interventions. These factor estimates are then incorporated into interpretable predictive models to identify which modules could treat and prevent maternal mental health issues. We have used expert mapping of individual items into modules and specific interventions expected to help with those items as initial estimates and as further longitudinal outcome data is collected they will be used to further tune the models. While our current efforts make use of item response theory as the primary statistical framework to efficiently estimate the impact of individual factors, future work could integrate data collection with prediction and treatment planning and integration of large language models. For instance, training LLMs with CAT allows it to be compared with humans [4]. This allows LLMs to modify the characteristics of the test questions which would allow it to intelligently select factors for more precise data collection and to incorporate client inputs into the decision-making process.

## III. Datasets

Through analysis of the literature and longitudinal cohorts of pregnant mothers and their children including the Growing Up in Singapore Towards Healthy Outcomes study (GUSTO) [5], we identified vulnerability factors associated with perinatal mental health. We further identified factors associated with response and outcome from psychotherapy and specifically iCBT from the Understanding the person, exploring change across psychotherapies study (Xchange) (Ethics application:2018/01184) recruiting patients at the Institute of Mental Health referred to or undergoing psychotherapy.

## IV. CAT

CAT originated from the area of education to efficiently select test items according to an individual's proficiency. The theoretical basis of CAT lies in Item Response Theory (IRT) [6], in which items that are most informative about the individual's abilities are selected in order to efficiently and accurately measure their ability level [7]. In a similar vein, the Graded Response Model (GRM), a specific type of IRT, models the relationship between an individual's latent trait and their responses to polytomous items. While CAT is conventionally applied as an academic assessment, it holds the potential for broader applications, including profiling individual vulnerabilities for depression as well as task data.

TABLE I. TABLE TYPE STYLE ITEM REDUCTION CHARACTERISTICS OF THE MEQ CAT, THEIR CORRELATION WITH FIXED-LENGTH FORM SCORES AND PRECISION OF MODEL

| **Scale** | **No. of items in original test** | **Stopping rule** | | **No. of items used** | | ***r*** | **RMSE** |
|---|---|---|---|---|---|---|---|
| | | ***SE*** | ***Min/ Max questions*** | ***Mean*** | ***SD*** | | |
| MEQ | 18 | <0.4 | 8/16 | 14.79 | 1.57 | 0.997 | 0.0669 |

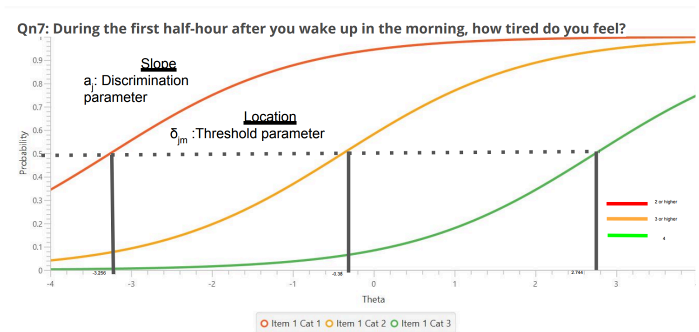


Fig. 2. Operating characteristic curves for 4 category example item 7 from MEQ scale under the Graded response model.

The simulation run on a CAT test for the MEQ using the Firestar 'R' package illustrates the utility of CAT in efficiently and accurately estimating latent traits. The MEQ originally comprises 18 questions, but administering the test with CAT reduced this number to a mean of 14. While fewer questions were administered, there was little compromise on the accuracy of the estimates, with a Pearson Correlation coefficient of 0.997 as well as a root mean squared error (RMSE) of 0.0669. With CAT, the MEQ could hence be administered with fewer questions while still maintaining a high level of accuracy.

## A. *Application of Graded Response Model in CAT*

In order to develop a CAT test, GRM was chosen to accurately assess the individual's vulnerability factors for depression through their responses to the polytomous psychometric items. To accomplish this, a dataset comprising existing responses to the aforementioned items by participants was first collated. This data was then passed into an objective function that calculates the log likelihood for each set of parameters, based on the GRM. This curve ( $_{ix}$) represents the probability of an individual's response ($x$) being correct when falling above or below a given category threshold ($j = 1...\ _i$), conditional on trait level/ability level (θ).

$$_{ix}\ \frac{_{i}\ _{ij}}{1+exp[\ _{i}(\theta-\ _{ij})]},$$

where $x = j = 1,...,\ _i$

Fig. 3. The equation for GRM [8]

Within the GRM, each item would be fit with a discrimination parameter, which would be similar to the slope of probability against theta and extremity parameters describing the boundaries or thresholds where there is a 50% chance of choosing between the categories on either side.

In addition, the data was also passed into a function generating an initial guess for the optimisation, as well as a constraint function, which would ensure that the beta values would be monotonically increasing. These parameters and constraints were then passed into the SciPy optimiser, and using its minimise function, the optimal parameters that maximise the log likelihood were calculated.

## B. *Application of GRM for direct observation of latent variables*

In traditional GRM applications, experts often use Marginal Maximum Likelihood Estimation (MMLE) to handle the latent variable—a hidden factor that affects how individuals respond to questions. This traditional method involves making educated guesses about this unseen factor and then statistically inferring the parameters of the model based on the responses [9], [10].

However, our approach deviates from this norm. In our project, we directly observe the latent variable in our dataset, eliminating the need to estimate it indirectly. This direct observation is a significant advantage, allowing for a more straightforward application of Maximum Likelihood Estimation (MLE) [11]. In our context, MLE works by closely examining the relationship between the observed latent variable and the responses from individuals. Here's a step-by-step overview of how MLE is applied in our research:

- Data Collection: We gather responses from individuals for a series of questions, alongside directly observing the latent variable for each individual.
- Parameter Specification: We aim to determine two main parameters:
- Discrimination Parameters: These indicate how well each question differentiates between individuals with different levels of the latent trait.
- Extremity Parameters: These define the threshold levels at which the likelihood of choosing certain answer categories changes significantly. We estimated extremity and discrimination parameters for each item across all questionnaires identified to be vulnerability factors for mental health in the GUSTO cohort, either through previous associations or as features in prediction models, such as shown for the BFI in Table 2.
- Likelihood Calculation: For each individual, we calculate the likelihood of their observed responses, given the observed level of the latent trait. This involves considering how the responses and the latent trait interact, based on the current estimates of the discrimination and extremity parameters.
- Optimization: We optimise the parameters (discrimination and extremity) to maximise the likelihood of the observed set of responses across all individuals.

## C. *Simulation of CAT using Firestar*

After developing the CAT test, we used Firestar to simulate the CAT, using different stopping criteria to minimise the number of questions used without compromising accuracy. These were the hyper-parameters specified when running the Firestar package, with the reason for a selected hyper-parameter in brackets [12].

- pop.dist: Population distribution type for simulated theta; "NORMAL" (normal distribution)

- max.cat: Maximum number of response categories across items; 5 (based on questionnaire)
- min.theta, max.theta: Minimum and maximum theta values; -4, 4 (default)
- min.NI, max.NI: Minimum and maximum number of items to administer; 4, 10 (default)
- exposure.control: Whether to invoke exposure control; FALSE (default)
- selection.method: Method for item selection; Maximum Posterior Weighted Information, “MPWI”
- prior.dist: Type of prior distribution; 1 (normal distribution)
- prior.mean, prior.sd: Prior distribution mean and standard deviation; 0, 1 (normal distribution)

TABLE II. EXTREMITY AND DISCRIMINATION PARAMETERS OF ITEMS 1- 3 FOR BIG FIVE INVENTORY (BFI)

| BFI Item | Extremity parameter | | | | Discrimination parameter |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| Item 1 | -6.139 | -2.509 | 0.187 | 5.679 | 0.343 |
| Item 2 | 1.94 | -0.906 | -4.009 | -10.847 | -0.386 |
| Item 3 | -4.314 | -2.386 | -1.248 | 0.893 | 0.939 |

By directly incorporating the observed latent variable into the MLE process, our research brings clarity and precision to the parameter estimation, potentially increasing the robustness and interpretability of the GRM analysis. This direct approach, diverging from traditional MMLE methods, leverages our unique dataset to deepen the understanding of how individual responses relate to the observed latent trait.

## V. VULNERABILITY FACTOR MAPPING

We have performed an initial expert-based mapping of items to interventions from a consensus of therapists before training predictive models on longitudinal outcome data. Table 3 shows 14 identified vulnerability factors for maternal depression and anxiety. These factors are addressed through the specific iCBT modules, ranging from 1 to 7, which teach users various self- help interventions.

TABLE III. EXAMPLE OF THE MAPPING OF VULNERABILITY FACTORS TO ICBT MODULES

| Maternal Self Help Intervention Module | Vulnerability Factor |
|---|---|
| Module 1: Introduction to iCBT | Sleep |
| Module 2: Values and goals | Parenting/experience with baby, self esteem, attachment (including past trauma), income/socioeconomic status |
| Module 3: Changing emotion-driven behaviours | Personality, attachment (including past trauma) |
| Module 4: Flexible thinking | Depression/anxiety, negative cognitions, personality (neuroticism), self esteem |
| Module 5: Behavioural experiments | Depression/anxiety, negative cognitions |
| Module 6: Mindful emotional awareness | General wellbeing, health/quality of life, stress |
| Module 7: Relationships and communication | Marital relationship/social support/family stress |

### CONTEXT-AWARE RECOMMENDER SYSTEM

All participants will complete module 1 where they will be provided with information about each module as well as given an overview of the various interventions. They will then be asked to rate the relevance or usefulness of each module as an initial estimate of the rating.

The recommendation problem is framed as a prediction problem in which, given a user profile and a target item, the recommender system attempts to predict that user’s rating or preference for that item estimating the rating function [13], [14]:

R : Users x Items ⇒ Ratings

We would add additional information or dimensions from the various vulnerability and therapy-related factors as contextual information:

R : Users x Items x Contexts ⇒ Ratings

We use a contextual modelling approach where the contextual features are defined as the factors from the CAT as well as additional data from EMR and service-provider inputs alongside the user x item data obtained from module 1.

## VI. IMPACT AND CONCLUSIONS

We present a framework for personalised profiling in the delivery of iCBT modules. We are implementing this for perinatal mental health across Women’s Mental Health Services across two hospitals at the National University Hospital Singapore and Institute of Mental Health and as self-help for healthy populations. This can be expected to improve client engagement by reducing subject burden and improving motivation and early response. The use of AI to tailor interventions may substantially reduce the workload and burden on experienced therapists while closing the treatment gap for mental health. By incorporating domain content as well as domain expertise into a self-help process, we empower patients and the support they have available to access personalised psychotherapy. We expect that this will be eventually utilised across health services, social services and voluntary organisations by a range of service providers and caregivers as well as self-help by mothers in the community.

## Acknowledgment

This research was supported by Translational Clinical Research (TCR) Flagship Program on Developmental Pathways to Metabolic Disease funded by the National Research Foundation (NRF) and administered by the National Medical Research Council (NMRC), Singapore-NMRC/TCR/004- NUS/2008 and A*STAR Brain–Body Initiative (BBI) (#21718), IMH Research Seed Funding (642–2018), LKCMed-NUSMed- NHG Collaborative Mental Health Research Pilot Grant Call 2020 (MHRPG/2003) FY21 PRENATAL / EARLY CHILDHOOD GRANT CALL (H22P0M0005). Additional

funding is provided by the Singapore Institute for Clinical Sciences – A*STAR. K.M.G. is supported by the National Institute for Health Research through the NIHR Southampton Biomedical Research Centre.

## References


[1] S. Shorey, C. Y. I. Chee, E. D. Ng, Y. H. Chan, W. W. San Tam, and Y. S. Chong, "Prevalence and incidence of postpartum depression among healthy mothers: A systematic review and meta-analysis," Journal of Psychiatric Research, vol. 104, pp. 235-248, 2018.

[2] S. H. X. Lu et al., "A randomised controlled trial of clinician-guided internet-based cognitive behavioural therapy for depressed patients in Singapore," Frontiers in Psychology, vol. 12, 668384, 2021.

[3] A. Waqas and A. Rahman, "Does one treatment fit all? Effectiveness of a multicomponent cognitive behavioral therapy program in data-driven subtypes of perinatal depression," Frontiers in Psychiatry, vol. 12, p. 736790, 2021.

[4] Zhuang, Yan, Qi Liu, Yuting Ning, Weizhe Huang, Rui Lv, Zhenya Huang, Guanhao Zhao, et al. "Efficiently Measuring the Cognitive Ability of LLMs: An Adaptive Testing Perspective." arXiv (Cornell University), January 1, 2023. https://doi.org/10.48550/arxiv.2306.10512.

[5] S.-E. Soh et al., "Cohort profile: Growing up in Singapore towards Healthy Outcomes (GUSTO) birth cohort study," International Journal of Epidemiology, vol. 43, no. 5, pp. 1401–1409, Aug. 2013. doi:10.1093/ije/dyt125

[6] M. O. Edelen and B. B. Reeve, "Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement," Quality of Life Research, vol. 16, no. S1, pp. 5–18, Mar. 2007. doi:10.1007/s11136-007-9198-0

[7] G. G. Kingsbury and A. R. Zara, "Procedures for selecting items for computerized adaptive tests," Applied Measurement in Education, vol. 2, no. 4, pp. 359–375, Oct. 1989. doi:10.1207/s15324818ame0204_6

[8] S. E. Embretson and S. P. Reise, Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum, 2000.

[9] M. Young, The Technical Writer's Handbook. Mill Valley: University Science Books, 1989. .

[10] F. Samejima, "Estimation of latent ability using a response pattern of graded scores," Psychometrika, vol. 34, no. S1, pp. 1–97, Mar. 1969. doi:10.1007/bf03372160

[11] W. J. van der Linden and R. K. Hambleton, Eds., "Handbook of Modern Item Response Theory," Springer Science & Business Media, 2013.

[12] G. Adomavicius and A. Tuzhilin, "Context-aware Recommender Systems," Recommender Systems Handbook, pp. 217–253, Oct. 2010. doi:10.1007/978-0-387-85820-3_7

[13] Seung W., Choi. "Firestar: Firestar - Computerized Adaptive Testing (CAT) Simulation... in Choi-PhD/Firestar: Computerized Adaptive Testing (CAT) Simulation Program." R Package Documentation, September 14, 2022. https://rdrr.io/github/choi-phd/Firestar/man/Firestar.html.

[14] A. Ansari, S. Essegaier, and R. Kohli, "Internet recommendation systems," Journal of Marketing Research, vol. 37, no. 3, pp. 363–375, Aug. 2000. doi:10.1509/jmkr.37.3.363.18779