.

PROTOCOL

Sleep Architecture as a Digital Biomarker for Postpartum Depression: Neuro-structural Validation and Cross-Cultural Transferability using UK Biobank Longitudinal Data and a Local Hong Kong Cohort

Principal investigator:

Dr Ying Dai
Research Assistant Professor
School of Nursing
The University of Hong Kong

Co-investigator:

Prof. Kris Lok
Associate Professor
School of Nursing
The University of Hong Kong

Dr Amelia Hui
Consultant
Department of Obstetrics and Gynaecology
Queen Mary Hospital
The University of Hong Kong

Dr. Hye Ri Choi
Research Assistant Professor
School of Nursing
The University of Hong Kong

Dr. Jie Zhong
Research Assistant Professor
School of Nursing
The University of Hong Kong

Dr. Xiangyu Zhai
Research Assistant Professor
School of Nursing
The University of Hong Kong

.

TABLE OF CONTENTS

.

I. SIGNIFICANCE

Postpartum Depression (PND) constitutes a severe public health crisis that affects approximately 10% to 15% of mothers worldwide [1], with prevalence rates frequently reported as higher in dense and high-stress urban environments such as Hong Kong [2]. The condition is associated with profound negative outcomes including impaired mother-infant bonding [3], developmental delays in the child [4], and in severe cases maternal suicide [5]. Despite the availability of effective treatments PND remains significantly underdiagnosed and undertreated. The current standard of care for screening relies almost exclusively on subjective self-report questionnaires such as the Edinburgh Postnatal Depression Scale (EPDS) [6]. While the EPDS is a validated instrument it possesses inherent limitations that hinder early and accurate detection. The scale is burden intensive as it requires active participation from exhausted new mothers and is subject to recall bias. Furthermore, there is a persistent stigma in many Asian cultures surrounding the reporting of mental health symptoms which leads many mothers to under-report their psychological distress on questionnaires [7]. Consequently, there is an urgent and unmet need for objective and passive biomarkers that can identify risk without relying solely on maternal self-disclosure.

The central scientific problem this project addresses is the lack of a validated "digital phenotype" for PND. It is well established that sleep disturbance is a core symptom of depression [8]. However, in the context of early motherhood generalized complaints of poor sleep are often dismissed as a normal consequence of infant care. This lack of specificity renders simple sleep duration a poor diagnostic tool. Previous small-scale clinical studies using polysomnography suggest that mothers with PND exhibit distinct and subtle disruptions in circadian rhythmicity such as specific fragmentation patterns or flattened circadian amplitude which are undetectable by self-reports [9, 10]. We operationally define these persistent accelerometry-derived patterns as physiological scars. To identify these minute physiological fingerprints, researchers require massive datasets to ensure statistical power. However previous large-scale studies have been limited by their cross-sectional nature making it impossible to determine if sleep disturbances are a transient state or a stable trait. Furthermore, the biological mechanism linking these sleep scars to long-term depressive outcomes remains obscure. This project seeks to bridge these gaps by capitalizing on the newly released UK Biobank Data Release v20 which provides unprecedented longitudinal accelerometry and brain imaging data combined with the clinical relevance of a local Hong Kong cohort [11].

Project Objectives

The primary objective of this project is to develop and validate a PND Risk Index which is an objective screening algorithm based on sleep accelerometry data that can be deployed via consumer wearables. The research is structured around three specific objectives updated to leverage the latest longitudinal data capabilities. The first objective is to identify a stable physiological scar of PND using retrospective discovery. By analyzing raw accelerometry data from female participants in the UK Biobank we will determine whether a history of PND is associated with distinct and long-term alterations

in sleep microstructure. Uniquely we will utilize the newly released Seasonal Repeat Accelerometry dataset to test the temporal stability of these biomarkers and hypothesize that the PND Sleep Signature such as high Intra-daily Variability remains consistent across different seasons and years confirming it is a trait marker rather than a transient environmental response.

The second objective is to investigate the neurobiological consequences of this sleep scar by using the new Longitudinal Brain Imaging phenotypes. We will examine whether the specific sleep architectural defects identified in the first objective are associated with accelerated neuro-structural changes over time specifically focusing on cortical thinning and hippocampal volume reduction in the longitudinal MRI subset. Previous literature indicates these regions are highly sensitive to cortisol dysregulation in pregnancy [12].

The third objective is to conduct a prospective validation through a pilot study with 60 postpartum mothers in Hong Kong. We will collect objective sleep data via wearables to determine if the stable and neuro-biologically validated sleep signatures identified in the UK Biobank are detectable during the acute phase of PND in a Chinese population. This translational step ensures that our Big Data findings are applicable to the specific cultural and environmental context of Hong Kong.

Significance and Innovation

This proposal offers significant clinical innovation by shifting the paradigm of maternal mental health screening from subjective questionnaires to passive Zero-Burden Screening. If successful, this research creates a pathway for clinical tools where wearable technology could flag PND risk automatically and facilitate earlier intervention. This is particularly significant for Hong Kong where the healthcare system is overburdened, and passive monitoring could triage high-risk mothers more effectively. Additionally, we address the unique challenge of the local sleeping environment. Unlike Western populations where solitary sleeping is common Hong Kong is characterized by high-density living where co-sleeping with infants is prevalent. This project will develop specific algorithms to distinguish maternal sleep from infant movement artifacts [13]. Methodologically this study is among the first to capitalize on the UK Biobank 2025 Data Release v20. By utilizing the Seasonal Repeat Accelerometry data we address the critical limitation of previous snapshot studies and offer the first rigorous assessment of the temporal stability of digital depression biomarkers. Furthermore, the integration of longitudinal neuro-imaging data allows us to move beyond simple correlation to explore mechanisms by testing the hypothesis that sleep fragmentation acts as a mediator for neuro-structural degradation in mothers with a history of depression.

II. AIMS

.

Primary Aim:

The primary objective of this project is to develop and validate a PND Risk Index which is an objective screening algorithm based on sleep accelerometry data that can be deployed via consumer wearables. The research is structured around three specific objectives updated to leverage the latest longitudinal data capabilities. The first objective is to identify a stable physiological sleep scar of PND using retrospective discovery. By analyzing raw accelerometry data from female participants in the UK Biobank, we will determine whether a history of PND is associated with distinct and long-term alterations in sleep microstructure. Uniquely, we will use the newly released Seasonal Repeat Accelerometry dataset to test the temporal stability of these biomarkers and hypothesize that the PND Sleep Signature such as high Intra-daily Variability remains consistent across different seasons and years, aiming to confirm it is a trait marker rather than a transient environmental response.

Secondary Aims:

The second objective is to investigate the neurobiological consequences of this sleep scar by using the new Longitudinal Brain Imaging phenotypes. We will examine whether the specific sleep architectural defects identified in the first objective are associated with accelerated neuro-structural changes over time specifically focusing on cortical thinning and hippocampal volume reduction in the longitudinal MRI subset. Previous literature indicates these regions are highly sensitive to cortisol dysregulation in pregnancy.

The third objective is to conduct a prospective validation through a pilot study with 60 postpartum mothers in Hong Kong. We will collect objective sleep data via wearables to determine if the stable and neuro-biologically validated sleep signatures identified in the UK Biobank are detectable during the acute phase of PND in a Chinese population. This translational step ensures that our Big Data findings are applicable to the specific cultural and environmental context of Hong Kong.

Hypothesis:

1. PND is associated with a stable and measurable disruption in sleep architecture.
2. Chronic sleep fragmentation accelerates structural changes in brain regions responsible for emotion and memory, specifically causing volume loss in the hippocampus.
3. These sleep alterations can be accurately detected in Hong Kong mothers by using specialized algorithms that account for local co-sleeping practices.

III. STUDY DESIGN

.

This study This research proposes a sequential study design that integrates high-dimensional Big Data discovery with Thick Data clinical validation. The scientific premise of this approach rests on addressing the Ecological Fallacy often present in large-scale epidemiological studies. In this context relying solely on UK Biobank data to design a screening tool for Hong Kong mothers would be fallacious as it assumes that sleep-depression associations derived from a predominantly Caucasian and middle-aged population sleeping in solitary environments apply perfectly to young Chinese mothers in high-density co-sleeping environments. Conversely a standalone clinical study in Hong Kong would lack the sample size required to isolate specific digital biomarkers from the noise of daily life. Therefore, we adopt a translational design comprised of four distinct but complementary analytical components. These components are a Retrospective Discovery Analysis using the UK Biobank to identify stable sleep scars, a Mechanistic Validation utilizing longitudinal neuro-imaging to establish biological plausibility, a Prospective Validation Pilot in Hong Kong to test the transferability of these markers, and a final Algorithm Development phase. This triangulation of methods ensures that the resulting PND Risk Index is not only statistically robust but also biologically plausible and culturally adaptable.

Phase I. Retrospective Discovery and Stability Analysis (UK Biobank)

Data Source Characteristics and Cohort Selection We will conduct a secondary analysis of the UK Biobank dataset which serves as the gold standard for large-scale biomedical research. The UK Biobank recruited over 500,000 participants aged 40 to 69 years between 2006 and 2010 from across the United Kingdom. For this specific study we will use the sub-cohort of participants who participated in the accelerometer study. Between 2013 and 2015 approximately 103,000 participants were sent a wrist-worn Axivity AX3 triaxial accelerometer to wear for seven consecutive days. This device captures continuous high-resolution movement data at a sample rate of 100 Hz. It is important to note that this is actigraphy data serving as a proxy for sleep-wake patterns and not Polysomnography (PSG). We will utilize two distinct accelerometry datasets from the v20 data release. The primary dataset Category 1001 serves as the baseline measurement. The secondary dataset is the Seasonal Repeat Accelerometry study Category 1008 which was collected between 2018 and 2019 where a subset of participants repeated the monitoring protocol.

Rigorous Definition of the Exposure Variable The primary exposure of interest is a history of Postpartum Depression. Defining this phenotype in a retrospective cohort requires a sophisticated algorithmic approach to minimize misclassification bias. We will synthesize data from multiple sources within the UK Biobank repository. The

primary source will be the Mental Health Questionnaire Category 136 specifically Field 20447 Lifetime number of depressed periods and Field 20544 Mental health problems ever diagnosed by a professional to identify women who explicitly reported a diagnosis of postnatal depression. To enhance the specificity of this classification we will cross-reference self-reports with linked hospital inpatient records using International Classification of Diseases ICD-10 codes specifically searching for code F53 Mental and behavioural disorders associated with the puerperium. We will also utilize data from the reproductive health questionnaire to verify the timing of deliveries relative to the reported depressive episodes. To ensure phenotype purity participants with conflicting data points among these three resources will be excluded. Women who report depression but have nulliparous status meaning they have never given birth will serve as a negative control group.

Construction of the Control Group using Propensity Score Matching A major challenge in retrospective analysis is confounding by indication and lifestyle factors. To address this we will construct a robust control group using Propensity Score Matching. We will estimate the propensity score for each participant using a logistic regression model that predicts the probability of having a history of PND based on a comprehensive set of covariates. These covariates will include age at recruitment, Townsend Deprivation Index, Body Mass Index, smoking status, alcohol intake frequency, menopausal status, and family history of depression. We will then match each PND case one-to-one with a control participant using the nearest neighbor matching algorithm with a caliper width of 0.2 standard deviations of the logit of the propensity score. This rigorous matching process ensures that the distribution of baseline characteristics is balanced between the case and control groups thereby isolating the independent effect of PND history on sleep architecture.

Accelerometry Signal Processing and Feature Extraction Pipeline The raw accelerometry data is stored in the compressed.cwa format. Processing this massive dataset requires a high-performance computing approach. We will utilize the UK Biobank Research Analysis Platform cloud infrastructure to execute our analysis pipeline which will be built using the R programming language. The core of our signal processing will rely on the GGIR package which is the industry-standard open-source tool for processing raw accelerometer data. The pipeline involves several critical steps. First the algorithm performs auto-calibration of the signal to local gravity to correct for sensor gain and offset errors. Second it calculates the Euclidean Norm Minus One which represents the magnitude of acceleration corrected for gravity expressed in milli-gravitational units. Negative values are rounded to zero. Third the algorithm detects non-wear time based on the standard deviation and value range of each axis over 15-minute windows.

.

Derivation of Novel Sleep Micro-structure Metrics We will calculate standard sleep metrics such as Total Sleep Time and Sleep Efficiency to allow for comparison with existing literature. However, we will focus primarily on extracting Non-Standard Circadian Metrics that describe the structural integrity of the rest-activity rhythm as standard metrics are often insufficiently sensitive to detect the subtle fragmentation seen in depression. We will calculate Intra-daily Variability (IV) which is a nonparametric metric that quantifies the fragmentation of activity rhythms. IV is calculated as the ratio of the mean square of the difference between consecutive activity hours to the mean square of the difference from the global mean. A higher IV indicates that the rest-activity cycle is broken up into smaller fragments serving as a proxy for sleep maintenance insomnia and circadian disorganization. We will also calculate Inter-daily Stability (IS) which measures the coupling strength of the sleep-wake rhythm to the 24-hour zeitgeber. Lower IS values indicate a lack of synchronization often seen in mood disorders. Additionally, we will extract the L5 onset time which marks the start time of the least active five hours of the day to assess phase delays or advances.

Statistical Modeling of Stability and Seasonality To determine the stability of these markers we will employ a Linear Mixed Model framework. The LMM allows us to handle the hierarchical structure of the data and to explicitly model the variance components. The model will include the sleep metric as the dependent variable with PND history, season of measurement, and time interval between measurements as fixed effects and participant ID as a random effect. We will specifically test the interaction term between PND history and Season to determine if the sleep scar is exacerbated by seasonal changes. Furthermore, we will calculate the Intraclass Correlation Coefficient for the key sleep metrics across the baseline and repeat measurements specifically within the PND group. An ICC threshold of greater than 0.6 will be used to define a metric as a stable trait suitable for use as a long-term biomarker in the PND Risk Index.

Phase II. Neuro-structural Mechanistic Validation (UK Biobank)

Rationale for Neuro-imaging Integration Establishing a statistical association between PND history and sleep fragmentation is necessary but not sufficient to claim causality or clinical significance. To strengthen the biological plausibility of our findings we will interrogate the newly released Longitudinal FreeSurfer phenotypes Category 530-537 from the UK Biobank imaging substudy. This dataset provides derived volumes and cortical thickness measures for a subset of approximately 4,500 participants who underwent repeat Magnetic Resonance Imaging scans with an interval of two to three years. This offers a unique opportunity to test the Neuro-toxic Scar Hypothesis which posits that chronic sleep fragmentation accelerates

neurodegeneration in specific brain regions associated with emotion regulation and memory.

Region of Interest (ROI) Selection and Hypothesis We will focus our analysis on a predefined set of Regions of Interest (ROIs) based on the established neurobiology of depression. The primary ROI is the hippocampus, specifically the cornu ammonis (CA) subfields and the dentate gyrus, as these regions are highly sensitive to cortisol dysregulation and sleep loss. Our secondary ROIs include the amygdala, which is critical for emotional processing, and the anterior cingulate cortex (ACC), which is involved in error monitoring and mood regulation. We hypothesize that the severity of the sleep architectural disturbance (e.g., the magnitude of Intra-daily Variability) identified in the accelerometry analysis will be a significant predictor of the rate of atrophy (volume loss) in these specific structures over the longitudinal imaging interval.

Longitudinal Mediation Analysis Framework We will employ a longitudinal mediation analysis to test the causal pathway. We hypothesize a specific directionality where PND history acts as the antecedent variable which causes persistent sleep fragmentation which in turn accelerates neurodegeneration. The path diagram consists of PND history as the independent variable X, the rate of hippocampal volume change as the dependent variable Y, and the sleep fragmentation metric as the mediator M. We will only include sleep metrics that demonstrated significant stability in the previous analysis phase as candidate mediators. We will first calculate the annualized percent change in hippocampal volume for each participant. Then we will construct a series of regression equations. Path A will estimate the effect of PND history on sleep fragmentation. Path B will estimate the effect of sleep fragmentation on hippocampal atrophy rate adjusting for PND history. Path C prime will estimate the direct effect of PND history on atrophy. We will use bootstrapping with 5,000 iterations to estimate the indirect effect and its confidence intervals.

Phase III. Prospective Validation in a Local Hong Kong Cohort

Clinical Feasibility and Recruitment Strategy The transition from Big Data to the clinic involves testing the identified biomarkers in an acute setting. We will conduct a prospective observational pilot study targeting 60 postpartum mothers in Hong Kong. Recruitment will be executed through a multi-channel strategy to ensure a sample that has various demographic characteristics. We have established a collaboration with Department of Obstetrics and Gynecology at Queen Mary hospital, which serve a diverse socioeconomic demographic. Research nurses will approach mothers at the postnatal ward at Queen Mary hospital. Additionally, we will use digital recruitment via established Hong Kong parenting portals such as "Baby Kingdom" and university

.

mass mailings. The inclusion criteria include mothers must be at least 18 years of age, within 6 months postpartum, and able to read Traditional Chinese. We will exclude mothers with diagnosed sleep apnea, those using CPAP machines, and those working permanent night shifts, as these factors introduce variance unrelated to mood. We will also exclude mothers with severe medical complications (e.g., pre-eclampsia requiring ongoing medication) to reduce physiological confounding.

Data Collection Protocol and Instrumentation Upon enrollment participants will undergo a baseline assessment where they will complete a battery of validated questionnaires. This includes the Edinburgh Postnatal Depression Scale to quantify depressive symptoms, the Pittsburgh Sleep Quality Index for subjective sleep quality, and the State-Trait Anxiety Inventory. Following the baseline assessment participants will be fitted with a GENEActiv Original accelerometer on their non-dominant wrist. The GENEActiv is chosen for its superior comparability to the Axivity device used in the UK Biobank as both devices utilize similar MEMS sensor technology. This ensures that the signal processing pipeline developed for the UK Biobank data can be applied to the local Hong Kong data with minimal modification. Participants will be instructed to wear the device for 14 consecutive days 24 hours a day. Concurrent with the objective monitoring participants will complete a simplified electronic sleep diary each morning via a secure WhatsApp link. We note that we will not be collecting MRI data in this local cohort due to budget constraints and the specific focus of this grant on validating the wearable-based risk index.

Cultural Adaptation: Addressing the Co-Sleeping Confound A significant methodological innovation of this study is the adaptation of Western-derived sleep algorithms to the Asian cultural context. In the UK Biobank population, solitary sleeping or sleeping with a partner in a large bed is the norm. In contrast, Hong Kong is characterized by extremely high-density living, and the cultural practice of bed-sharing (co-sleeping) with the infant is highly prevalent. This proximity creates a technical challenge known as "motion artifacts," where the accelerometer on the mother's wrist detects the movements of the co-sleeping infant or partner, potentially leading the algorithm to misclassify sleep as wakefulness. To address this, we will implement a "Context-Aware" processing pipeline. We will stratify our analysis based on the self-reported sleeping arrangement (Solitary, Room-sharing, Bed-sharing). For the Bed-sharing group, we will develop and test an "Adaptive Thresholding" algorithm. Instead of using a fixed gravity-subtracted acceleration threshold (e.g., 40mg) to define wakefulness, we will test a dynamic threshold that adjusts based on the background noise level of the signal during confirmed sleep periods. We will also use the sleep diary data to manually flag intervals where the mother reports being asleep but the device registers high-frequency, low-amplitude

movements characteristic of infant motion. This calibration process is essential for ensuring that the PND Risk Index is valid for the local population.

Statistical Analysis Plan for the Local Cohort The analysis of the pilot data will focus on establishing Concurrent Validity and Diagnostic Accuracy. We will calculate the Pearson correlation coefficients between the objective sleep metrics (derived from the GENEActiv) and the subjective EPDS scores. We expect a moderate-to-strong positive correlation between the Sleep Fragmentation Index and the EPDS total score. To assess diagnostic accuracy, we will perform a Receiver Operating Characteristic (ROC) curve analysis. We will classify participants as "At Risk" if their EPDS score is 10 or higher, which is the standard clinical cutoff in Hong Kong. We will then calculate the Area Under the Receiver Operating Characteristic Curve (AUROC) for each sleep metric (IV, IS, SFI) to determine its ability to discriminate between high-risk and low-risk mothers. We aim for an AUROC of greater than 0.75, which would indicate good diagnostic utility. We will also calculate the Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) at various cut-points to identify the optimal threshold for screening purposes.

Phase IV. Algorithm Development and Machine Learning Integration

The final analytical phase involves the synthesis of the "Big Data" stability findings and the "Thick Data" cultural validations into the "PND Risk Index." We will employ a Supervised Machine Learning approach to build this predictive model. We will test three classes of models: Logistic Regression (as a baseline interpretable model), Random Forest (to capture non-linear interactions), and Gradient Boosting Machines (XGBoost). The input features will include the stable objective sleep metrics (IV, IS, L5 onset), derived circadian features, and key demographic risk factors (e.g., primiparity, history of depression). The target variable will be the PND risk status. We will train the models using Leave-One-Out Cross-Validation (LOOCV) given the small sample size of the pilot study. Feature selection will be performed using Recursive Feature Elimination (RFE) to identify the minimal set of variables required for accurate prediction, ensuring that the final algorithm is computationally efficient enough to run on consumer-grade wearable devices. The performance of the final model will be evaluated using the F1-score, which balances precision and recall, ensuring that we do not miss potential cases of depression (high recall) while avoiding excessive false alarms (high precision). This finalized algorithm will serve as the core intellectual property and the primary deliverable for the startup seed fund.

IV. SETTING

Recruitment will be executed through a multi-channel strategy to ensure a sample that has various demographic characteristics. A research assistant will approach mothers at the postnatal ward at Queen Mary hospital. Additionally, we will use digital recruitment via established Hong Kong parenting portals such as "Baby Kingdom" and university mass mailings. For those who agree to participate, the research assistant will contact the participant via phone for recruitment.

V. PARTICIPANTS

This study involves two distinct groups of participants. The first group consists of a retrospective cohort from the UK Biobank, where the researchers will select female participants who have provided accelerometry data. Within this large dataset, the team focuses on a specific subgroup of women with a documented history of PND, identified through mental health questionnaires, professional diagnoses, and hospital inpatient records. To ensure a clean comparison, the study also includes a control group of women who have never experienced this condition, matched for factors such as age, education level, body mass index, and lifestyle habits.

The second group comprises a local prospective cohort of 60 postpartum mothers recruited in Hong Kong. For this local clinical pilot, the inclusion and exclusion criteria are as follows:

The inclusion criteria

a) Mothers who are at least 18 years of age and are within 6 months of giving birth.
b) Participants must be capable of reading Traditional Chinese, Mandarin, or English, to ensure they can provide informed consent and complete the required study documentation.

The exclusion criteria

a) Individuals with specific conditions that might interfere with the measurement of sleep patterns related to mood will be excluded. This includes mothers diagnosed with sleep apnea or those who use CPAP machines.
b) To reduce physiological noise in the data, women working permanent night shifts or those with severe medical complications, such as pre-eclampsia requiring ongoing medication, will also be excluded.

Sample size: The total sample size is 1,060 women, including approximately 1,000 from the UK Biobank and 60 from Queen Mary Hospital. For women from the UK biobank dataset, it is estimated that 900 to 1000 women have a confirmed history of PND, valid 7-day sleep records, and structural brain MRI data. For the Hong Kong

cohort, power analysis showed that a total of 60 women reach the minimum sample size needed to detect an Area Under the Receiver Operating Characteristic Curve of 0.75 against a random baseline of 0.50, with a power of 80% and a 5% significance level to compute the standard error required to prove the results are not due to chance.

Participation incentives:
The incentive for study participation is a cash coupon valued at HKD 200 once they complete baseline questionnaire and wear actigraphy for 14 days to collect objective sleep data.

VI. Data collection

Research assistants will identify and contact eligible participants through Department of Obstetrics and Gynaecology at Queen Mary Hospital, Hong Kong parenting portals such as "Baby Kingdom" and HKU university mass mailings. Prior to beginning the study, potential participants will be presented with an Information Sheet on the first page of the printed questionnaires. This document will detail the purpose and scope of the SADBPD research, the expected time commitment for data collection, steps taken to ensure participant anonymity and data confidentiality, their right to withdraw from the study at any time without prejudice or negative consequences, the secure data storage protocols and retention period, and contact information for support and questions.

Upon enrollment participants will undergo a baseline assessment where they will complete a battery of validated questionnaires. This includes the Edinburgh Postnatal Depression Scale to quantify depressive symptoms, the Pittsburgh Sleep Quality Index for subjective sleep quality, and the State-Trait Anxiety Inventory. Following the baseline assessment participants will be fitted with a GENEActiv Original accelerometer on their non-dominant wrist. Participants will be instructed to wear the device for 14 consecutive days 24 hours a day. Participants will complete a simplified electronic sleep diary each morning via a secure WhatsApp link. We will not collect MRI data in Hong Kong cohort due to budget constraints and the specific focus of this grant on validating the wearable-based risk index.

Data Protection and Storage
All collected data will be stored securely for 5 years following study completion, in accordance with institutional research data management policies. Access to the Qualtrics survey responses will be strictly limited to the principal investigator, co-

investigators and research assistants. Data will be stored in password-protected files on encrypted servers, and all identifying information will be removed during data analysis and reporting. After the 5-year retention period, all data will be permanently deleted by overriding each file.

VII. Measurement

a) The Edinburgh Postnatal Depression Scale (EPDS) will be used to quantify depressive symptoms.
b) The Pittsburgh Sleep Quality Index (PSQI) will be used to measure subjective sleep quality.
c) The State-Trait Anxiety Inventory (STAI) will be used to measure anxiety trait and state.
d) The GENEActiv Original accelerometer will be weared on the participant's non-dominant wrist to measure objective sleep health.

VIII. Data analysis

We will proceed with the data analysis in distinct phases using both high-performance cloud computing and clinical statistical validation. For the retrospective UK Biobank analysis, we will process raw accelerometry data using the R programming language and the GGIR package to calibrate signals and remove non-wear time. We will calculate specific circadian metrics, including Intra-daily Variability and Inter-daily Stability, to quantify the fragmentation and synchronization of sleep rhythms. To determine if these markers represent stable traits rather than transient states, we will employ Linear Mixed Models to test for seasonal interactions and calculate Intraclass Correlation Coefficients across repeat measurements.

For the neuro-structural validation, we will use a longitudinal mediation analysis framework to investigate the biological mechanism linking sleep and depression. We will construct a series of regression equations to estimate the effect of sleep fragmentation on the rate of hippocampal atrophy, defining sleep metrics as the mediator between the history of depression and brain volume loss. We will use bootstrapping with 5,000 iterations to estimate the confidence intervals of this indirect effect, ensuring the statistical robustness of the proposed causal pathway.

In the local Hong Kong cohort, our analysis will focus on establishing concurrent validity and diagnostic accuracy. We will calculate Pearson correlation coefficients to assess the relationship between objective sleep metrics and subjective depression

scores. To evaluate the screening utility of the markers, we will perform Receiver Operating Characteristic analysis and calculate the Area Under the Curve to determine sensitivity and specificity. Finally, we will integrate these findings into a predictive algorithm using Supervised Machine Learning models, such as Random Forest and Gradient Boosting Machines, and we will validate these models using Leave-One-Out Cross-Validation to prevent overfitting given the small sample size.

IX. ETHICAL APPROVAL

The protocol will be submitted to the Institutional Review Board of the University of Hong Kong/ Hospital Authority Hong Kong West Cluster for approval. The study will be conducted according to the ethical principles of "Declaration of Helsinki".

X. POTENTIAL HAZARDS TO PARTICIPANTS

There are no known potential hazards to participants. We will protect subjects by implementing an active safety protocol that triggers immediate psychiatric referral for any participants showing signs of severe depression or self-harm, ensuring faster intervention than standard care. Additionally, we will protect participant privacy by pseudonymizing all data and isolating personal identifiers on an offline work computer.

XI. SOURCE OF FUNDING

This study was sponsored by the new staff startup seed fund by the University of Hong Kong.

.

XII. REFERENCES

[1] Wang, Z., Liu, J., Shuai, H., Cai, Z., Fu, X., Liu, Y., ... & Yang, B. X. (2021). Mapping global prevalence of depression among postpartum women. *Translational Psychiatry*, 11(1), 543.

[2] Lee, D. T., Yip, A. S., Chiu, H. F., Leung, T. Y., & Chung, T. K. (2001). A psychiatric epidemiological study of postpartum Chinese women. *American Journal of Psychiatry*, 158(2), 220-226.

[3] Field, T. (2010). Postpartum depression effects on early interactions, parenting, and safety practices: a review. *Infant Behavior and Development*, 33(1), 1-6.

[4] Grace, S. L., Evindar, A., & Stewart, D. E. (2003). The effect of postpartum depression on child cognitive development and behavior: a review and critical analysis of the literature. *Archives of Women's Mental Health*, 6(4), 263-274.

[5] Lindahl, V., Pearson, J. L., & Colpe, L. (2005). Prevalence of suicidality during pregnancy and the postpartum. *Archives of Women's Mental Health*, 8(2), 77-87.

[6] Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British Journal of Psychiatry*, 150(6), 782-786.

[7] Klainin-Yobas, P., Cho, M. A. A., & Creedy, D. (2009). Postnatal depression in Asian cultures: A literature review. *International Journal of Nursing Studies*, 46(10), 1355-1373.

[8] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

[9] Posmontier, B. (2008). Sleep quality in women with and without postpartum depression. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 37(6), 722-737.

[10] Bei, B., Milgrom J., Ericksen, J., & Trinder, J. (2010). Subjective perception of sleep, but not objective sleep parameters, is associated with immediate postpartum mood disturbances in healthy women. *Sleep*, 33(4), 531-538.

[11] Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., ... & Wareham, N. J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank study. *PloS One*, 12(2), e0169649.

[12] Frodl, T., & O'Keane, V. (2013). How does the brain deal with cumulative stress? A review with focus on developmental stress, HPA axis function and hippocampal structure in humans. Neurobiol Dis, 52: 24-37.

.

[13] Sadeh, A., Acebo, C., Seifer, R., Aytur, S., & Carskadon, M. A. (1995). Activity-based assessment of sleep-wake patterns during the first year of life. *Sleep*, 18(5), 329-337.

XIII. ATTACHMENTS

Appendix I Questionnaires