Persistent Inequality in Maternal Mental Health*

Jacob Hazen† and Fernando Saltiel‡

January 30, 2025

Abstract

In this paper, we analyze persistent inequality in maternal mental health by examining whether and how mothers' latent mental health can be compared across different stages of motherhood for different groups. We introduce a measurement model for binary mental health indicators that draws on multiple measures across survey rounds, establishing identification conditions for valid cross-group and longitudinal comparisons. We implement this framework using rich longitudinal data from Peru (Young Lives Study), covering mothers' mental health over a fifteen-year period and study inequality in mental health across Indigenous and non-Indigenous mothers. We find that mental health measures exhibit strict measurement invariance, enabling direct comparisons across groups and over time. Our results reveal that while both Indigenous and non-Indigenous mothers experience improvements in mental health after the postpartum period, the gains are substantially larger for non-Indigenous mothers, leading to widening disparities as children age. We show that differences in educational attainment and family composition explain most of the initial gap in mental health, but early postpartum mental health becomes increasingly important in driving persistent inequality through children's adolescence.

Keywords: Maternal Mental Health; Factor Models; Measurement Invariance

JEL Codes: C38, I14, J13, J15

*We thank Nicolas Ajzenman, Claire Boone, Jerome Lariviere, Doan Hai Ma, Fabian Lange, Brennan McLachlan, Iones Kelanemer Holban, and Nicolas Pinsonneault, as well as seminar participants at McGill University for useful comments and suggestions.

†McGill University, Department of Economics. Email: jacob.hazen@mcgill.ca

‡McGill University, Department of Economics. Email: fernando.saltiel@mcgill.ca

1 Introduction

Maternal mental health disorders are highly prevalent worldwide, with postpartum depression affecting approximately 13% of mothers in high-income countries and nearly 19% of mothers in low- and middle-income countries (Woody et al., 2017; Shorey et al., 2018). Moreover, these mental health challenges often persist throughout motherhood, with the burden disproportionately falling on socioeconomically disadvantaged groups (O'hara and Swain, 1996). Poor maternal mental health is particularly concerning given its association with various risk factors, including lower socioeconomic status, limited social support, and exposure to stressful life events (Beck, 2001).

At the same time, understanding the extent to which mental health problems prevail across groups and over time presents significant measurement challenges. For instance, mothers from different socioeconomic and cultural backgrounds may interpret and respond to the same mental health questions differently, even when experiencing similar underlying mental health states. Similarly, questions that accurately reflect mental health disorders in the postpartum period may not capture the same underlying construct as children age and mothers face different challenges. As such, understanding the extent of inequality in mental health and the degree to which mental health issues persist over time requires ensuring that mental health measures are comparable both across groups and over time.

In this paper, we study the determinants of persistent inequality in maternal mental health. We first aim to understand the extent to which underlying mental health can be compared across different moments in motherhood and across groups with different characteristics. To this end, we introduce a measurement system that posits that observed multiple measures of mental health are a function of an underlying factor of mental health and measurement error, which can vary by period. The measures are observed for the same individual across multiple periods, for a group of respondents who may pertain to different groups. As such, a key consideration in this context is whether these measures truly capture the same underlying mental health construct for different subpopulations and over time. In particular, this amounts to analyzing whether the measurement system exhibits measurement invariance (Millsap and Yun-Tein, 2004). Establishing such invariance ensures that differences in observed responses reflect true variation in underlying mental health rather than systematic differences in how different groups interpret and respond to these questions.

Mental health measures are often categorical, and in many cases binary, requiring specific methodological considerations for establishing measurement invariance.[1] Drawing on Millsap and Yun-Tein (2004), Wu and Estabrook (2016), and Liu et al. (2017), we introduce the necessary conditions for identification of the measurement model with binary measures. A key challenge in this context is that the variance of binary measures is a deterministic function of their mean, thus reducing the set of moments available for identification. We show how, under basic identification conditions, the measurement system parameters can be recovered non-parametrically. However, to enable comparisons of the latent factors across groups and over time, we establish the necessary conditions for loading invariance.[2]

[1] Binary mental health measures are common screening tools for mental health disorders, through scales such as the Self-Reporting Questionnaire (Beusenberg et al., 1994), the Kessler Psychological Distress Scale (Kessler et al., 2002), and the General Health Questionnaire (Goldberg and Williams, 1988).

[2] We additionally allow for the measurement errors of the same measure to be correlated over time, which can pick up the fact that a person may be consistently more likely to say yes to a specific question in multiple periods. Moreover, strict

We implement this framework by drawing on unique data from the Young Lives Study in Peru, which follows mothers and their children from 2002 through 2016. The study administered the Self-Reporting Questionnaire (SRQ-20) to mothers when their children were aged 1, 5, 12 and 15, providing a consistent measure of mental health across this fifteen-year period.[3] Given substantial socioeconomic disparities between Indigenous and non-Indigenous populations in Peru (Freire et al., 2015), we implement our framework to understand the extent to which mothers' mental health is comparable across time and between these groups. Even in raw measures, we observe substantial disparities, with Indigenous mothers reporting 5.95 symptoms of mental distress in the baseline survey compared to 5.35 symptoms among non-Indigenous mothers. While both groups show improvements over time, these gaps persist, with Indigenous mothers still reporting 1.2 more symptoms than their non-Indigenous counterparts by the endline round.

The implementation of our measurement model reveals that the SRQ-20 exhibits strict invariance across periods and between Indigenous and non-Indigenous mothers. This finding indicates that mothers' responses to these questions reflect the same underlying mental health construct across groups and over time, enabling meaningful comparisons of latent mental health. Using the estimated model parameters, we find that Indigenous mothers exhibit substantially worse mental health than their non-Indigenous counterparts at baseline, with an average difference of 0.12 standard deviations. This gap is particularly pronounced at the bottom of the distribution, with Indigenous mothers at the 10th percentile scoring 0.25 standard deviations below their non-Indigenous counterparts, while differences at the 90th percentile are not statistically significant.

These mental health disparities evolve differently across the distribution as children age. While both groups experience improvements in mental health over time, these gains are more substantial for non-Indigenous mothers, whose average mental health improves by 0.7 standard deviations between baseline and Round 2, compared to 0.4 standard deviations for Indigenous mothers. As a result, mental health disparities widen over time, with the gap in average mental health increasing from 0.12 to 0.37 standard deviations between baseline and Round 5. These patterns are particularly pronounced at the top of the distribution, where Indigenous mothers move from having similar mental health as their non-Indigenous counterparts at baseline to exhibiting substantially worse outcomes by Round 5.

We further use the estimated model to understand the extent of persistence in maternal mental health issues. We find substantial persistence in mental health across survey rounds, particularly for non-Indigenous mothers. Initial mental health exhibits strong persistence through age 5, as a one standard deviation increase in mental health at baseline is associated with a 0.48 standard deviation increase in mental health in Round 2 for non-Indigenous mothers, compared to 0.29 for Indigenous mothers. Furthermore, this persistence remains even through their children turning 15, as baseline mental health is associated with a 0.36 standard deviation increase in mental health for non-Indigenous mothers and 0.21 for Indigenous mothers fourteen years after the initial survey.

In light of the sizeable disparities in mental health, we aim to understand the factors driving these

invariance additionally allows us to posit that any changes in the response patterns for a measure (and across groups) is driven by differences in the factor.

[3]Postpartum depression comprises the period from four weeks after giving birth up to one year (Miller, 2002). Since baseline round was conducted when the child was one year old, we often refer to baseline mental health as postpartum mental health.

gaps through a RIF regression approach Firpo et al. (2009, 2018). This decomposition method allows us to examine the extent to which gaps at different percentiles of the distribution are explained by background characteristics. We find that ethnic differences in postpartum mental health at the 10th percentile are fully explained by differences in observed characteristics, with mother's education and the number of children playing particularly important roles. While gaps are close to zero at the upper end of the distribution, we still find that differences in these variables would lead to a substantial reduction in the gap at the median and through the 90th percentile.

We further examine how the drivers of inequality in maternal mental health evolve over time by estimating the RIF decomposition in subsequent survey rounds while incorporating the role of prior mental health. By Round 2, mental health from the first survey round explains between 15-30% of the total gap across the distribution, while socioeconomic characteristics become less predictive of mental health disparities. Furthermore, by the endline survey round, we find that lagged mental health accounts for a larger share of the total gap across the distribution, underscoring how early disparities in mental health persist over a fifteen-year period . The large unexplained component, particularly at the top of the distribution, points to cumulative advantages in non-Indigenous mothers' mental health trajectories that are not fully explained by baseline socioeconomic characteristics. Our measurement framework thus reveals how early inequalities in maternal mental health can have persistent impacts, serving as a key mechanism through which initial socioeconomic disparities translate into long-term inequalities between Indigenous and non-Indigenous mothers.

Our paper makes several contributions to the literature. First, we introduce a measurement framework that establishes the identification conditions under which binary indicators can be compared across groups and over time. While the economics literature in economics has extensively studied the identification and estimation of latent factors with continuous measures (Carneiro et al., 2003), particularly in the context of skill formation (Cunha and Heckman, 2008; Cunha et al., 2010), less attention has been paid to the specific challenges that arise when identifying measurement systems when using binary measures. As such, our work directly fits in with Attanasio et al. (2020c), who examine measurement invariance in socioemotional skills using categorical constructs across cohorts. Our framework additionally helps us understand the drivers of inequality in maternal mental health, further contributing to a growing literature in economics examining the determinants of mothers' psychological well-being (Baranov et al., 2020; Sevim et al., 2024). At the same time, we remark that our framework can be applied to other domains where binary measures are prevalent, such as measuring socioemotional skills or analyzing responses to parental investment questions.

Second, we contribute to both the psychometric and health literature by establishing measurement invariance of maternal mental health across groups and over time. Our approach builds on methods for establishing measurement invariance with categorical variables (Millsap and Yun-Tein, 2004; Wu and Estabrook, 2016), their longitudinal extensions (Liu et al., 2017), and testing procedures for establishing valid comparisons across groups (Putnick and Bornstein, 2016). While prior work has examined measurement invariance of mental health constructs separately across socioeconomic groups and over time, we show that the SRQ-20 exhibits strict invariance both longitudinally and between Indigenous and non-Indigenous populations. This finding is particularly relevant given extensive evidence documenting socioeconomic disparities in maternal mental health (Woody et al., 2017; O'hara and Swain, 1996) and the persistence of these disparities throughout motherhood (Vliegen et al., 2014; Moore Simas et al., 2019).

Finally, we contribute to understanding the determinants and persistence of inequality in maternal mental health. Through distributional decomposition methods (Firpo et al., 2009, 2018; Attanasio et al., 2020c), we document how socioeconomic characteristics shape mental health disparities across the distribution and over time. Our results reveal that early postpartum mental health becomes increasingly important in explaining persistent disparities, building on work examining the long-term trajectories of maternal mental health (Wang et al., 2011; Jackson et al., 2017). By establishing both the strong persistence in mental health trajectories and their systematic variation across groups, we show how initial inequalities compound over time, potentially contributing to intergenerational transmission of disadvantage (Burke, 2003; Aizer et al., 2016).

The rest of the paper is structured as follows. Section 2 introduces our measurement framework for modeling multiple binary indicators of maternal mental health and derives the identification conditions required for valid comparisons across groups and over time. Section 3 describes the Young Lives Study data from Peru and provides summary statistics on mothers' backgrounds and SRQ-20 responses. In Section 4, we test for measurement invariance, present the estimation results, and discuss how the SRQ-20 can be used to compare mental health outcomes across Indigenous and non-Indigenous mothers over a fifteen-year period. Section 5 documents the persistent inequalities in maternal mental health that emerge from our estimates, while Section 6 explores how these disparities evolve and decompose their drivers over time. Finally, Section 7 concludes.

2 Methodological Framework

In this section, we introduce a measurement system for a set of observed binary mental health measures that reflect respondents' underlying (unobserved) mental health. The measures are observed for the same individual across multiple periods, for a group of respondents who may pertain to different groups. As such, a key consideration in this context is whether these binary measures truly capture the same underlying mental health factor across different subpopulations and over time – that is, whether the measurement system exhibits *measurement invariance*. Establishing such invariance with binary indicators is particularly challenging because the variance of the observed measures is a direct function of the mean, thus reducing the set of moments that can be used for model identification.

In this context, we draw on Millsap and Yun-Tein (2004), Millsap et al. (2012), Wu and Estabrook (2016), and Liu et al. (2017), and introduce the necessary conditions under which these binary measures can be used to establish longitudinal and cross-group invariance. As such, we show how this framework can be used to make valid and consistent comparisons of mothers' latent mental health across groups and over time when only observing responses to binary measures.

2.1 Measurement System of Mental Health

We consider a set of observed measures of mental health $\{X_{ijgt}\}$, where i represents individuals, j captures different measurement items, g indexes distinct groups (comprising subpopulations, or cohorts), and t indexes time (or age). We observe p binary measures of maternal mental health for each group in each time period t, such that $X_{ijgt} \in \{0,1\}$ for $j \in \{1,\ldots,p\}$, for $g \in \{G_1,...,G\}$ and $t \in \{1,\ldots,T\}$. In this context, we posit the observed indicators of mental health to be a manifestation of mothers' latent mental health, albeit measured with error.[4]

In the factor model for binary measures (Millsap and Yun-Tein, 2004; Wu and Estabrook, 2016; Liu et al., 2017), the observed measures are assumed to be determined by latent responses X^*_{ijgt} that underlie each binary measure, X_{ijgt}. The latent responses are assumed to be continuous and are linked to the observed measures through the following equation:

$$X_{ijgt} \ = \begin{cases} 1, & \text{if } X^*_{ijgt} \ > \ \tau_{jgt}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where τ_{jgt} is a threshold parameter that is allowed to vary across individuals, groups and over time. As such, the observed binary measures of mental health are a deterministic function of the latent response, such that when the latent response of person i for measure j exceeds the corresponding threshold ($X^*_{ijgt} > \tau_{jgt}$), the person answers affirmatively to that question.

We follow the existing literature on latent factor models with categorical variables, and posit that the latent response X^*_{ijgt} is a linear function of mothers' latent mental health, which is given by ξ_{igt}, in the following linear specification:

$$X^*_{ijgt} \ = \ v_{jgt} + \lambda_{jgt}\,\xi_{igt} + \delta_{ijgt}, \tag{2}$$

where v_{jgt} is an intercept term, λ_{jgt} is a factor loading capturing how strongly item j relates to the latent factor at group g and time t, and δ_{ijgt} reflects an item-specific error term.

In equation (2), latent mental health ξ_{igt} is allowed to vary across individuals, groups, and over time, with a mean of κ_{gt} and variance ϕ_{gt}. Importantly, to capture the persistence of mother's latent mental health, we allow for ξ_{igt} to be correlated across periods, where the covariance between mental health in periods t_0 and t_1 is given by $\phi_{gt_0t_1}$. As such, the variance-covariance matrix of the latent factors is given by:

$$\Phi_g = \begin{bmatrix} \phi_{g1} & \cdots & \phi_{g1T} \\ \vdots & \ddots & \vdots \\ \phi_{g1T} & \cdots & \phi_{gT} \end{bmatrix}$$

where ϕ_{gts} ($t \neq s$) denotes the covariance between ξ_{igt} and ξ_{igs}, capturing how maternal mental health evolves across time in group g. We further assume that the error terms are mean zero with variance θ_{jgt}. As such, the mean and the variance of the latent responses $\{X_{ijgt}\}^*$ are given by:

$$E(X^*_{ijgt}) = \mu_{jgt} = v_{jgt} + \lambda_{jgt}\kappa_{gt} \quad , \quad \text{Var}(X^*_{ijgt}) = \Sigma_{jgt} = \lambda^2_{jgt}\,\phi_{gt} + \theta_{jgt} \tag{3}$$

where μ_{jgt} and Σ_{jgt} denote the mean and the variance of the continuous latent measure j for group g in period t, respectively.

We follow the existing literature on latent factor models and assume the error terms to be in-

[4] For expositional simplicity, we focus on a single-factor specification, yet the framework can incorporate multiple dimensions of the latent factors (Millsap and Yun-Tein, 2004).

dependent from the latent factors, both contemporaneously ($\xi_{igt} \perp \delta_{ijgt}$), as well as across periods ($\xi_{igt} \perp \delta_{ijgs}$ for $t \neq s$). At the same time, we assume that the measurement errors are independent across measures in the same period ($\delta_{ijgt} \perp \delta_{ikgt}$ for $j \neq k$), but, we allow for the error terms pertaining to the same measure to be correlated *across* periods, following Millsap et al. (2012) and Liu et al. (2017). As such, if a respondent systematically underreports whether they felt unhappy across multiple survey rounds, this flexible error structure would account for such response patterns, instead of attributing it to changes in the latent factor. The $pT \times pT$ variance-covariance matrix of the measurement errors for group g (Θ_g) is thus given by:

$$\Theta_g = \begin{bmatrix} \Theta_{g,11} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Theta_{g,pp} \end{bmatrix} \quad \text{where} \quad \Theta_{g,jj} = \begin{bmatrix} \theta_{jg1} & \cdots & \theta_{jg1T} \\ \vdots & \ddots & \vdots \\ \theta_{jgT1} & \cdots & \theta_{jgT} \end{bmatrix} \tag{4}$$

where the each diagonal block $\Theta_{g,jj}$ is a $T \times T$ matrix capturing the covariances of the measurement errors for item j across T periods.[5] The off-diagonal blocks are equal to zero, reflecting the assumption that measurement errors are uncorrelated across different items either in the same period or across periods.

Altogether, the parameters in the measurement system presented in this section – comprising the latent factors, the thresholds, loadings, intercepts, and error terms – are allowed to vary across groups and over time. However, the full set of model parameters is not identified even for a particular group in a specific time period unless we impose additional restrictions on the measurement system.

2.2 Model Identification

The measurement system presented in equations (1)-(4) requires additional restrictions to be identified. An essential challenge in a model that draws on binary measures is that the variance of the observed measures is a deterministic function of the mean of the measures, as $\text{Var}(X_{ijgt}) = E[X_{ijgt}]\left(1 - E[X_{ijgt}]\right)$. As such, the econometrician can only draw on the means of the observed measures and on the covariance across measures for identification. In this context, this yields

$$\frac{pT(pT+1)}{2}$$

observed moments for identification, comprised of pT item-level means and $\frac{pT(pT-1)}{2}$ covariance across the measures for each group.[6]

At the same time, the model introduced in Section 2.1 contains a large number of parameters to be identified for each group. In particular, identifying the model presented in equations (1)-(4) requires the identification of pT item-level intercepts (ν_{jgt}), pT factor loadings (λ_{jgt}), T latent factor means (κ_{gt}), T factor variances (ϕ_{gt}), and $\frac{T(T-1)}{2}$ factor covariances across periods (ϕ_{gts}). Moreover, the flexible structure on the measurement errors introduced in equation (4) requires identifying pT variances (θ_{jgt}) and $\frac{pT(T-1)}{2}$ off-diagonal covariances for the same item j across periods (θ_{jgts}). Lastly,

[5] The diagonal elements of this sub-matrix reflect the variance of the error term for measure j in period t. The off-diagonal measures capture the covariance of the error term across periods t and s for $t \neq s$, θ_{jgts}.

[6] In contexts with continuous measures, or when using categorical measures with more than two responses, we could additionally draw on the pT item-level variances as moments for identification.

to link the observed binary measure (X_{ijgt}) to the continuous latent response (X^*_{ijgt}), we must additionally recover pT item-level thresholds (τ_{jgt}). Altogether, identification of the model introduced in equations (1)-(4) requires identifying

$$(2p+1)\cdot(2T)+\frac{(p+1)\cdot T\cdot(T-1)}{2}$$

model parameters for each group.[7]

In this context, we introduce a set of assumptions required to identify the model parameters from the set of available moments in the data. We draw on Millsap and Yun-Tein (2004) and Wu and Estabrook (2016)'s arguments for model identification for measurement models with categorical variables in cross-sectional data, and on (Liu et al., 2017)'s identification arguments for categorical variables in a longitudinal setting. As such, we present the set of assumptions required to identify a model with a set of binary measures observed for multiple groups across multiple periods. In particular, we posit that the following restrictions are sufficient for identification:

Assumption 1 (Measurement Location)**.**

- $v_{jgt} = 0$ *for all* $j \in \{1,\ldots,p\},\ \ g \in \{G_1,\ldots,G\},\ \ t \in \{1,\ldots,T\}$.
- $\kappa_{G_1,1} = 0$ *for* $g = \{G_1\}$ *and* $t = \{1\}$.

Assumption 2 (Threshold Invariance)**.** $\tau_{jgt} = j$ *for all* $j \in \{1,\ldots,p\}, g \in \{G_1,\ldots,G\}, t \in \{1,\ldots,T\}$.

Assumption 3 (Measurement Scale)**.** $\lambda_{1gt} = 1$ *for* $j = 1$, *for all* $g \in \{G_1,\ldots,G\}$, $t \in \{1,\ldots,T\}$.

Assumption 4 (Measurement Error Structure)**.**

- $\theta_{j,G1,1} = 1$ *for* $g = G_1$, *t=1*, *for all* $j \in \{1,\ldots,p\}$.
- $\theta_{1gt} = 1$ *for all* $g \in \{G_1,\ldots,G\}$, $t \in \{1,\ldots,T\}$.

These assumptions introduce a set of restrictions on the measurement system parameters. Assumption 1 sets all intercepts to zero, as both intercepts and thresholds cannot be simultaneously identified in models with binary measures, and normalizes the mean of the latent factor for the reference group ($g = \{G_1\}$) in the first period to zero.[8] Assumption 2 imposes threshold invariance across groups and over time, which ensures the equivalence of the mapping between the observed binary measure (X_{ijgt}) and the underlying continuous response (X^*_{ijgt}) across different subpopulations and periods. Assumption 3 introduces a standard scale normalization by setting the factor loading of the first item to one. Assumption 4 imposes two restrictions the measurement error structure. First, the variances of the error terms are equal to one in the first period for all groups (Liu et al., 2017). Second, the variance of the reference item ($j = 1$) must equal unity for all periods for the reference group. We impose this additional restriction on the variance of the reference item across all periods,

[7] The set of parameters to be identified in a one-period model is substantially smaller, amounting to $4p+2$ model parameters, as the off-diagonal covariances in the factors and in the error terms are equal to zero.

[8] With binary outcomes, the item intercepts and thresholds cannot be freely estimated at the same time, since shifting both would leave the model-implied probabilities unchanged (see also Millsap and Yun-Tein, 2004).

as we cannot impose restrictions on a second threshold in the context of binary measures (Millsap and Yun-Tein, 2004; Wu and Estabrook, 2016).[9]

Upon imposing Assumptions 1–4, the set of model parameters to be identified equals $pT(T+3)+T^2-T-(p-1)$ for the reference group and $pT(T+3)+T^2-T$ for other groups.[10] Given the observed moments in the data, the model parameters are identified as long as:[11]

$$pT\,(pT+1) \;\geq\; (1+p)\,T^2+(3p-1)\,T.$$

As such, in a cross-sectional specification ($T=1$), identification requires observing at least three measurement items ($p \geq 3$), a standard requirement in cross-sectional measurement models with binary variables (Millsap and Yun-Tein, 2004), as well as in measurement systems considered in the Economics literature with continuous variables (Carneiro et al., 2003). In a longitudinal setting, while models with two periods ($T=2$) still require three measurement items for identification, the conditions become less restrictive as we observe individuals across more periods, requiring only two measurement items ($p \geq 2$) when the model includes more than two time periods.

While Assumptions 1–4 allow us to identify the full set of measurement system parameters, they do not provide sufficient structure to make meaningful comparisons of mothers' latent mental health across groups. First, Assumption 2 requires threshold invariance, which assumes the mapping between the observed binary measure (X_{ijgt}) and the underlying continuous response (X^*_{ijgt}) to be equivalent across different subpopulations and periods. With binary measures, this assumption cannot be tested, as it is required for model identification (Millsap and Yun-Tein, 2004; Wu and Estabrook, 2016). Second, even with threshold invariance, comparisons of the latent factor means (κ_{gt}) or variances (ϕ_{gt}) across groups remain invalid, as differences in these parameters could reflect variation in the factor loadings (λ_{jgt}) rather than true differences in mothers' latent mental health. As such, additional restrictions on the measurement system are required to establish that observed differences in mothers' mental health reflect true variation in their latent factor, rather than differences in how the binary measures relate to mental health across groups.

2.3 Measurement Invariance

Measurement invariance determines whether the binary measures exhibit equivalent psychometric properties across groups and time periods, ensuring they capture the same underlying construct. Testing for measurement invariance involves progressively imposing stronger restrictions on the measurement system parameters to ensure that observed differences in mothers' responses reflect true variation in their latent mental health rather than differences in how these measures relate to the underlying factor and comparing the relative fit to less restrictive models (Putnick and Bornstein, 2016). This nested-model approach to testing measurement invariance, in which increasingly restrictive con-

[9] Despite imposing restrictions on the variances of the measurement errors, the correlation in the measurement errors across periods for the same item remain unrestricted.

[10] For all groups, we impose pT restrictions on the intercepts ($\nu_{jgt}=0$), T restrictions on the factor loadings for the reference item ($\lambda_{1gt}=1$), and T restrictions on the error variances for the reference item ($\theta_{1gt}=1$). Additionally, for the reference group ($g=\{G_1\}$), we restrict one factor mean ($\kappa_{G_1,1}=0$) and impose $p-1$ additional restrictions on the error variances in the first period ($\theta_{j,G1,1}=1$).

[11] Since other groups have $p-1$ more parameters to identify than the reference group, this condition ensures identification for all groups if it holds for $g \neq \{G_1\}$.

straints are imposed on the measurement system parameters, is essential for validating comparisons of mothers' mental health across groups and time periods.

In our setting with binary measures, we draw on Millsap and Yun-Tein (2004) and Wu and Estabrook (2016) to directly test for the joint invariance of thresholds and factor loadings, as threshold invariance is required to identify the measurement system (Assumption 2). We then assess the equivalence of measurement errors across groups and time, evaluating how model fit changes with each additional set of restrictions (Liu et al., 2017).

Threshold and Loading (Scalar) Invariance. The threshold and loading invariant model imposes equality restrictions on the factor loadings across groups and over time, while maintaining the threshold invariance required for identification.[12] Formally, we introduce the following assumption:

Assumption 5 (Loading Invariance)**.** *The factor loadings are equal across groups and time periods:*

$$\lambda_{jgt} = \lambda_j \textit{ for all } j \in \{1,...,p\}, g \in \{G_1,...,G\}, t \in \{1,...,T\}.$$

Combined with threshold invariance (Assumption 2), these restrictions ensure that the mapping between mothers' latent mental health (ξ_{igt}) and their binary responses (X_{ijgt}) is equivalent across groups and time periods. That is, for a given level of the latent factor, the probability of mothers answering affirmatively to item j should not differ across groups or time due to shifts in the factor loading (λ_{jgt}) or threshold (τ_{jgt}).

Under scalar invariance, differences in the observed binary responses across groups and time must reflect differences in the latent factor distribution (ξ_{igt}), as the mapping between the latent factor and observed measures is equivalent across groups. Specifically, when two mothers from different groups g and g' or time periods t and t' have the same latent mental health ($\xi_{igt} = \xi_{ig't'}$), they must have the same probability of answering affirmatively to item j, since $\lambda_{jgt} = \lambda_{jg't'}$ and $\tau_{jgt} = \tau_{jg't'}$. This allows us to interpret differences in the factor means (κ_{gt}) and variances (ϕ_{gt}) as true variation in mothers' mental health across groups and over time (Millsap and Yun-Tein, 2004; Wu and Estabrook, 2016; Liu et al., 2017). As such, when scalar invariance holds, changes in the share of mothers who answer affirmatively to specific questions are not explained by changes in item-level parameters, are instead driven by changes in the distribution of latent mental health.

This property is crucial in our setting, as it enables us to assess how mothers' mental health differs across groups in each period and to compare the evolution of mental health across different subpopulations. With multiple items per round, we can test whether loadings and thresholds remain constant over time and across groups, and then examine the resulting factor means to interpret how maternal mental health differs by group g in each round t. Scalar invariance is critical for ensuring that cross-group or longitudinal comparisons in means reflect genuine changes in mothers' latent mental health rather than to differences in how mothers respond to the same item.

[12] Scalar invariance captures whether differences in the means of observed measures stem from differences in the latent construct (Putnick and Bornstein, 2016). In a context with continuous, or categorical variables with multiple responses, it is possible to separately test loading (metric) and threshold (scalar) invariance. However, threshold invariance is a necessary assumption for model identification with binary measures, implying that metric and scalar invariance are not separately testable in our setting (Millsap and Yun-Tein, 2004).

Threshold, Loading and Error (Strict) Invariance. While scalar invariance ensures comparability of the latent factor means and variances, we can consider additional restrictions on the measurement system to compare the changes in responses to individual items. In particular, the strict (threshold, loading and error) invariant model further restricts the variance of the measurement errors to be equivalent across groups and over time by assuming:

Assumption 6 (Measurement Error Invariance)**.** *The variance of the measurement errors is equal across groups and time periods:*

$$\theta_{jgt} = \theta_j \textit{ for all } j \in \{1,...,p\}, g \in \{G_1,...,G\}, t \in \{1,...,T\}.$$

Under these restrictions, each binary measure exhibits the same residual variation across groups and over time. A measurement system that exhibits strict invariance implies that two mothers with the same level of latent mental health (ξ_{igt}) would have have identical probabilities of answering affirmatively to a given question.[13]

3 Data Sources and Summary Statistics

3.1 Young Lives Study

In this study, we use data from the Young Lives study, which tracks children across four low- and middle-income countries — Ethiopia, India, Peru, and Vietnam — over a two-decade period. Our work focuses on the Younger Cohort, who were on average one year old in the baseline survey carried out in 2002, and who were subsequently followed in follow-up surveys carried out in 2006 (Round 2), 2009 (Round 3), 2012 (Round 4), and 2016 (Round 5). Each survey round collected extensive information on parental outcomes, including on their educational attainment, health outcomes, and general household socioeconomic characteristics, as well as information on children's educational and health outcomes through parental, child and community questionnaires.[14]

In our analysis, we use Young Lives data from Peru, as it is the only country within the study that collects measures of maternal mental health across all survey rounds. The sampling design in Peru followed a multi-stage approach, as 20 districts were randomly selected (excluding the wealthiest 5% based on district-level poverty measures), followed by selection of villages within districts, and systematic sampling of households with children born in 2001-2002 (Escobal and Flores, 2008). This process identified approximately 100 eligible households per district, generating a baseline sample of 2,052 index children in the Younger Cohort. The cumulative attrition rate through the fifth survey round remained notably low at about 9 percent (Sánchez and Escobal, 2020), ensuring consistent longitudinal tracking of both children and caregivers.

To understand the potential drivers of maternal mental health, we draw on various socioeconomic and demographic measures collected in the baseline survey round. First, we consider mother's educational attainment, measured by the years of completed schooling. We use the Young-Lives constructed wealth index, which is an average of measures capturing housing quality, consumer durables,

[14]Note that the Young Lives data has been used extensively in the economics literature, see Attanasio et al. (2017, 2020a); Singh (2020); Lopez et al. (2024), among others.

and access to services into a single metric. We also include information on the mother's health status, through their height and weight, as well as the number of children the mother had prior to the index child and whether the father resides in the home in the baseline round. These measures enable us to examine how various dimensions of household characteristics relate to maternal mental health outcomes.[15]

As discussed in the introduction, we additionally examine differences in maternal mental health between Indigenous and non-Indigenous mothers in Peru. Indigenous populations in Peru face substantial economic and social disparities, with far higher poverty rates, lower access to public services and significantly lower educational attainment than their non-Indigenous counterparts (Freire et al., 2015). In our context, we follow the literature and classify mothers to be Indigenous if they speak an Indigenous language, primarily Quechua or Aymara, and to be non-Indigenous otherwise (Arteaga and Glewwe, 2019; Sánchez et al., 2020; Posso, 2023).

3.2 Maternal Mental Health Measures

The Self-Reporting Questionnaire (SRQ-20) is a screening tool of 20 dichotomous items ("yes" or "no") designed to identify symptoms associated with anxiety and depression over the past 30 days. Developed by the World Health Organization (WHO), the SRQ-20 is intended to detect "common mental disorders" (CMDs) in resource-constrained settings, with questions targeting symptoms such as poor sleep, loss of appetite, and nervousness (Beusenberg et al., 1994). Due to its simplicity, low cost, and ease of administration, the SRQ-20 has become a widely used approach for mental health screening in low- and middle-income countries.

The SRQ-20 has been extensively employed and validated across developing countries (Harpham et al., 2003; Tuan et al., 2004; Netsereab et al., 2018). Furthermore, various development agencies, government organizations and non-governmental organizations include the SRQ-20 in surveys to monitor mental health trends, evaluate policy interventions, and to implement mental health screening initiatives. While the SRQ-20 has been shown to exhibit reliability and cultural adaptability in low-resource settings, this construct does not explicitly differentiate different dimensions of mental health, such as anxiety and depression, as it serves primarily as a general screening instrument.

Within the Young Lives study, the SRQ-20 was administered longitudinally to caregivers and, in later rounds, specifically to mothers, allowing us to track maternal mental health over time, comprising a fifteen year period across the five survey rounds.[16] The questionnaire comprises 20 items that address symptoms experienced in the last 30 days, as listed in Table A1. The SRQ-20 questions focus on physical and emotional symptoms commonly associated with psychological distress and mental health challenges. The questions range from physical manifestations like headaches and poor appetite to emotional symptoms such as unhappiness and difficulty making decisions.[17]

[15] We additionally consider measures of mother's social capital (constructed from five questions on community ties and trust), and whether the pregnancy of the index child had been desired.

[16] We restrict the sample to caregivers who were also the index child's mothers in the baseline round, which was the case for 98.7% of the sample. In our empirical analysis, we do not include responses from Round 3, as the identity of the respondent to the mental health questions was not directly discernible from the survey design.

[17] The optimal cutoff scores for the SRQ-20, above which an individual is classified as likely having a mental health disorder, vary across populations and settings, yet commonly fall in the range of 6-7 positive responses (Harpham et al., 2003; van der Westhuizen et al., 2016).

3.3 Sample Selection and Descriptive Statistics

To study the evolution of maternal mental health over time, we restrict our sample to mothers who were between 16 and 45 years old in the baseline survey and who answered the full set of SRQ-20 across all survey rounds. As such, we observe a consistent measure of maternal mental health across all survey rounds. Upon imposing these restrictions, our final sample consists of 1,270 mothers, of which 429 are indigenous and 841 are non-indigenous.[18]

The first column of Table 1 presents summary statistics for the mothers included in the main sample. Only 32% of mothers had completed a high school degree by the baseline survey round, fitting in with the Young Lives survey design. Mothers in our sample were 27 years old on average and had already given birth to an average of 1.6 children prior to having the index child. Moreover, 87% of these mothers lived with the father at the time of the baseline survey. The second and third columns show that Indigenous mothers had significantly lower educational attainment, having completed an average of five years of schooling by the baseline round, compared to 8.7 years for their non-Indigenous counterparts. The wealth gap between these groups is also substantial, reaching 0.7 standard deviations (Escobal, 2012). Indigenous mothers had also had more children by the baseline survey round, and were more likely to be living with the father of the index child.

In Table 1, we additionally show the evolution of SRQ-20 scores across survey rounds. In the baseline round, mothers reported 5.5 symptoms of mental distress, on average. Nonetheless, mothers' mental health significantly improved by the time their index children had turned five, as the number of symptoms of mental distress declined to 3.5, on average. At the same time, mental health appeared to slightly better at later periods, as the average SRQ-20 score decreased to 4.3 and 4.2 in Rounds 4 and 5, respectively. Altogether, these results fit in with extensive evidence in the literature showing improvements in maternal mental health during early childhood (Schmied et al., 2013; Baranov et al., 2020; Sevim et al., 2024).[19]

At the same time, we find substantial disparities in mental health between indigenous and non-indigenous mothers. In the baseline survey, indigenous mothers reported an average of 5.9 symptoms of mental distress, compared to 5.3 symptoms among non-indigenous mothers. This gap persisted throughout the survey period, with indigenous mothers consistently reporting between 0.9 and 1.2 more symptoms than their non-indigenous counterparts. By Round 5, when the index children were 15 years old, indigenous mothers still reported 1.21 more symptoms on average.[20] Lastly, Figure A1 presents the distribution of SRQ-20 scores by ethnicity over time, showing that mental health disparities extend beyond the mean, as Indigenous mothers report more mental health symptoms than their non-Indigenous peers across the SRQ-20 distribution.[21]

While the SRQ-20 results presented in Table 1 indicate an improvement in maternal mental health

[18]The baseline SRQ-20 survey is answered by 1,966 mothers aged 16-45 in the baseline survey. We further restrict the sample to mothers who answered the survey in Rounds 2, 4 and 5, which further reduces our sample to 1,270 mothers.

[19]Table A2 shows that Indigenous mothers most frequently reported feeling tense (62%), unhappy (52%), and frightened (51%) in the first round. In comparison, non-Indigenous mothers most frequently reported feeling tense (65%), unhappy (40%), and having difficulties with decision-making (48%) in the same round. Most symptoms showed substantial improvements in subsequent rounds for both groups.

[20]Tables A3-A4 show positive correlations across all SRQ-20 measures at baseline, with similar patterns for both groups.

[21]Such distributional disparities also emerge when we capture mental health disorders using the 7-point cutoff suggested in the literature. 39% of Indigenous mothers are depressed at baseline in this definition, and while the rate falls to 30% by endline, it consistently exceeds the corresponding rate for non-Indigenous mothers by 5-8 percentage points.

over time along with substantial disparities across groups, these reported measures capture mothers' underlying mental health with error. Moreover, the extent of measurement error may vary across groups and over time, making it difficult to draw conclusions from the descriptive evidence presented above. As such, to properly assess mental health disparities in our sample, we next implement the framework introduced in Section 2.

4 Measurement Invariance and Latent Mental Health

4.1 Model Implementation and Estimation

In this section, we estimate the measurement model from the framework introduced in Section 2. Specifically, we use the SRQ-20 measures to estimate a measurement system with a one-factor structure across the four survey rounds for Indigenous and non-Indigenous mothers.[22,23] Following the minimally identifiable model outlined in Section 2.2, we set all measurement intercepts equal to zero, set the item-thresholds to be equal across periods and groups, and assume the variance of the error terms to be equal to one — which corresponds to the 'Theta' parameterization.[24]

To estimate the measurement model, we follow the literature and assume multivariate normality of the unobserved latent responses (Muthén, 1984). We estimate the model using mean- and variance-adjusted Diagonally Weighted Least Squares (DWLS), implemented in *lavaan* (Rosseel, 2012).[25] DWLS reduces computational demands by using only diagonal elements of the weight matrix while maintaining desirable asymptotic properties (Liu et al., 2017), and is the most commonly used estimator in the literature (Toppeta, 2022; Attanasio et al., 2020c, 2024).

4.2 Testing for Measurement Invariance

As discussed in Section 2.3, testing for measurement invariance is essential to establish whether a construct is measured consistently across different groups and over time. In our context, we analyze whether differences in mothers' responses to specific SRQ-20 questions across survey rounds and between Indigenous and non-Indigenous populations reflect changes in their underlying mental health or whether they stem from systematic differences in how these groups interpret and respond to the questions. For instance, the question 'Did you cry more than usual?' might capture different aspects of mental health for Indigenous and non-Indigenous mothers due to cultural differences in expressing emotions. Similarly, this question might reflect different dimensions of mental health over time – capturing post-natal depression when children are one year old, but potentially becoming less salient as children grow older and mothers face different challenges.

[22] For ease of interpretation, we reverse the SRQ-20 measures, allowing us to interpret positive values as reflecting better mental health and lower values as worse mental health.

[23] We determine the number of latent factors supported by the SRQ-20 data (Table A5). Various factor retention methods that are suitable for binary measures (Finch, 2023) recommend retaining one factor across groups and over time. We therefore proceed with a single-factor specification in the estimation of our empirical analysis.

[24] We additionally allow the variance of the error terms to be correlated across periods for the same measure.

[25] Common estimation approaches for categorical factor models include Maximum Likelihood, Weighted Least Squares (WLS), Diagonally Weighted Least Squares (DWLS), and Unweighted Least Squares (ULS). WLS, while optimal for large samples, is computationally intensive and prone to non-convergence in complex models.

To test for measurement invariance, we first estimate the minimally comparable (configural) model presented in Section 2.2. In a context with binary measures, this entails imposing equality of thresholds across groups and setting the variance of the error terms in the first period to equal one, while allowing the remaining measurement system parameters to vary across groups and over time. Following Putnick and Bornstein (2016), we consider the model to exhibit an acceptable fit when the Comparative Fit Index (CFI) exceeds 0.90, while a CFI above 0.95 indicates good fit. If the configural model fits the data well, mothers' responses to the SRQ-20 can be explained by the same unifying factor structure across groups and over time. However, making meaningful comparisons of the latent factor variances or means require additional assumptions in this context.

We therefore impose additional constraints on the measurement system to test whether observed changes in SRQ-20 responses are driven solely by differences in respondents' underlying mental health. First, we test the fit of a 'Threshold and Loading Invariance' (scalar) model, which imposes the equality of factor loadings across groups and over time. Given that measurement invariance is assessed through the relative fit of nested models (Wu and Estabrook, 2016), we examine whether imposing loading invariance across groups and time significantly worsens model fit relative to the configural model. As discussed in Section 2.3, establishing scalar invariance in this context implies that any differences in latent mental health across groups and over time would be explained by changes in mothers' responses to the observed measures. We lastly test a model that imposes the equality of measurement error terms across groups and over time — corresponding to the strict invariance model (Liu et al., 2017). If strict invariance holds, it implies that any observed differences in item responses can be attributed purely to disparities in the underlying maternal mental health factor, rather than to measurement discrepancies or systematic shifts in error variances.

To assess measurement invariance in this longitudinal context, we rely on approximate fit indices (AFI) rather than traditional fit measures such as the χ^2, which can be sensitive to sample size and model complexity.[26] Following Cheung and Rensvold (2002), we consider a change in CFI (ΔCFI) of less than 0.01 between nested models as evidence of invariance. Chen (2007) proposes additional thresholds for the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) of 0.015 and 0.030, respectively. In our context, we focus primarily on the CFI and RMSEA given that the SRMR has not been systematically validated for longitudinal models (Little and Card, 2024). As such, we consider both whether the ΔCFI and the RMSEA fall within the acceptable thresholds to establish evidence of invariance across the nested models discussed above.

Empirical Evidence. Table 2 presents evidence on the measurement invariance of the SRQ-20 across survey rounds and between Indigenous and non-Indigenous mothers. The configural model exhibits strong fit with a CFI of 0.9590, exceeding the 0.95 threshold for a good model fit (Putnick and Bornstein, 2016). The RMSEA of 0.0148 further supports that mothers' responses to the SRQ-20 are well-explained by a common factor structure across groups and over time, lending strong support to the one-factor structure assumed in our model.

To assess measurement invariance, we examine changes in model fit as we impose additional restrictions. The threshold and loading invariant (scalar) model, which assumes loading invariance

[26] The chi-squared test has a high type 1 error rate in contexts with large samples and complex models, potentially leading to false rejections of invariance (Cheung and Rensvold, 2002).

across groups and time, yields a CFI of 0.951, along with a ΔCFI of 0.0083 relative to the configural model, which provides strong evidence of scalar invariance. The threshold, loading and error (strict) invariant model produces a CFI of 0.943, with the additional decrease in CFI (ΔCFI = 0.0081) further supporting strict invariance in this measurement system.[27] The RMSEA remains below 0.017 across all specifications, providing further evidence that even our most restrictive model adequately represents the data. These results establish that the SRQ-20 exhibits both scalar and strict invariance, enabling direct comparisons of maternal mental health across Indigenous and non-Indigenous mothers across the four survey rounds included in our analysis.

These results establish that the SRQ-20 is invariant across groups and over time, enabling meaningful comparisons of maternal mental health between Indigenous and non-Indigenous mothers across survey rounds. The strict invariance of our measurement system implies that differences in observed SRQ-20 responses reflect true variation in the underlying mental health factor rather than differences in how mothers respond to these questions in different groups and over time. Our empirical results thus comprise novel evidence by establishing both longitudinal and cross-group invariance in maternal mental health while drawing on binary measures.[28] This finding is particularly relevant in a context where mental health measures can be sensitive to life events and transitions, demonstrating that the SRQ-20 serves as a robust instrument for examining the evolution of mental health disparities between Indigenous and non-Indigenous mothers over this fifteen-year period.

4.3 Estimated Model Parameters

Given that our measurement system supports strict invariance, we estimate a model where loadings, thresholds, and the variance of the error terms are constrained to be equal across groups and over time. Table A6 presents the estimated factor loadings and thresholds for each measure in the SRQ-20. The factor loadings range from 0.60 to 0.96, indicating that all measures in the SRQ-20 are informative of mothers' underlying mental health, albeit with varying degrees of influence on mothers' latent mental health.

Furthermore, we use the estimated measurement system parameters to compute the signal-to-noise ratio for each underlying continuous variable (X^*_{ijgt}), which captures the share of variance in the latent measure that is attributable to variation in the latent mental health factor.[29] We present these ratios in Table A7. The measures contain substantial signal, with ratios ranging from 0.26 to 0.62 across groups and time periods. Moreover, while these ratios exhibit some variation across survey rounds — with generally higher ratios in later rounds when children are older — the differences are relatively modest, suggesting that the SRQ-20 measures maintain their information content throughout this fifteen-year period.

In Table A8, we present the estimated covariances of error terms across rounds for the same question. The correlations are substantial for several measures, with items related to sleep issues, being

[27] While the χ^2 values are statistically significant, this reflects the test's sensitivity to sample size and model complexity rather than meaningful differences in fit.

[28] Prior research has established measurement invariance of mental health constructs separately across socioeconomic groups (Joshanloo et al., 2013) and over time (Scholte et al., 2011; Moehring et al., 2021).

[29] Despite the equality of loadings and error variances across groups and time, the signal-to-noise ratios can vary with the variance of the latent factor (ϕ_{gt}), allowing us to assess how the information content of measures evolves across groups and over time. The signal-to-noise ratio for measure j in group g and period t is given by $s_{jgt} = \lambda^2_{jgt}\phi_{gt}/(\lambda^2_{jgt}\phi_{gt} + \theta_{jgt})$.

frightened, and having poor appetite exhibiting correlations of 0.2-0.4 across adjacent survey rounds. These patterns highlight the importance of allowing for a flexible structure in the measurement system, as they indicate significant persistence in measurement error at the individual level. A model that instead assumed the item-specific covariances to equal zero across periods would attribute the persistence in response patterns to the underlying latent mental health factor — thus overestimating the correlation in mental health over time.

5 Persistent Inequality in Maternal Mental Health

5.1 Evolution of Maternal Mental Health

Given that our measurement system exhibits strict invariance across groups and over time, we can analyze differences in latent maternal mental health between Indigenous and non-Indigenous mothers as their children age. Table 3 presents the estimated parameters for the latent factors from our measurement system, including the variance-covariance matrix of the factors and their means for each group and time period. At baseline, when children are one year old, Indigenous mothers exhibit substantially worse mental health than their non-Indigenous counterparts, with an average difference of 0.12 standard deviations. This gap in latent mental health is smaller than the 0.15 SD difference in the raw SRQ-20 measures, highlighting the importance of accounting for measurement error when examining mental health disparities.

To further understand the evolution of maternal mental health across the distribution, we simulate mothers' mental health using the estimated parameters of the latent factor distribution.[30] Figure 1 presents the evolution of maternal mental health across survey rounds, showing the mean along with the 10th and 90th percentiles of the distribution for each group. The distribution of mental health at baseline reveals substantial differences across groups. While Indigenous mothers generally exhibit worse mental health throughout the distribution, the gap is particularly pronounced at the bottom, with Indigenous mothers at the 10^{th} percentile scoring 0.25 standard deviations below their non-Indigenous counterparts. On the other hand, the gap at the 90^{th} percentile is close to zero.[31] This pattern aligns with the higher variance in maternal mental health observed among non-Indigenous mothers in the baseline round, as shown in Figure 2.

Figure 1 further shows that non-Indigenous mothers experience substantial improvements in mental health as their children age. Their average mental health increases by 0.68 standard deviations between baseline and Round 2. While mothers' mental health slightly worsens after their children turn five, it remains 0.50 standard deviations higher than at baseline by the time their children turn 15. Indigenous mothers similarly exhibit improvements in mental health over time, with average mental health improving by 0.39 standard deviations between baseline and Round 2. However, these gains are more modest than those experienced by non-Indigenous mothers, as their average mental health in Round 5 is just 0.13 standard deviations higher than at baseline.

[30] We simulate outcomes for 100,000 mothers by drawing from a multivariate normal distribution of the moments of the latent mental health distribution presented in Table 3 (Del Bono et al., 2020; Wiswall and Agostinelli, 2024).

[31] Figure A2 presents additional moments of the distribution, showing that the median exhibits patterns similar to the mean, while the 25th and 75th percentiles closely track the evolution at the 10th and 90th percentiles, respectively.

As a result, mental health disparities between Indigenous and non-Indigenous mothers widen over time, with the gap in average mental health increasing from 0.12 to 0.37 standard deviations between baseline and Round 5. These disparities evolve differently across the distribution. At the 90th percentile, while Indigenous mothers exhibit similar mental health to their non-Indigenous counterparts at baseline, by Round 5 non-Indigenous mothers show substantially better mental health (in the range of 0.5 standard deviations). Conversely, at the bottom of the distribution, the large initial gap at the 10th percentile gradually closes, with both groups showing similar levels of mental health by Round 5.[32]

Altogether, by estimating the model introduced in Section 2 and given that maternal mental health exhibits strict invariance, we are able to make meaningful comparisons of maternal mental health across groups and over time. The results reveal not only systematic disparities between Indigenous and non-Indigenous mothers, but also how these disparities evolve differently across the distribution as children age.

5.2 Persistence in Maternal Mental Health

The variance-covariance matrix presented in Table 3 indicates substantial persistence in maternal mental health over time. In particular, the correlation between latent mental health at baseline and at endline equals 0.34 for non-Indigenous mothers compared to 0.26 for their Indigenous peers (Table A9), thus indicating greater persistence in mental health for non-Indigenous mothers vis-a-vis their Indigenous counterparts that extends through their children's adolescence.

To better understand these patterns of persistence in maternal mental health, we examine how mothers' initial mental health relates to their subsequent outcomes across the distribution. Specifically, we use our simulated dataset to place mothers at their percentile in the round-specific mental health distribution that comprises both Indigenous and non-Indigenous mothers. Figure 3 presents mothers' mental health outcomes at age 5 by their position in the age 1 mental health distribution. Initial mental health exhibits strong persistence across these four years, as a one SD increase in mental health at baseline is associated with a 0.48 SD increase in mental health in Round 2 for non-Indigenous mothers, compared to 0.29 for Indigenous mothers. Furthermore, this persistence in maternal mental health remains even through their children turning 15. Panel B shows that a one standard deviation increase in baseline mental health is associated with a 0.36 SD increase in mental health for non-Indigenous mothers and 0.21 for Indigenous mothers fourteen years later after the initial survey. Altogether, these results indicate stronger persistence in initial mental health for non-Indigenous mothers.[33]

To further understand how early mental health shapes long-term outcomes, we examine how maternal mental health at baseline and by the time their children had turned 5 jointly determine mental health fifteen years later. Figure 4 demonstrates that postpartum mental health significantly influences latent mental health in Round 5, even after controlling for mental health at age 5. For non-Indigenous

[32]These distributional changes are reflected in the evolution of inequality within each group. Figure A3 shows that the 90/10 and 50/10 ratios increase more substantially for Indigenous mothers, particularly after age 5.

[33]Figure A4 shows similar patterns of differential persistence across other periods. Between Rounds 4 and 5, a one SD increase in mental health is associated with increases of 0.55 and 0.19 standard deviations for non-Indigenous and Indigenous mothers, respectively, again reflecting stronger persistence for non-Indigenous mothers.

mothers in the middle decile of the Round 2 mental health distribution, moving from the bottom of the baseline mental health distribution to the top decile increases their position in the age-15 mental health distribution by nearly 30 percentiles. These patterns are similarly present for mothers in the top decile of the Round 2 distribution. Furthermore, initial mental health similarly shapes long-run outcomes for Indigenous mothers (Panel B). Mothers in the middle decile of the Round 2 distribution who move from the bottom to top decile of the baseline mental health distribution improve their position in the age-15 distribution by about 25 percentiles.[34] Altogether, these results indicate the persistent impact of postpartum mental health on long-term maternal mental health which persists even upon controlling for intermediate mental health outcomes.[35]

The persistent influence of postpartum mental health on long-term mental health may emerge through various mechanisms (Beck, 2001; Wang et al., 2011; Shorey et al., 2018). First, the postpartum period represents a critical window during which mothers establish enduring patterns of emotional regulation and stress response, fundamentally shaping their long-term psychological resilience (Baranov et al., 2020). Second, early maternal mental health challenges permanently alter household dynamics and social support structures in ways that subsequent improvements in mental health cannot fully reverse (Burke, 2003; Moore Simas et al., 2019). Third, postpartum mental health may reveal underlying vulnerabilities that shape long-term trajectories of maternal well-being (Vliegen et al., 2014).

Given the persistent influence of postpartum mental health on mothers' long-term outcomes, we additionally examine the relationship with their children's own long-term mental health. In Table A10, we show that mothers' latent mental health at baseline predicts their children's anxiety (GAD-7) and depression (PHQ-8) scores nearly two decades later for both Indigenous and non-Indigenous mothers.[36] All in all, these results remark the important role that postpartum mental health plays in shaping both mothers' and children's long-run mental health outcomes.

6 Inequality in Maternal Mental Health

6.1 Background Characteristics and Latent Mental Health

To understand the drivers of maternal mental health, we first construct factor scores using the estimated measurement system parameters from Section 4. Since our observed mental health measures are binary in nature, we estimate the conditional probability of each response pattern to derive consistent factor scores (Bartholomew et al., 2011). We estimate the factor scores following an Empirical Bayes Modal approach (Skrondal and Rabe-Hesketh, 2004), which combines information from both the distribution of the latent variables and the observed response patterns to estimate the scores, and then adjust these scores to align with the model-implied variance-covariance matrix (Ta-

[34] Figure A5 shows similar patterns when examining mental health outcomes in Round 4, confirming the robustness of these relationships across different time horizons.

[35] This finding parallels evidence from Attanasio et al. (2020b), who show that cognitive and socioemotional skills from two periods prior significantly influence current skill outcomes for children.

[36] Mental health outcomes for children were measured using the Generalized Anxiety Disorder-7 (GAD-7) and Patient Health Questionnaire-8 (PHQ-8) instruments during a supplementary phone survey conducted between August and October 2020 (Favara et al., 2022).

ble A9).[37] Given the invariance of our measurement system, the scored mental health factors are in fact comparable both over time and across groups.

To understand the potential drivers of maternal mental health, we examine the correlation between the scored factors and the main background characteristics in our setting. In Table A11, we show that more educated and higher-wealth mothers tend to have better mental health, while older mothers and those with more children have worse mental health at baseline across both groups. These patterns persist through subsequent periods. These results broadly fit in with existing evidence in the literature (Burke, 2003; Wang et al., 2011; Di Cesare et al., 2013; Aizer et al., 2016; Jackson et al., 2017). Given that Indigenous mothers in our sample tend to be older and have more children by the time the reference child was born (Table 1), these patterns may help explain the mental health gaps documented in Section 5.

To understand the relative importance of these variables in shaping mothers' baseline mental health, we estimate regressions of mental health in the first round on background characteristics separately for Indigenous and non-Indigenous mothers. The first two columns of Table 4 show that mothers with fewer children at baseline and those with a father present in the household tend to have better mental health, even upon controlling for other background characteristics. To examine how these relationships evolve over time, we estimate similar specifications for the following survey rounds, and include lagged mental health as an additional variable to account for persistence. Altogether, the results reveal substantial persistence in mental health across rounds, particularly for non-Indigenous mothers.[38] While most baseline characteristics become less predictive of mental health over time, the number of baseline children is negatively associated with maternal mental health even twelve years after the birth of the reference child.[39]

6.2 Drivers of Inequality in Maternal Mental Health

Methodological Approach. Given the significant gaps in latent mental health between Indigenous and non-Indigenous women that emerge across the distribution and persist across multiple periods, we examine whether differences in background characteristics may explain these patterns. To this end, we employ a RIF regression approach that generalizes the Oaxaca–Blinder decomposition to distributional statistics (Firpo et al., 2009, 2018). For a specific quantile, $v(\cdot)$, we define a counterfactual distribution F_Y^C by reweighting one group's characteristics so that they resemble the other group's distribution (DiNardo et al., 1996) in:

$$F_Y^C = F_{Y|\mathbf{X},G=0}\,\omega(\mathbf{X})\,dF_{\mathbf{X}|G=1} \simeq \int F_{Y|\mathbf{X},G=0}\,\omega(\mathbf{X})\,dF_{\mathbf{X}|=G=0}$$

where the reweighting factor ($\omega(\mathbf{X})$) is given by: $\omega(\mathbf{X}) = \frac{1-P}{P}\frac{P(G=1|\mathbf{X})}{1-P(G=1|\mathbf{X})}$, where P is the overall share of the population in group $G=1$. Upon recovering F_Y^C, we construct the counterfactual quantile:

$$v_c = \mathbb{E}\left[\mathrm{RIF}\{y, v(F_Y^C)\}\right] = \mathbf{X}^{c\prime}\hat{\beta}_c$$

[37] To account for uncertainty in the estimation procedure, we bootstrap the model 1,000 times and calculate standard errors based on these replications (Attanasio et al., 2020c).

[38] Results remain similar when including an expanded set of baseline covariates (Table A12).

[39] Note that we restrict our analysis to background characteristics measured at baseline to avoid including variables measured in later survey rounds (i.e. household wealth) that could be influenced by mothers' mental health at baseline.

where RIF$\{\cdot\}$ is the recentered influence function for $v(\cdot)$, and $\mathbf{X}^c$ and $\hat{\beta}_c$ denote the characteristics and coefficients under the counterfactual quantile v, respectively. The overall gap in quantile v between the two groups is given by:

$$\Delta v = \underbrace{\mathbf{X}^{1\prime}(\hat{\beta}_1 - \hat{\beta}_c)}_{\Delta v_S^p} + \underbrace{(\mathbf{X}^1 - \mathbf{X}^c)'\hat{\beta}_c}_{\Delta v_S^e} + \underbrace{(\mathbf{X}^c - \mathbf{X}^0)'\hat{\beta}_0}_{\Delta v_X^p} + \underbrace{\mathbf{X}^{c\prime}(\hat{\beta}_c - \hat{\beta}_0)}_{\Delta v_X^e}.$$

where $\Delta v_S^p + \Delta v_S^e$ capture the 'coefficient effect' (returns to characteristics), while $\Delta v_X^p + \Delta v_X^e$ represent the 'composition effect' (differences in characteristics). By regressing the RIF on the full set of covariates ($\mathbf{X}$) for each group, we can measure how differences in characteristics and their returns contribute to the gap at each quantile of the distribution of latent mental health.[40]

Empirical Results. We estimate the RIF decomposition to understand the extent to which mothers' socioeconomic characteristics contribute to the inequality in the distribution of maternal mental health in the baseline survey round. Figure 5 presents the gap in mental health across various percentiles of the distribution, and the extent to which these differences are explained by mothers' educational attainment, age, wealth, and their family composition. We find, for instance, that the gap in mental health at the 10^{th} percentile is fully explained by differences in the characteristics of Indigenous and non-Indigenous mothers, with mother's education and the number of children playing a significant role.[41] While at the upper end of the distribution, the gaps in latent mental health are close to zero, we still find that closing gaps in socioeconomic characteristics would result in a relative improvement of Indigenous' mothers mental health at the median and through the 90^{th} percentile. These findings suggest that reducing socioeconomic disparities could substantially close the inequality in postpartum mental health between Indigenous and non-Indigenous mothers.

We next examine how these distributional gaps evolve over time by estimating the RIF decomposition to subsequent survey rounds. Given the strong persistence in maternal mental health documented in Section 5.2, we include one-period lagged mental health as an explanatory variable to further capture the role of prior mental health status beyond background characteristics in driving persistent inequalities. In Figure 6, we show that postpartum mental health explains between 15–30% of the total gap across the distribution by the time their children then reach age five. At the same time, we find that the overall share of the gap that is explained by background characteristics is significantly smaller in Round 2 than at baseline. In particular, while wealth differences play a particularly important role at the lower tail, explaining about 40% of the total gap at the 10^{th} percentile, differences in mothers' educational attainment play a minimal role in explaining the distributional gaps at this age.[42]

By Round 5, when children have reached age 15, the picture gradually changes. As shown in the second panel of Figure 6 (Panel B), not only do the total gaps widen above the median, but lagged mental health becomes a more important explanatory factor. Specifically, lagged mental health accounts for a larger share of the total gap across the distribution, underscoring how early

[40] We could alternatively estimate the DiNardo et al. (1996) decomposition, yet this flexible estimation approach typically requires a larger sample size than what is available in this context.

[41] Table A13 presents the contribution of each covariate in explaining disparities across the distribution.

[42] The complete set of RIF regression estimates for Rounds 2, 4 and 5 are presented in Appendix Table A14.

disparities become entrenched over the long run.[43] Meanwhile, the direct contribution of socioeconomic characteristics becomes less significant, with mother's education and baseline wealth showing mostly insignificant effects by age 1 — suggesting their influence may operate primarily through their impact on earlier mental health. The large unexplained component, particularly at the top of the distribution, thus underscores the cumulative advantages in non-Indigenous mothers' mental health trajectories that are not fully explained by the set of baseline socioeconomic characteristics included in the decomposition.

All in all, these findings highlight how early inequalities in maternal mental health can have persistent impacts. While baseline disparities are largely explained by differences in mother's education and family size, the growing importance of lagged mental health in explaining later-life gaps points to a process of cumulative disadvantage. Combined with the differential persistence patterns documented earlier, this suggests that early-life maternal mental health is a key mechanism through which initial socioeconomic disparities translate into long-term inequalities between Indigenous and non-Indigenous mothers.

7 Conclusion

In this paper, we have introduced a measurement framework that establishes the conditions under which binary mental health measures can be validly compared across populations and over time. We implement this framework using longitudinal data from Peru, where we observe mothers' mental health from their children's infancy through adolescence. Our measurement model reveals that the Self-Reporting Questionnaire (SRQ-20) exhibits strict invariance between Indigenous and non-Indigenous mothers over this fifteen-year period, indicating that observed disparities reflect true differences in mothers' latent mental health rather than systematic differences in how these measures are interpreted. Using this result, we demonstrate that while both groups experience improvements in mental health after the postpartum period, these gains are substantially larger for non-Indigenous mothers, leading to a widening gap in mental health that persists through their children's adolescence.

Altogether, our results remark how socioeconomic disparities shape the evolution of mental health inequality. We find that differences in mothers' education and family composition explain most of the initial gap in postpartum mental health between Indigenous and non-Indigenous mothers. However, as children age, we show that early mental health status becomes increasingly important in explaining these disparities, pointing to a process of cumulative disadvantage that operates through the strong persistence in maternal mental health trajectories. These results suggest that targeted interventions aimed at improving postpartum mental health for mothers in disadvantaged groups could have significant impacts on their long-term mental health outcomes.

[43] The RIF decomposition results for Round 4, presented in Figure A6, show that lagged mental health still plays an important role in shaping inequalities across the mental health distribution.

References

Aizer, Anna, Laura Stroud, and Stephen Buka (2016) 'Maternal stress and child outcomes: Evidence from siblings.' *Journal of Human Resources* 51(3), 523–555

Arteaga, Irma, and Paul Glewwe (2019) 'Do community factors matter? an analysis of the achievement gap between indigenous and non-indigenous children in peru.' *International Journal of Educational Development* 65, 80–91

Attanasio, Orazio, Áureo De Paula, and Alessandro Toppeta (2024) 'Intergenerational Mobility in Socio-emotional Skills.' Technical Report w27823, National Bureau of Economic Research, Cambridge, MA, June

Attanasio, Orazio, Costas Meghir, and Emily Nix (2020a) 'Human Capital Development and Parental Investment in India.' *The Review of Economic Studies* 87(6), 2511–2541

Attanasio, Orazio, Costas Meghir, Emily Nix, and Francesca Salvati (2017) 'Human capital growth and poverty: Evidence from Ethiopia and Peru.' *Special Issue on Human Capital and Inequality* 25, 234–259

Attanasio, Orazio, Raquel Bernal, Michele Giannola, and Milagros Nores (2020b) 'Child development in the early years: Parental investment and the changing dynamics of different dimensions.' Technical Report, National Bureau of Economic Research

Attanasio, Orazio, Richard Blundell, Gabriella Conti, and Giacomo Mason (2020c) 'Inequality in socio-emotional skills: A cross-cohort comparison.' *Journal of Public Economics* 191, 104171

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko (2020) 'Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial.' *American economic review* 110(3), 824–859

Bartholomew, David J, Martin Knott, and Irini Moustaki (2011) *Latent variable models and factor analysis: A unified approach* (John Wiley & Sons)

Beck, Cheryl Tatano (2001) 'Predictors of postpartum depression: An update.' *Nursing Research* 50, 275–285

Beusenberg, M, John H Orley, and World Health Organization. Division of Mental Health (1994) 'A User's guide to the self reporting questionnaire'

Burke, L. (2003) 'The impact of maternal depression on familial relationships.' *International Review of Psychiatry* 15(3), 243–255

Carneiro, Pedro, Karsten T Hansen, and James J Heckman (2003) '2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice.' *International Economic Review* 44(2), 361–422

Chen, Fang Fang (2007) 'Sensitivity of goodness of fit indexes to lack of measurement invariance.' *Structural equation modeling: a multidisciplinary journal* 14(3), 464–504

Cheung, Gordon W, and Roger B Rensvold (2002) 'Evaluating goodness-of-fit indexes for testing measurement invariance.' *Structural equation modeling* 9(2), 233–255

Cunha, Flavio, and James J Heckman (2008) 'Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation.' *Journal of human resources* 43(4), 738–782

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010) 'Estimating the technology of cognitive and noncognitive skill formation.' *Econometrica* 78(3), 883–931

Del Bono, Emilia, Josh Kinsler, and Ronni Pavan (2020) 'Skill formation and the trouble with child non-cognitive skill measures.' Technical Report, IZA Discussion Papers

Di Cesare, Mariachiara, Ricardo Sabates, and Keith M Lewin (2013) 'A double prevention: How maternal education can affect maternal mental health, child health and child cognitive development.' *Longitudinal and Life Course Studies* 4(3), 166–179

DiNardo, John, Nicole M Fortin, and Thomas Lemieux (1996) 'Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach.' *Econometrica* 64(5), 1001

Escobal, Javier (2012) *Multidimensional poverty and inequality of opportunity in Peru: taking advantage of the longitudinal dimension of Young Lives* (Young Lives)

Escobal, Javier, and Eva Flores (2008) 'An assessment of the young lives sampling approach in peru'

Favara, Marta, Richard Freund, Catherine Porter, Alan Sanchez, and Douglas Scott (2022) 'Young lives, interrupted: short-term effects of the covid-19 pandemic on adolescents in low-and middle-income countries.' *The Journal of Development Studies* 58(6), 1063–1080

Finch, W Holmes (2023) 'A comparison of methods for determining the number of factors to retain with exploratory factor analysis of dichotomous data.' *Psych* 5(3), 1004–1018

Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux (2009) 'Unconditional quantile regressions.' *Econometrica* 77(3), 953–973

Firpo, Sergio P, Nicole M Fortin, and Thomas Lemieux (2018) 'Decomposing wage distributions using recentered influence function regressions.' *Econometrics* 6(2), 28

Freire, German Nicolas, Steven Daniel Schwartz Orellana, Melissa Zumaeta Aurazo, Damasceno Costa Costa, Jonna Maria Lundvall, Martha Celmira Viveros Mendoza, Leonardo Ramiro Lucchetti, Laura Moreno, and Liliana Do Couto Sousa (2015) 'Indigenous latin america in the twenty-first century: the first decade.' Technical Report, The World Bank

Goldberg, David, and Paul Williams (1988) *A User's Guide to the General Health Questionnaire* (Windsor, UK: NFER-Nelson)

Harpham, Trudy, Michael Reichenheim, Rebecca Oser, Elizabeth Thomas, Narmeen Hamid, Surinder Jaswal, Ana Ludermir, and Magna Aidoo (2003) 'Measuring mental health in a cost-effective manner.' *Health policy and planning* 18(3), 344–349

Jackson, Margot I, Kathleen Kiernan, and Sara McLanahan (2017) 'Maternal education, changing family circumstances, and children's skill development in the United States and UK.' *The ANNALS of the American Academy of Political and Social Science* 674(1), 59–84

Joshanloo, Mohsen, Marié P. Wissing, Itumeleng P. Khumalo, and Sanne M.A. Lamers (2013) 'Measurement invariance of the Mental Health Continuum-Short Form (MHC-SF) across three cultural groups.' *Personality and Individual Differences* 55(7), 755–759

Kessler, Ronald C., Gavin Andrews, Lisa J. Colpe, Eva Hiripi, Daniel K. Mroczek, Sharon-Lise T. Normand, Ellen E. Walters, and Alan M. Zaslavsky (2002) 'Short screening scales to monitor population prevalences and trends in non-specific psychological distress.' *Psychological Medicine* 32(6), 959–976

Little, Todd D., and Noel A. Card (2024) *Longitudinal Structural Equation Modeling* Methodology in the Social Sciences, second edition ed. (New York, NY: The Guilford Press, a division of Guildford Publications, Inc.)

Liu, Yu, Roger E Millsap, Stephen G West, Jenn-Yun Tein, Rika Tanaka, and Kevin J Grimm (2017) 'Testing measurement invariance in longitudinal data with ordered-categorical measures.' *Psychological methods* 22(3), 486

Lopez, Jennifer, Jere Behrman, Santiago Cueto, Marta Favara, and Alan Sánchez (2024) 'Late-childhood foundational cognitive skills predict educational outcomes through adolescence and into young adulthood: Evidence from Ethiopia and Peru.' *Economics of Education Review* 100, 102543

Miller, Laura J (2002) 'Postpartum depression.' *Jama* 287(6), 762–765

Millsap, RE, H Cham, B Laursen, TD Little, and NA Card (2012) 'Handbook of developmental research methods.' *Investigating factorial invariance in longitudinal data* pp. 109–127

Millsap, Roger E, and Jenn Yun-Tein (2004) 'Assessing factorial invariance in ordered-categorical measures.' *Multivariate behavioral research* 39(3), 479–515

Moehring, Anne, Diana Guertler, Kristian Krause, Gallus Bischof, Hans-Juergen Rumpf, Anil Batra, Susanne Wurm, Ulrich John, and Christian Meyer (2021) 'Longitudinal measurement invariance of the patient health questionnaire in a German sample.' *BMC Psychiatry* 21(1), 386

Moore Simas, Tiffany A., Ming-Yi Huang, Cody Patton, Marcia Reinhart, Anita J. Chawla, Christine Clemson, and Adi Eldar-Lissai (2019) 'The humanistic burden of postpartum depression: A systematic literature review.' *Current Medical Research and Opinion* 35(3), 383–393

Muthén, Bengt (1984) 'A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators.' *Psychometrika* 49(1), 115–132

Netsereab, Tesfit Brhane, Meron Mehari Kifle, Robel Berhane Tesfagiorgis, Sara Ghebremichael Habteab, Yosan Kahsay Weldeabzgi, and Okbazghi Zaid Tesfamariam (2018) 'Validation of the WHO self-reporting questionnaire-20 (SRQ-20) item in primary health care settings in Eritrea.' *International journal of mental health systems* 12, 1–9

O'hara, Michael W, and Annette M Swain (1996) 'Rates and risk of postpartum depression—a meta-analysis.' *International review of psychiatry* 8(1), 37–54

Posso, Alberto (2023) 'Bilingual education and child labor: Lessons from peru.' *Journal of Economic Behavior & Organization* 212, 840–872

Putnick, Diane L., and Marc H. Bornstein (2016) 'Measurement invariance conventions and reporting: The state of the art and future directions for psychological research.' *Developmental Review* 41, 71–90

Rosseel, Yves (2012) '**Lavaan** : An *R* Package for Structural Equation Modeling.' *Journal of Statistical Software*

Sánchez, Alan, and Javier Escobal (2020) 'Survey attrition after 15 years of tracking children in four developing countries: The young lives study.' *Review of Development Economics* 24(4), 1196–1216

Sánchez, Alan, Guido Meléndez, and Jere r Behrman (2020) 'Impact of the juntos conditional cash transfer program on nutritional and cognitive outcomes in peru: comparison between younger and older initial exposure.' *Economic Development and Cultural Change* 68(3), 865–897

Schmied, Virginia, Maree Johnson, Norell Naidoo, Marie-Paule Austin, Stephen Matthey, Lynn Kemp, Annie Mills, Tanya Meade, and Anthony Yeo (2013) 'Maternal mental health in Australia and New Zealand: A review of longitudinal studies.' *Women and Birth* 26(3), 167–178

Scholte, Willem F, Femke Verduin, Anouk Van Lammeren, Theoneste Rutayisire, and Astrid M Kamperman (2011) 'Psychometric properties and longitudinal validation of the self-reporting questionnaire (SRQ-20) in a Rwandan community setting: A validation study.' *BMC Medical Research Methodology* 11(1), 116

Sevim, Dilek, Victoria Baranov, Sonia Bhalotra, Joanna Maselko, and Pietro Biroli (2024) 'Trajectories of Early Childhood Skill Development and Maternal Mental Health.' *Journal of Human Resources* 59(S), S365–S401

Shorey, Shefaly, Cornelia Yin Ing Chee, Esperanza Debby Ng, Yiong Huak Chan, Wilson Wai San Tam, and Yap Seng Chong (2018) 'Prevalence and incidence of postpartum depression among healthy mothers: A systematic review and meta-analysis.' *Journal of psychiatric research* 104, 235–248

Singh, Abhijeet (2020) 'Learning more with every year: School year productivity and international learning divergence.' *Journal of the European Economic Association* 18(4), 1770–1813

Skrondal, Anders, and Sophia Rabe-Hesketh (2004) *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models* (Chapman and Hall/CRC)

Toppeta, Alessandro (2022) 'Skill formation with siblings.' Technical Report

Tuan, Tran, Trudy Harpham, and Nguyen Tuo Huong (2004) 'Validity and reliability of the self-reporting questionnaire 20 items in Vietnam.' *Hong Kong Journal of Psychiatry* 14(3), 15

van der Westhuizen, Claire, Gail Wyatt, John K Williams, Dan J Stein, and Katherine Sorsdahl (2016) 'Validation of the self reporting questionnaire 20-item (srq-20) for use in a low-and middle-income country emergency centre setting.' *International journal of mental health and addiction* 14, 37–48

Vliegen, Nicole, Sara Casalin, and Patrick Luyten (2014) 'The Course of Postpartum Depression: A Review of Longitudinal Studies.' *Harvard Review of Psychiatry* 22(1), 1–22

Wang, Liang, Tiejian Wu, James L. Anderson, and James E. Florence (2011) 'Prevalence and Risk Factors of Maternal Depression During the First Three Years of Child Rearing.' *Journal of Women's Health* 20(5), 711–718

Wiswall, Matthew, and Francesco Agostinelli (2024) 'Estimating the Technology of Children's Skill Formation.' *Journal of Political Economy* p. 733396

Woody, CA, AJ Ferrari, DJ Siskind, HA Whiteford, and MG Harris (2017) 'A systematic review and meta-regression of the prevalence and incidence of perinatal depression.' *Journal of affective disorders* 219, 86–92

Wu, Hao, and Ryne Estabrook (2016) 'Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes.' *Psychometrika* 81(4), 1014–1045

8 Tables and Figures

Table 1: Summary Statistics

Panel A: Background Characteristics	Overall Mean	Non-Ind. Mean	Ind. Mean
High School Complete	0.32	0.46	0.17
Mother's Education	6.83	8.66	5.01
Mother's Age	27.40	26.78	28.01
Mother's Height	149.68	150.79	148.57
Mother's Weight	55.42	56.74	54.11
Wealth Index	0.41	0.49	0.34
Father in Household	0.87	0.86	0.89
Number of Children	1.64	1.30	2.32
Panel B: Mean SRQ Values	Overall Mean	Non-Ind. Mean	Ind. Mean
Child Age 1	5.55	5.34	5.95
Child Age 5	3.53	3.22	4.12
Child Age 12	4.27	3.87	5.05
Child Age 15	4.20	3.79	5.01
Observations	1270	841	429

Note: The table presents background characteristics (Panel A) and mean SRQ values (Panel B) for the sample. Panel A provides descriptive statistics for time-invariant variables. The time invariant variables are taken at baseline, when the children are 1 years old. The wealth index is a score based on housing quality, consumer durables, and access to services, its values range from 0 to 1. Height and weight are in centimeters and kilograms, respectively. Education is measured in years. Father in household is a binary variable indicating whether the father is present in the household. Number of children is the total number of children in the household. Panel B shows the mean SRQ scores across rounds, with the overall mean presented first, followed by disaggregated means for non-Indigenous and Indigenous groups.

Table 2: Measurement Invariance Testing

Model	χ^2	CFI	RMSEA	SRMR	$\Delta\chi^2$	Δ CFI	Δ RMSEA	Δ SRMR
Threshold	6731.16	0.9590	0.0148	0.0883				
Thr. + Loading	7025.02	0.9507	0.0161	0.0896	293.85	0.0083	-0.0013	-0.0013
Thr. + Load. + Error	7328.57	0.9426	0.0172	0.0942	303.55	0.0081	-0.0011	-0.0046

Note: The table presents the χ^2 test, Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR) for the configural (Thresholds), scalar (Thresholds + Loadings), and strict (Thresholds + Loadings + Errors) models. The model is estimated using the mean and variance-adjusted Diagonally Weighted Least Squares (DWLS) estimator.

Table 3: Variance-Covariance Matrix and Means for Maternal Mental Health Latent Factor

Mental Health	Age 1	Age 5	Age 12	Age 15
		Non-Indigenous		
Variance-Covariance Matrix				
Age 1	1.00			
Age 5	0.52	1.30		
Age 12	0.55	0.55	1.44	
Age 15	0.45	0.55	0.83	1.75
Mean (κ)	0.00	0.78	0.56	0.66
		Indigenous		
Variance-Covariance Matrix				
Age 1	1.32			
Age 5	0.42	1.28		
Age 12	0.64	0.42	1.84	
Age 15	0.33	0.35	0.37	1.24
Mean (κ)	-0.12	0.44	0.23	0.14

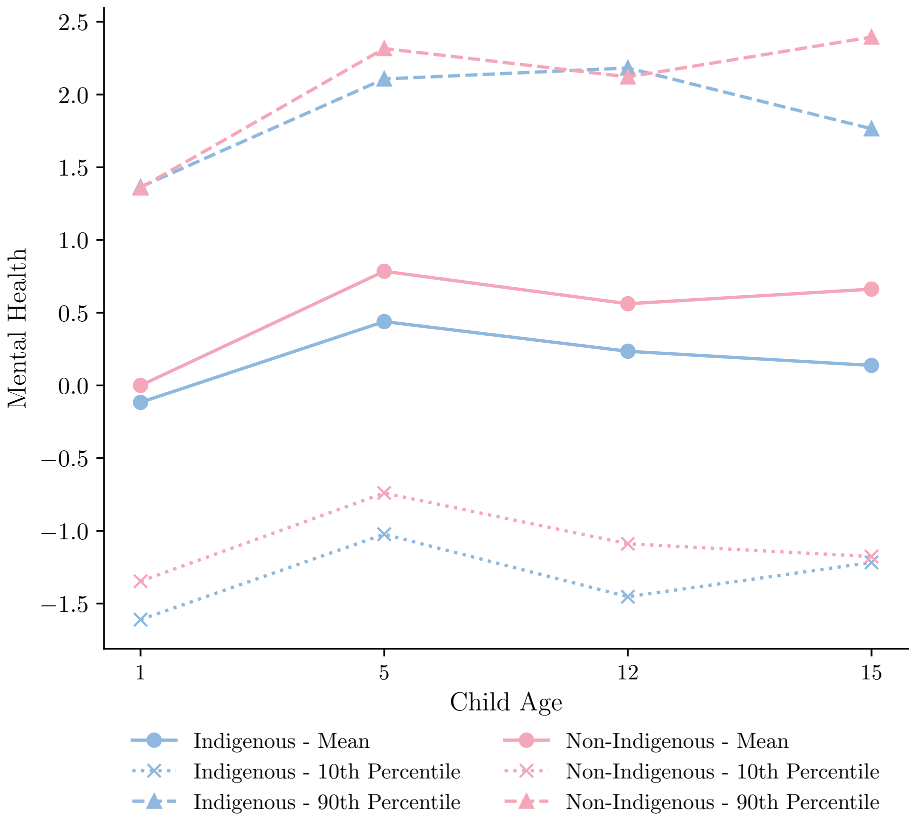
Note: The table presents the variance-covariance matrix and means for the maternal mental health latent factor by non-Indigenous and Indigenous groups. Age refers to the child's age at the time of the survey. The CFA model is estimated using mean- and variance-adjusted DWLS. Higher values indicate better mental health.

Table 4: Regression Analysis of Maternal Mental Health Factor Scores

	Round 1		Round 2		Round 4		Round 5	
	Non-Ind.	Ind.	Non-Ind.	Ind.	Non-Ind.	Ind.	Non-Ind.	Ind.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's	0.029**	-0.002	-0.009	-0.002	0.025**	-0.007	0.007	0.006
Education	(0.012)	(0.015)	(0.011)	(0.012)	(0.010)	(0.015)	(0.010)	(0.013)
Wealth	-0.011	0.040	0.031	0.026	-0.083**	0.029	-0.030	0.001
Index	(0.042)	(0.073)	(0.039)	(0.060)	(0.037)	(0.073)	(0.038)	(0.056)
Mother's	-0.002	0.002	-0.013*	0.007	0.009	-0.012	-0.011	-0.006
Age	(0.007)	(0.012)	(0.007)	(0.010)	(0.007)	(0.011)	(0.007)	(0.009)
Num. of	-0.078***	-0.094***	-0.021	-0.099***	-0.028	-0.032	0.047*	-0.039
Children	(0.030)	(0.033)	(0.030)	(0.032)	(0.032)	(0.034)	(0.028)	(0.030)
Father in	0.282***	0.374**	-0.009	0.170	-0.067	0.158	-0.098	0.038
Household	(0.096)	(0.160)	(0.092)	(0.134)	(0.095)	(0.172)	(0.094)	(0.126)
Lagged			0.470***	0.251***	0.387***	0.269***	0.568***	0.175***
Mental Health			(0.032)	(0.041)	(0.031)	(0.053)	(0.034)	(0.042)
R^2	0.0438	0.0340	0.2178	0.1323	0.1691	0.0811	0.2708	0.0689
Observations	841	429	841	429	841	429	841	429

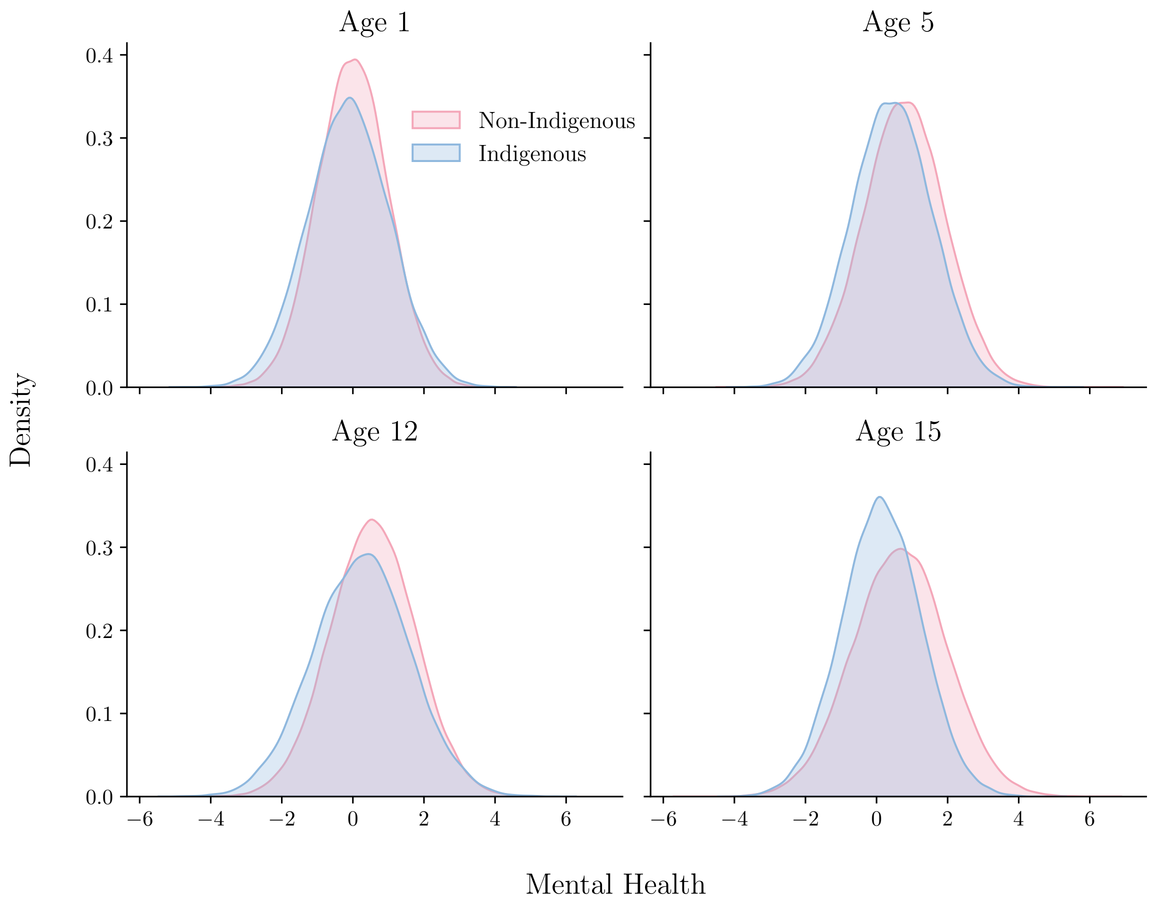
Note: The table presents the results of the regression analysis for maternal mental health factor scores at different survey rounds. The table shows the coefficients and standard errors for the non-Indigenous and Indigenous groups. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Figure 1: Evolution of Maternal Mental Health Latent Factor



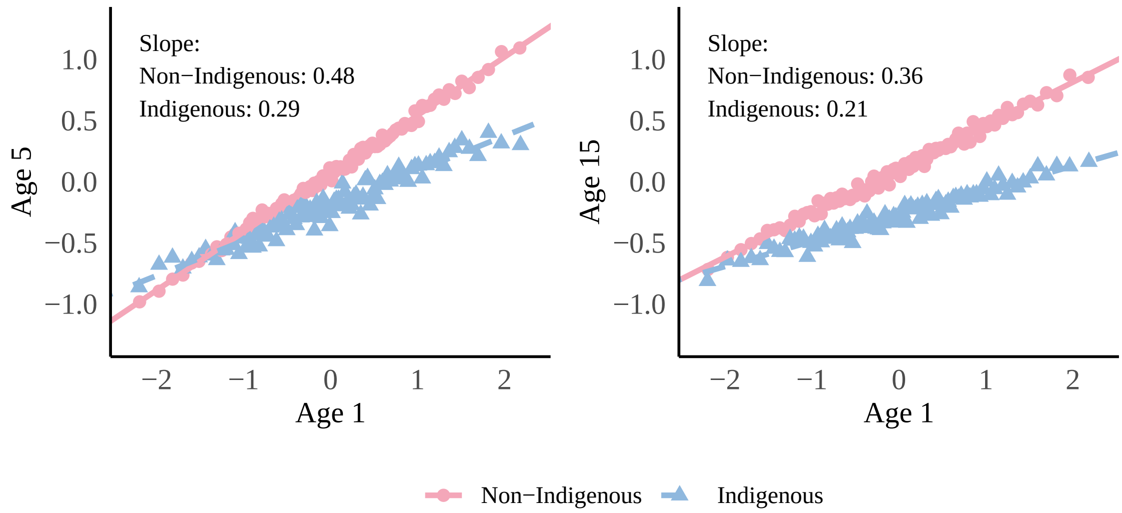
Note: The plot shows the 90th, mean, and 10th percentile of the distribution of maternal mental health latent factor by Indigenous status. Age represents the age of the child at the time of the survey. The model is estimated using mean- and variance-adjusted DWLS. Higher values represent *better* mental health.

Figure 2: Density of Mental Health Latent Factor by Indigenous Status



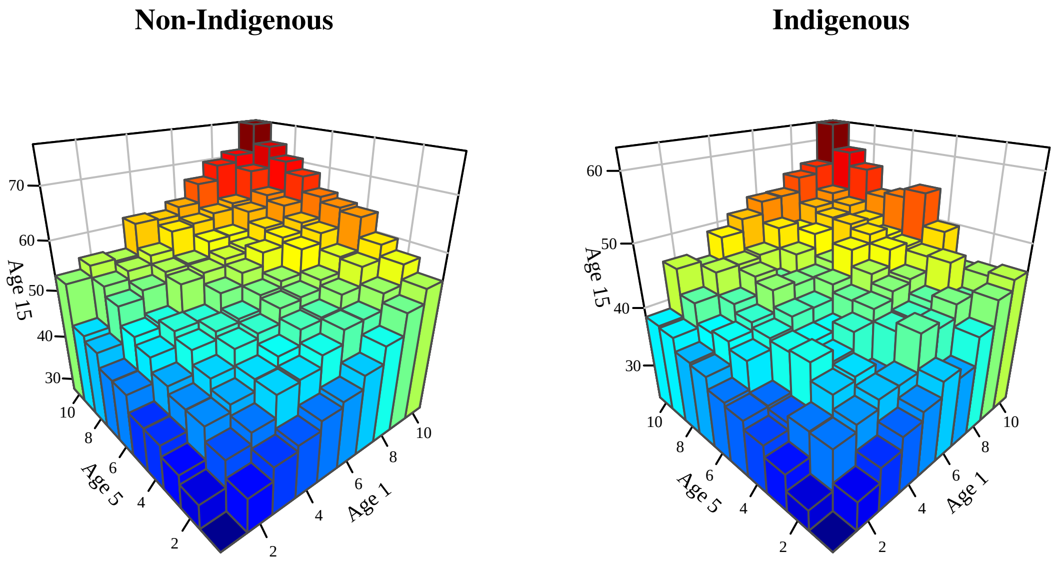
Note: The plot shows the density of the maternal mental health latent factor by Indigenous status. Age represents the age of the child at the time of the survey. The model is estimated using mean- and variance-adjusted DWLS.

Figure 3: Persistence of Maternal Mental Health Latent Factor



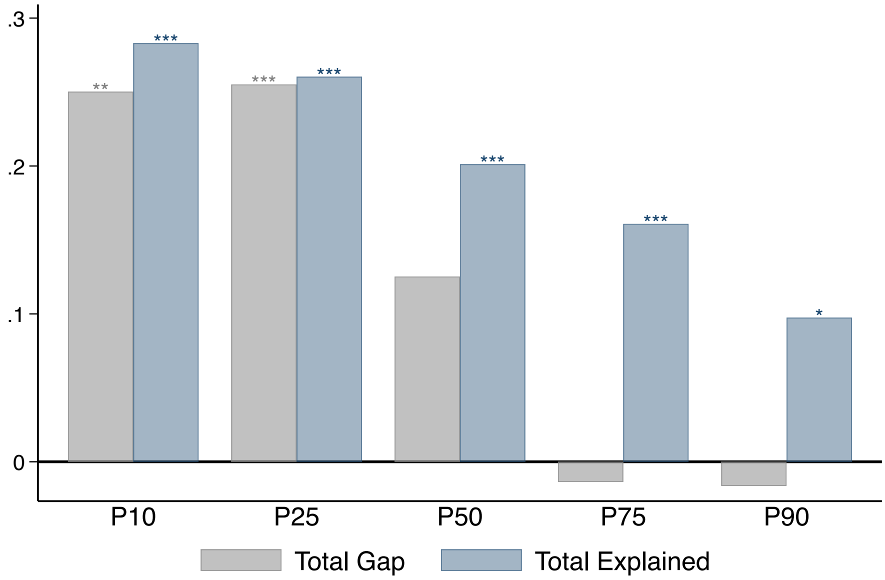
Note: The plot shows latent mental health standardized at age 15 versus standardized latent mental health at age 1, separated by Indigenous status. We plot 100 binned averages (1 percentile), and the lines are linear fits for each group. The latent mental health factor is estimated using mean- and variance-adjusted DWLS. The slope is the persistence of latent mental health from age 1 to age 5 for the left figure and from age 1 to age 15 for the right figure.

Figure 4: Joint Distribution of Maternal Mental Health Latent Factor at Ages 1, 5, and 15



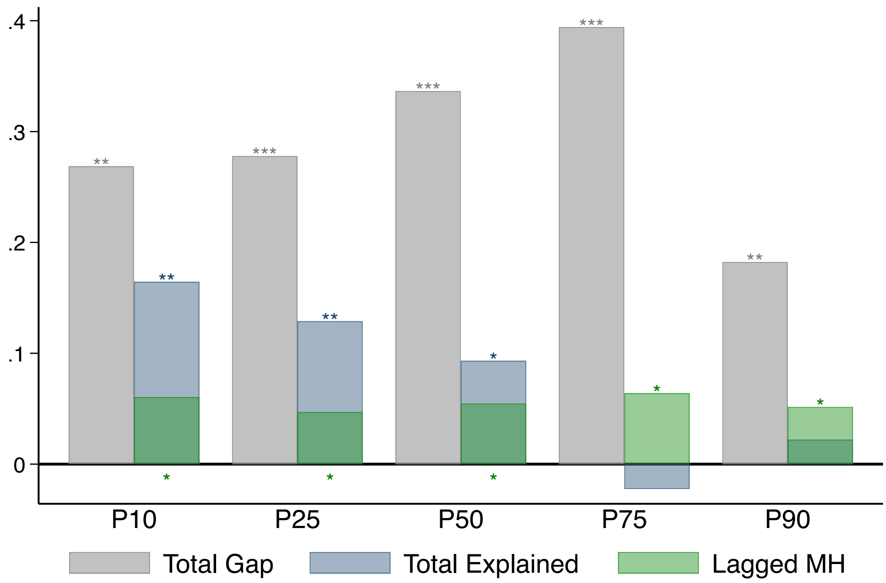
Note: The plot shows the joint distribution of maternal mental health latent factor at ages 1, 5, and 15, separated by Indigenous status. The model is estimated using mean- and variance-adjusted DWLS. Higher values represent *better* mental health.

Figure 5: RIF Decomposition of Gaps in Maternal Mental Health (Round 1)

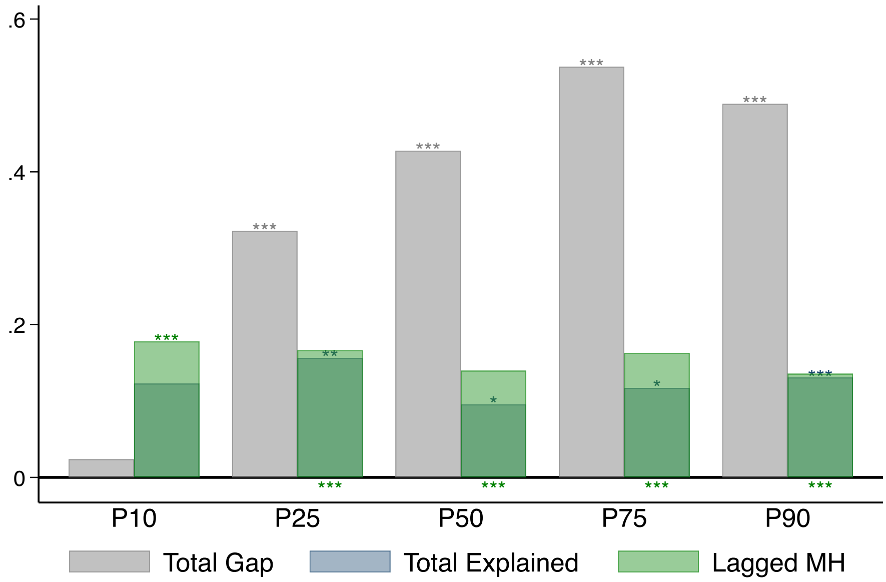


The plot shows the Recentered Influence Function (RIF) decomposition of the gaps in maternal mental health factor scores at the 10th, 25th, 50th, 75th, and 90th percentiles for Round 1. The total gap is the difference in mental health factor scores between Indigenous and non-Indigenous mothers. The explained component is the part of the gap explained by the covariates, and the unexplained component is the part of the gap not explained by the covariates. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health. *p<0.1, **p<0.05, ***p<0.01.

Figure 6: RIF Decomposition of Gaps in Maternal Mental Health (Rounds 2 and 5)



(a) RIF Decomposition: Round 2



(b) RIF Decomposition: Round 5

The plot shows the Recentered Influence Function (RIF) decomposition of the gaps in maternal mental health factor scores at the 10th, 25th, 50th, 75th, and 90th percentiles for (a) Round 2 and (b) Round 5. The total gap is the difference in mental health factor scores between Indigenous and non-Indigenous mothers. The explained component is the part of the gap explained by the covariates, and the unexplained component is the part of the gap not explained by the covariates. The green bar explains the share of the gap at different percentiles explained by lagged maternal mental health. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health. *p<0.1, **p<0.05, ***p<0.01.

A Appendix

Table A1: SRQ-20 Items Measuring Maternal Mental Health Problems within the Last 30 Days.

Title	Questions
Unhappy	Did you feel unhappy?
Poor Appetite	Was your appetite poor?
Sleep	Did you sleep badly?
Frightened	Were you easily frightened?
Nervous/Tense	Did you feel nervous, tense or worried?
Digestion	Was your digestion poor?
Thinking	Did you have trouble thinking clearly?
Cry	Did you cry more than usual?
Enjoyment	Did you find it difficult to enjoy your daily activities?
Work	Did your daily work suffer?
Usefulness	Were you unable to play a useful part in life?
Lost Interest	Did you lose interest in things?
Worthless	Did you feel you were a worthless person?
Giving Up	Were things so bad that you felt that you just couldn't go on?
Always Tired	Did you feel tired all the time?
Stomach Issues	Did you have uncomfortable feelings in your stomach?
Headache	Did you often have headaches?
Handshake	Did your hands shake?
Easily Tired	Were you easily tired?
Decision-Making	Did you find it difficult to make decisions?

Note: The SRQ-20 is a self-report questionnaire that measures mental health problems. The questionnaire consists of 20 items that ask about mental health problems within the last 30 days. The items are dichotomous (yes/no) and are used to measure the presence of mental health problems.

Table A2: Proportion of Mothers Reporting Mental Health Problems by Round and Indigenous Status

	Indigenous				Non-Indigenous			
Measure	Age 1	Age 5	Age 12	Age 15	Age 1	Age 5	Age 12	Age 15
Unhappy	52.4	43.1	53.6	52.4	48.8	36.1	39.4	38.3
Poor Appetite	22.1	22.4	25.9	26.1	23.1	16.8	18.4	14.7
Sleep	17.0	20.3	28.9	28.7	16.8	18.3	25.1	25.9
Frightened	51.0	34.7	33.8	34.3	42.3	22.5	25.1	24.7
Tense	62.2	66.2	64.6	66.0	64.9	60.5	59.0	57.4
Digestion	21.4	16.1	19.1	14.2	16.2	9.5	12.7	14.4
Thinking	28.0	11.9	17.7	15.2	24.7	8.0	13.3	11.1
Crying	29.6	19.1	27.7	30.3	22.8	14.3	15.7	19.1
Enjoyment	27.7	9.8	16.3	20.0	27.3	7.5	14.3	15.3
Work	18.6	9.3	16.1	12.1	21.3	4.5	9.0	8.8
Usefulness	16.8	5.4	11.7	8.4	17.1	4.4	8.8	5.9
Lost Interest	15.6	5.8	12.1	10.3	14.4	4.4	8.9	7.6
Worthless	15.2	8.6	9.6	6.5	11.3	4.3	4.9	4.3
Giving Up	6.1	4.7	7.5	6.3	3.0	3.4	3.3	4.5
Always Tired	31.2	16.3	24.2	31.0	23.4	10.8	19.4	19.1
Stomach Issues	24.7	23.8	21.9	17.0	27.9	18.1	21.2	16.4
Headache	48.7	48.0	45.9	54.3	41.5	44.5	37.6	41.1
Handshake	19.6	12.4	12.4	14.9	10.5	7.6	6.4	6.4
Easily Tired	44.8	22.1	30.3	35.2	29.1	13.6	20.3	21.8
Decision-Making	42.0	12.4	25.9	17.9	47.9	13.3	24.6	21.8

Note: The table shows the proportion of mothers responding "yes" to the mental health problems by round and indigenous status. The SRQ-20 items are dichotomous (yes/no) and are used to measure the presence of mental health problems.

Table A3: Non-Indigenous Mothers Correlation Matrix for SRQ-20 Mental Health Measures

Variable	Unhp	App	Slp	Frgt	Tns	Dgst	Thnk	Cry	Enjy	Wrk	Usfl	Intr	Wrth	GvUp	Tird	Stom	Hdch	HndS	ETir	Decs
Unhappy	1.00																			
Appetite	0.46	1.00																		
Sleep	0.53	0.41	1.00																	
Fright	0.37	0.28	0.32	1.00																
Tense	0.53	0.41	0.47	0.33	1.00															
Digest	0.47	0.37	0.42	0.29	0.42	1.00														
Think	0.28	0.22	0.25	0.17	0.25	0.22	1.00													
Cry	0.50	0.38	0.44	0.30	0.44	0.39	0.23	1.00												
Enjoy	0.48	0.37	0.43	0.30	0.43	0.38	0.23	0.40	1.00											
Work	0.41	0.32	0.37	0.25	0.37	0.33	0.19	0.34	0.33	1.00										
Useful	0.44	0.34	0.39	0.27	0.39	0.35	0.20	0.36	0.35	0.30	1.00									
Interest	0.45	0.35	0.40	0.28	0.40	0.36	0.21	0.37	0.36	0.31	0.33	1.00								
Worth	0.44	0.34	0.39	0.27	0.40	0.35	0.21	0.37	0.36	0.31	0.32	0.33	1.00							
Give-up	0.49	0.38	0.43	0.30	0.44	0.39	0.23	0.41	0.40	0.34	0.36	0.37	0.36	1.00						
Always Tired	0.54	0.41	0.47	0.33	0.48	0.43	0.25	0.45	0.43	0.37	0.39	0.40	0.40	0.44	1.00					
Stomach	0.48	0.37	0.42	0.29	0.43	0.38	0.22	0.40	0.39	0.33	0.35	0.36	0.35	0.39	0.43	1.00				
Headache	0.49	0.38	0.43	0.30	0.43	0.39	0.23	0.40	0.39	0.34	0.36	0.37	0.36	0.40	0.44	0.39	1.00			
Handshake	0.41	0.32	0.36	0.25	0.37	0.33	0.19	0.34	0.33	0.28	0.30	0.31	0.30	0.34	0.37	0.33	0.33	1.00		
Easily Tired	0.45	0.34	0.40	0.27	0.40	0.35	0.21	0.37	0.36	0.31	0.33	0.34	0.33	0.37	0.40	0.36	0.36	0.31	1.00	
Decision	0.38	0.30	0.34	0.24	0.34	0.31	0.18	0.32	0.31	0.27	0.28	0.29	0.29	0.32	0.35	0.31	0.31	0.26	0.29	1.00

Note: The table shows the correlation matrix for SRQ-20 mental health measures using round 1 items for non-indigenous mothers. The correlation matrix shows the relationship between the mental health measures.

Table A4: Indigenous Mothers Correlation Matrix for SRQ-20 Mental Health Measures

Variable	Unhp	App	Slp	Frgt	Tns	Dgst	Thnk	Cry	Enjy	Wrk	Usfl	Intr	Wrth	GvUp	Tird	Stom	Hdch	HndS	ETir	Decs
Unhappy	1.00																			
Appetite	0.45	1.00																		
Sleep	0.44	0.31	1.00																	
Fright	0.47	0.33	0.32	1.00																
Tense	0.70	0.48	0.47	0.50	1.00															
Digest	0.61	0.42	0.42	0.44	0.65	1.00														
Think	0.37	0.25	0.25	0.27	0.39	0.34	1.00													
Cry	0.54	0.37	0.36	0.39	0.57	0.50	0.30	1.00												
Enjoy	0.50	0.35	0.34	0.36	0.53	0.47	0.28	0.41	1.00											
Work	0.43	0.30	0.29	0.31	0.46	0.41	0.24	0.36	0.33	1.00										
Useful	0.44	0.30	0.30	0.31	0.47	0.41	0.25	0.36	0.33	0.29	1.00									
Interest	0.48	0.33	0.33	0.35	0.51	0.45	0.27	0.39	0.37	0.32	0.32	1.00								
Worth	0.55	0.38	0.37	0.40	0.59	0.52	0.31	0.45	0.42	0.36	0.37	0.40	1.00							
Give	0.61	0.42	0.41	0.44	0.65	0.57	0.34	0.50	0.47	0.40	0.41	0.45	0.51	1.00						
Always Tired	0.62	0.43	0.42	0.45	0.66	0.58	0.35	0.51	0.47	0.41	0.41	0.45	0.52	0.58	1.00					
Stomach	0.55	0.38	0.37	0.40	0.58	0.51	0.31	0.45	0.42	0.36	0.36	0.40	0.46	0.51	0.52	1.00				
Headache	0.55	0.38	0.37	0.40	0.58	0.51	0.31	0.45	0.42	0.36	0.37	0.40	0.46	0.51	0.52	0.46	1.00			
Handshake	0.50	0.35	0.34	0.36	0.54	0.47	0.28	0.41	0.38	0.33	0.33	0.37	0.42	0.47	0.47	0.42	0.42	1.00		
Easily Tired	0.53	0.37	0.36	0.38	0.57	0.50	0.30	0.43	0.41	0.35	0.35	0.39	0.45	0.49	0.50	0.44	0.44	0.41	1.00	
Decision	0.41	0.28	0.28	0.29	0.43	0.38	0.23	0.33	0.31	0.27	0.27	0.30	0.34	0.38	0.38	0.34	0.34	0.31	0.33	1.00

Note: The table shows the correlation matrix for SRQ-20 mental health measures using round 1 items for indigenous mothers. The correlation matrix shows the relationship between the mental health measures.

Table A5: Number of Factors Retained by Different Methods

Approach	Age 1	Age 5	Age 12	Age 15
Parallel analysis	3	3	3	3
Velicer MAP	2	2	1	1
VSS Compl. 1	1	2	1	1
VSS Compl. 2	2	2	2	2
Acceleration factor	1	1	1	1

Note: The table shows the number of factors retained by different methods for each round. We compute the tetracholoric correlation matrix for the SRQ-20 items due to the binary nature of the items. Parallel analysis compares observed eigenvalues with those from random data to determine the number of factors. Velicer's Minimum Average Partial (MAP) method selects the number of factors by minimizing the average squared partial correlation. Very Simple Structure (VSS) assesses the fit of simplified factor structures. Complexity 1 and 2 refer to different ways of measuring factor simplicity. The acceleration factor identifies the optimal number of factors based on the rate of change in eigenvalues.

Table A6: Factor Loadings and Thresholds for Maternal Mental Health Construct

Item	Loadings	Thresholds
Unhappy	0.94	0.13
Poor Appetite	0.69	-0.81
Sleep	0.73	-0.72
Frightened	0.60	-0.35
Nervous/Tense	0.84	0.75
Digestion	0.70	-1.10
Thinking	0.62	-1.02
Cry	0.80	-0.82
Enjoyment	0.91	-1.06
Work	0.78	-1.31
Usefulness	0.72	-1.44
Lost Interest	0.88	-1.56
Worthless	0.86	-1.73
Giving Up	0.76	-2.00
Always Tired	0.96	-0.88
Stomach Issues	0.73	-0.78
Headache	0.66	0.06
Handshake	0.61	-1.34
Easily Tired	0.81	-0.62
Decision-Making	0.67	-0.55

Note: The table shows the factor loadings and thresholds for the maternal mental health construct. The factor loadings indicate the strength of the measure on the latent factor. The thresholds indicate the level of the latent factor at which the item has a 50% probability of being endorsed.

Table A7: Signal-to-Noise Ratios for SRQ-20 Mental Health Measures by Round and Indigenous Status

	Non-Indigenous				Indigenous			
	$s_{t=1}$	$s_{t=2}$	$s_{t=3}$	$s_{t=4}$	$s_{t=1}$	$s_{t=2}$	$s_{t=3}$	$s_{t=4}$
Unhappy	0.472	0.537	0.562	0.610	0.540	0.533	0.621	0.525
Poor Appetite	0.324	0.385	0.408	0.457	0.387	0.381	0.469	0.373
Sleep	0.346	0.407	0.432	0.480	0.410	0.404	0.493	0.395
Frightened	0.264	0.318	0.340	0.385	0.320	0.314	0.397	0.307
Nervous/Tense	0.415	0.480	0.505	0.554	0.483	0.476	0.566	0.468
Digestion	0.329	0.390	0.414	0.462	0.392	0.386	0.474	0.378
Thinking	0.279	0.335	0.358	0.404	0.338	0.331	0.416	0.324
Cry	0.390	0.454	0.478	0.527	0.457	0.450	0.540	0.441
Enjoyment	0.454	0.520	0.545	0.593	0.523	0.516	0.605	0.507
Work	0.380	0.444	0.469	0.517	0.447	0.440	0.530	0.431
Usefulness	0.345	0.406	0.431	0.479	0.409	0.402	0.491	0.394
Lost Interest	0.435	0.500	0.525	0.574	0.503	0.496	0.586	0.488
Worthless	0.426	0.491	0.516	0.565	0.494	0.487	0.577	0.479
Giving Up	0.365	0.428	0.453	0.502	0.431	0.424	0.514	0.416
Always Tired	0.477	0.543	0.568	0.615	0.546	0.539	0.627	0.530
Stomach Issue	0.350	0.412	0.436	0.485	0.415	0.408	0.497	0.400
Headache	0.302	0.360	0.383	0.431	0.363	0.356	0.443	0.348
Handshake	0.272	0.327	0.349	0.395	0.329	0.323	0.407	0.316
Easily Tired	0.394	0.458	0.483	0.532	0.461	0.454	0.544	0.446
Decision-Making	0.313	0.372	0.395	0.443	0.375	0.368	0.455	0.360

Note: The table shows signal-to-noise ratios for the SRQ-20 mental health measures by round and indigenous status. The signal-to-noise ratio for measure j in group g and period t is given by $s_{jgt} = \lambda_{jgt}^2 \phi_{gt} / (\lambda_{jgt}^2 \phi_{gt} + \theta_{jgt})$.

Table A8: Uniqueness Values for SRQ-20 Mental Health Measures by Indigenous Status

	Non-Indigenous			Indigenous		
Measure	Period 1–2	Period 1–4	Period 1–5	Period 1–2	Period 1–4	Period 1–5
Unhappy	-0.04	0.24	0.20	0.61	0.30	-0.04
Poor Appetite	0.29	0.13	0.11	0.39	0.31	-0.16
Sleep	0.21	0.11	0.08	0.22	-0.18	0.29
Frightened	0.20	0.34	0.30	0.23	0.19	0.13
Nervous/Tense	0.14	0.02	0.21	0.34	0.29	0.00
Digestion	0.41	0.44	0.39	0.27	0.36	0.22
Thinking	0.02	-0.04	-0.02	-0.02	0.04	-0.01
Crying	0.21	0.19	0.27	0.08	0.06	0.31
Enjoyment	-0.05	0.07	-0.08	-0.41	0.10	0.08
Work	-0.00	0.09	0.02	-0.25	-0.15	-0.05
Usefulness	0.07	0.09	0.14	-0.05	0.12	0.03
Lost Interest	0.01	0.09	0.05	-0.18	-0.48	0.04
Worthless	0.19	0.22	-0.05	0.12	0.05	0.24
Giving Up	0.37	0.31	0.31	-0.05	0.47	0.41
Always Tired	0.31	-0.05	0.10	0.06	0.02	0.03
Stomach Issues	0.29	0.34	0.12	0.16	0.16	0.08
Headaches	0.37	0.38	0.20	0.21	0.15	0.25
Handshake	0.28	0.11	0.17	0.21	0.19	0.04
Easily Tired	0.09	0.02	0.01	-0.04	0.03	0.17
Decision-Making	0.12	0.27	0.08	-0.26	0.06	0.15

Note: The table shows comparison of uniqueness values between non-Indigenous and Indigenous groups across periods. This table shows the first round residual covariance across time for each mental health measure. We select only the first round to compare the uniqueness values across periods. Other rounds are available upon request. Note, the Theta parameterization sets the first round residual for each measure to 1 (i.e Period 1 – 1 is set to 1).

Table A9: Correlations for Maternal Mental Health Construct by Indigenous Status

Mental Health	Age 1	Age 5	Age 12	Age 15
	Non-Indigenous			
Age 1	1.00			
Age 5	0.46	1.00		
Age 12	0.46	0.41	1.00	
Age 15	0.34	0.37	0.52	1.00
	Indigenous			
Age 1	1.00			
Age 5	0.32	1.00		
Age 12	0.41	0.28	1.00	
Age 15	0.26	0.28	0.24	1.00

Note: This table shows the strict CFA model (Thresholds + Loadings + Errors) correlations for maternal mental health latent factor by non-Indigenous and Indigenous for each round. Age refers to the child's age at the time of the survey. The CFA model is estimated using mean- and variance-adjusted DWLS. Higher values indicate better mental health.

Table A10: Regression Analysis of Patient Health Questionnaire (PHQ-8) on Baseline Maternal Mental Health

	GAD-7		PHQ-8	
	Non-Ind.	Ind.	Non-Ind.	Ind.
	(1)	(2)	(3)	(4)
Mother's Baseline MH	-0.108**	-0.077*	-0.109***	-0.047
	(0.043)	(0.042)	(0.041)	(0.048)
Mother's Education	0.030**	0.029**	0.036***	0.051***
	(0.013)	(0.014)	(0.012)	(0.015)
Wealth Index	-0.045	0.062	0.035	-0.013
	(0.051)	(0.067)	(0.047)	(0.072)
Mother's Age	0.003	-0.005	0.002	0.008
	(0.009)	(0.010)	(0.008)	(0.011)
Num. Children	0.024	0.032	0.043	0.009
	(0.037)	(0.031)	(0.036)	(0.031)
Father in HH	-0.196	-0.058	-0.097	0.026
	(0.120)	(0.157)	(0.114)	(0.175)
R^2	0.0131	0.0208	0.0221	0.0344
Observations	750	370	750	371

Note: The table presents the results of the regression analysis of Patient Health Questionnaire (PHQ-8) on Baseline Maternal Mental Health. The table shows the coefficients and standard errors for the non-Indigenous and Indigenous groups. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table A11: Correlation Matrix for Maternal Mental Health and Socioeconomic Factors by Indigenous Status

	Mother's Education	Wealth Index	Mother's Age	Num. of Children	Father in HH	Mental Health R.1	Mental Health R.2	Mental Health R.4	Mental Health R.5
					Non-Indigenous				
Correlation Matrix									
Mother's Education	1.000								
Wealth Index	0.578	1.000							
Mother's Age	-0.049	0.030	1.000						
Num. Children	-0.383	-0.284	0.641	1.000					
Father in HH	-0.014	-0.042	0.119	0.173	1.000				
MH (Round 1)	0.164	0.091	-0.090	-0.164	0.079	1.000			
MH (Round 2)	0.074	0.059	-0.145	-0.159	0.015	0.460	1.000		
MH (Round 4)	0.096	0.013	-0.039	-0.091	-0.017	0.455	0.406	1.000	
MH (Round 5)	0.036	-0.027	-0.047	-0.025	-0.037	0.341	0.366	0.521	1.000
					Indigenous				
Correlation Matrix									
Mother's Education	1.000								
Wealth Index	0.546	1.000							
Mother's Age	-0.087	0.154	1.000						
Num. Children	-0.384	-0.182	0.719	1.000					
Father in HH	-0.053	0.030	0.106	0.188	1.000				
MH (Round 1)	0.082	0.071	-0.120	-0.181	0.071	1.000			
MH (Round 2)	0.110	0.090	-0.149	-0.247	0.035	0.321	1.000		
MH (Round 4)	0.044	0.034	-0.155	-0.174	0.035	0.410	0.277	1.000	
MH (Round 5)	0.084	0.037	-0.157	-0.186	-0.005	0.260	0.279	0.243	1.000

Note: The table shows the correlation matrix for maternal mental health and socioeconomic factors by non-Indigenous and Indigenous status. The correlation matrix includes maternal mental health (MH) scores at different survey rounds and key socioeconomic variables.

Table A12: Regression Analysis of Maternal Mental Health on Additional Covariates

	Round 1		Round 2		Round 4		Round 5	
	Non-Ind.	Ind.	Non-Ind.	Ind.	Non-Ind.	Ind.	Non-Ind.	Ind.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother's Education	0.029**	-0.003	-0.012	-0.004	0.026**	-0.009	0.006	0.003
	(0.012)	(0.015)	(0.011)	(0.013)	(0.011)	(0.015)	(0.011)	(0.013)
Wealth Index	0.038	0.050	0.030	0.042	-0.053	0.069	-0.007	0.017
	(0.045)	(0.080)	(0.042)	(0.067)	(0.038)	(0.077)	(0.041)	(0.062)
Mother's Age	-0.004	0.004	-0.014**	0.005	0.009	-0.009	-0.012*	-0.003
	(0.007)	(0.012)	(0.007)	(0.011)	(0.007)	(0.011)	(0.007)	(0.010)
Num. Children	-0.070**	-0.077**	-0.009	-0.096***	-0.038	-0.022	0.044	-0.039
	(0.031)	(0.034)	(0.032)	(0.033)	(0.033)	(0.035)	(0.029)	(0.030)
Father in HH	0.234**	0.427***	-0.017	0.175	-0.073	0.192	-0.116	0.076
	(0.096)	(0.161)	(0.093)	(0.139)	(0.096)	(0.175)	(0.095)	(0.127)
Mother's Weight	0.002	-0.016**	-0.003	0.001	0.004	-0.014*	0.001	-0.008
	(0.004)	(0.008)	(0.004)	(0.007)	(0.003)	(0.008)	(0.003)	(0.007)
Mother's Height	-0.002	0.030**	0.005	-0.007	-0.007	0.021*	0.007	0.018*
	(0.007)	(0.012)	(0.006)	(0.012)	(0.006)	(0.011)	(0.006)	(0.010)
Social Capital	0.126***	-0.037	0.029	0.026	0.051	0.005	0.056*	-0.043
	(0.035)	(0.055)	(0.034)	(0.049)	(0.033)	(0.055)	(0.033)	(0.044)
Desired Pregnancy	0.140**	0.282***	0.068	0.029	-0.007	0.217**	0.022	0.034
	(0.066)	(0.106)	(0.065)	(0.094)	(0.064)	(0.101)	(0.063)	(0.084)
Coastal Region	-0.076	0.246	0.101	-0.075	-0.119*	0.108	-0.098	0.004
	(0.071)	(0.192)	(0.066)	(0.139)	(0.066)	(0.188)	(0.067)	(0.140)
Lagged Mental Health			0.470***	0.258***	0.385***	0.260***	0.565***	0.157***
			(0.032)	(0.042)	(0.031)	(0.053)	(0.034)	(0.043)
R^2	0.0628	0.0565	0.2193	0.1268	0.1718	0.0885	0.2780	0.0622
Observations	834	425	834	425	834	425	834	425

Note: The table shows the results of the regression of mental health factor scores at different survey rounds on maternal and household characteristics. This table contains additional covariates. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table A13: Recentered Influence Function (RIF) decomposition for maternal mental health factor at the 10th, 25th, 50th, 75th, and 90th percentile

	P10	P25	P50	P75	P90
	(1)	(2)	(3)	(4)	(5)
Total Difference	0.250**	0.255***	0.125	-0.014	-0.016
	(0.099)	(0.093)	(0.087)	(0.096)	(0.088)
Explained	0.283***	0.260***	0.201***	0.161***	0.097*
	(0.060)	(0.053)	(0.047)	(0.050)	(0.053)
Unexplained	-0.033	-0.005	-0.076	-0.175	-0.114
	(0.113)	(0.103)	(0.096)	(0.107)	(0.102)
Mother's Education	0.128*	0.126**	0.138**	0.089	0.011
	(0.069)	(0.059)	(0.054)	(0.059)	(0.064)
Wealth Index	0.061	0.034	-0.030	-0.009	0.021
	(0.048)	(0.042)	(0.038)	(0.042)	(0.045)
Mother's Age	0.014	0.008	-0.001	-0.007	-0.017
	(0.016)	(0.014)	(0.012)	(0.013)	(0.015)
Num. Children	0.088	0.105**	0.107**	0.096**	0.089*
	(0.055)	(0.048)	(0.044)	(0.048)	(0.051)
Father in HH	-0.007	-0.013	-0.013	-0.009	-0.007
	(0.007)	(0.009)	(0.009)	(0.007)	(0.006)
Observations	1,270	1,270	1,270	1,270	1,270

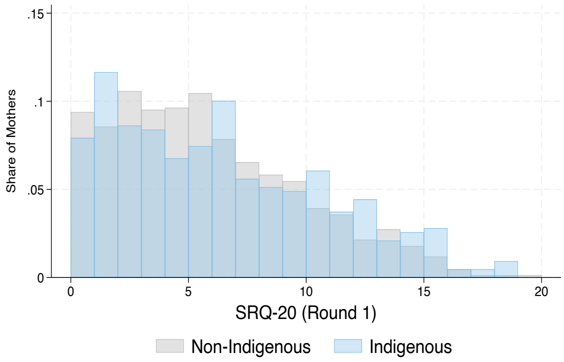
Note: The table presents the results of the Recentered Influence Function (RIF) decomposition analysis for maternal mental health factor scores at different percentiles. The table shows the total difference, explained and unexplained components, and the contribution of each covariate to the explained component. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table A14: Recentered Influence Function (RIF) decomposition for maternal mental health factor at rounds 2, 4, and 5

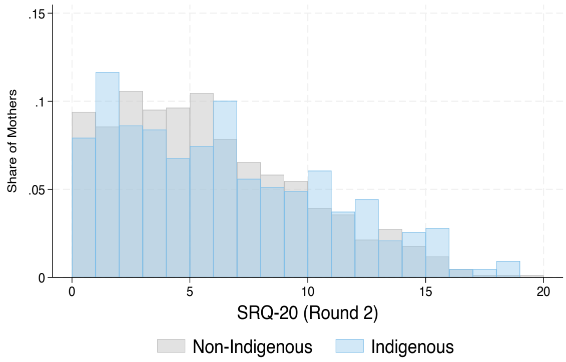
	Round 2					Round 4					Round 5				
	P10	P25	P50	P75	P90	P10	P25	P50	P75	P90	P10	P25	P50	P75	P90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Total Difference	0.269**	0.278***	0.337***	0.394***	0.182**	0.306**	0.393***	0.280***	0.221**	-0.050	0.024	0.322***	0.428***	0.537***	0.489***
	(0.105)	(0.091)	(0.075)	(0.083)	(0.086)	(0.126)	(0.090)	(0.077)	(0.092)	(0.079)	(0.096)	(0.084)	(0.073)	(0.085)	(0.078)
Explained	0.164**	0.129**	0.093*	-0.022	0.022	0.269***	0.192***	0.114**	0.145***	0.175***	0.123	0.156**	0.095*	0.117*	0.131***
	(0.073)	(0.055)	(0.053)	(0.061)	(0.047)	(0.073)	(0.058)	(0.047)	(0.051)	(0.047)	(0.080)	(0.066)	(0.055)	(0.063)	(0.049)
Unexplained	0.104	0.149	0.243***	0.417***	0.160*	0.036	0.201**	0.165*	0.076	-0.225	-0.099	0.166*	0.332***	0.420***	0.358***
	(0.121)	(0.099)	(0.083)	(0.094)	(0.090)	(0.140)	(0.102)	(0.085)	(0.099)	(0.088)	(0.118)	(0.100)	(0.085)	(0.096)	(0.084)
Lagged	0.061*	0.047*	0.055*	0.064*	0.052*	0.123***	0.137***	0.123***	0.125***	0.112***	0.178***	0.166***	0.140***	0.163***	0.136***
Mental Health	(0.035)	(0.027)	(0.031)	(0.036)	(0.029)	(0.031)	(0.031)	(0.027)	(0.028)	(0.025)	(0.046)	(0.042)	(0.035)	(0.041)	(0.034)
Share Explained by Covariates															
Mother's	-0.049	-0.008	-0.009	-0.077	0.025	0.122	0.164***	0.045	0.063	0.097*	0.004	0.009	0.030	0.027	0.018
Education	(0.076)	(0.056)	(0.052)	(0.060)	(0.045)	(0.079)	(0.061)	(0.048)	(0.053)	(0.050)	(0.082)	(0.064)	(0.053)	(0.059)	(0.044)
Wealth Index	0.109**	0.019	0.033	-0.040	-0.071**	-0.062	-0.111***	-0.050	-0.069*	-0.018	-0.023	-0.001	-0.028	-0.075*	-0.013
	(0.054)	(0.040)	(0.037)	(0.042)	(0.032)	(0.056)	(0.043)	(0.034)	(0.037)	(0.035)	(0.058)	(0.045)	(0.037)	(0.042)	(0.031)
Mother's Age	0.035*	0.022	0.014	0.010	0.006	-0.020	-0.010	-0.014	-0.013	-0.002	0.010	0.005	0.010	0.019	0.017
	(0.020)	(0.014)	(0.012)	(0.014)	(0.010)	(0.019)	(0.014)	(0.012)	(0.013)	(0.011)	(0.019)	(0.014)	(0.012)	(0.014)	(0.011)
Num. Children	0.013	0.050	0.002	0.014	0.005	0.111*	0.012	0.004	0.033	-0.023	-0.055	-0.027	-0.062	-0.021	-0.028
	(0.060)	(0.045)	(0.041)	(0.047)	(0.035)	(0.064)	(0.047)	(0.038)	(0.041)	(0.039)	(0.065)	(0.050)	(0.042)	(0.046)	(0.035)
Father in HH	-0.005	-0.001	-0.001	0.006	0.006	-0.005	0.000	0.005	0.007	0.008	0.009	0.003	0.005	0.004	0.001
	(0.007)	(0.004)	(0.004)	(0.006)	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	(0.005)	(0.005)	(0.005)	(0.004)
Observations								1,270							

Note: The table presents the results of the Recentered Influence Function (RIF) decomposition analysis for maternal mental health factor scores at different percentiles. The table shows the total difference, explained and unexplained components, and the contribution of each covariate to the explained component. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

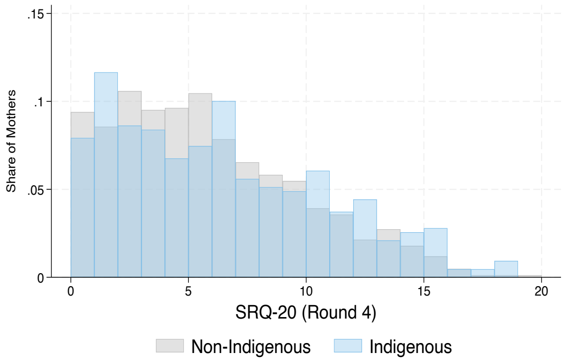
Figure A1: Distribution of Raw SRQ-20 Scores Across Groups and Over Time



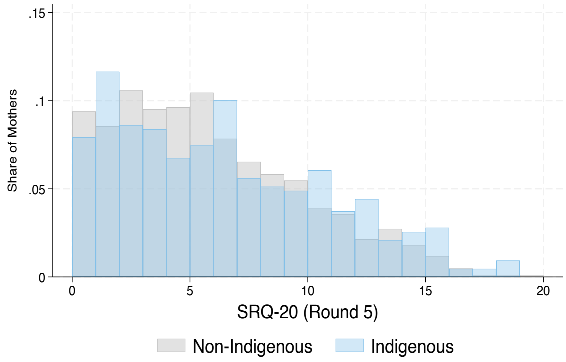
(a) SRQ-20 Scores: Round 1



(b) SRQ-20 Scores: Round 2



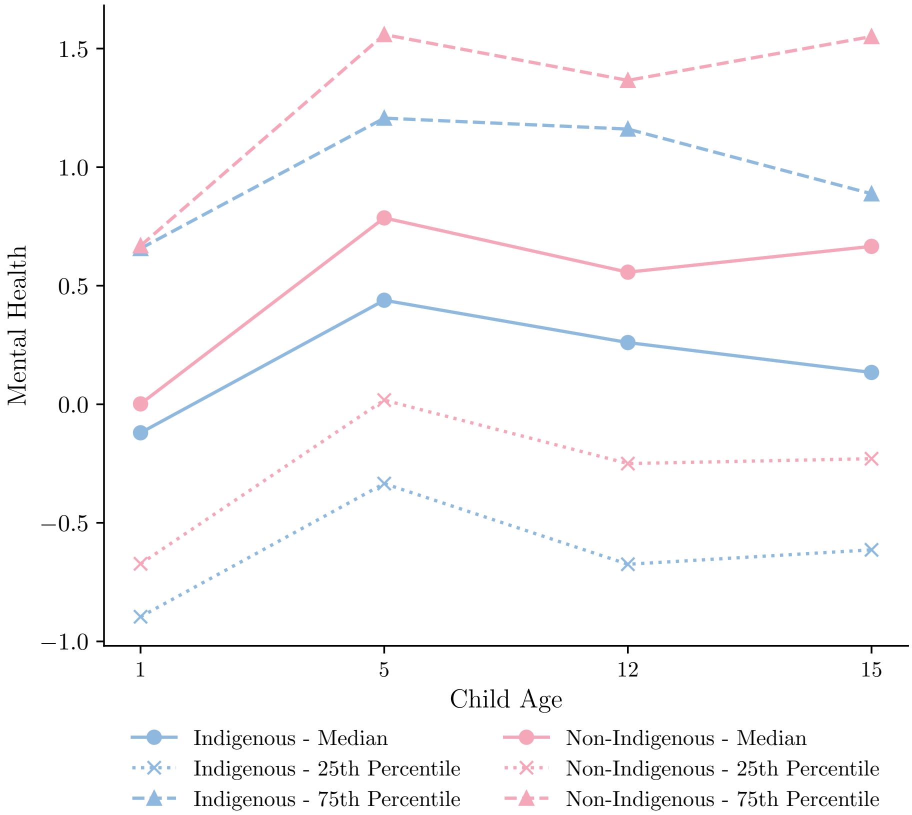
(c) SRQ-20 Scores: Round 4



(d) SRQ-20 Scores: Round 5

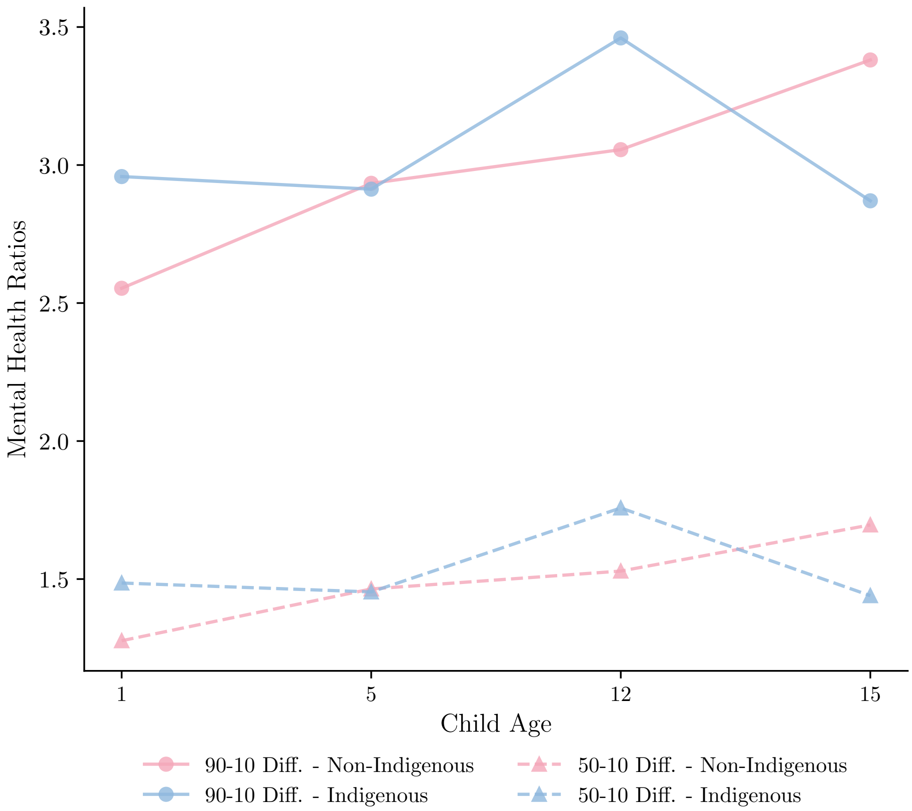
This figure displays the distribution of raw SRQ-20 scores for Indigenous and non-Indigenous mothers in Rounds 1, 2, 4 and 5 for the main sample of mothers included in the paper.

Figure A2: Evolution of Maternal Mental Health by Indigenous Status



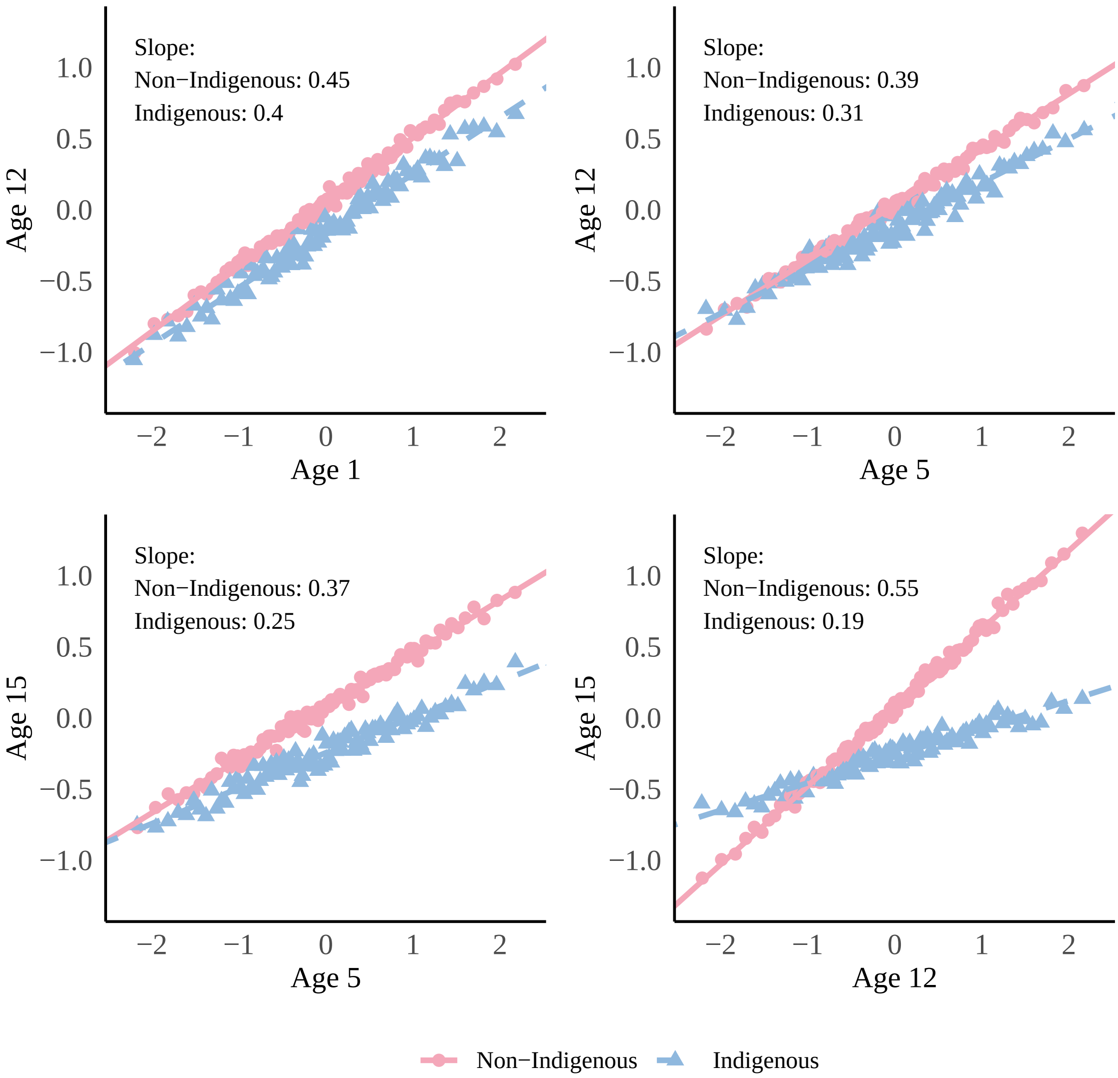
The plot shows the 75th, Median, and 25th percentile of the distribution of maternal mental health factor scores. Age represents the age of the child at the time of the survey. The scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health.

Figure A3: Percentile differences of Maternal Mental Health by Indigenous Status



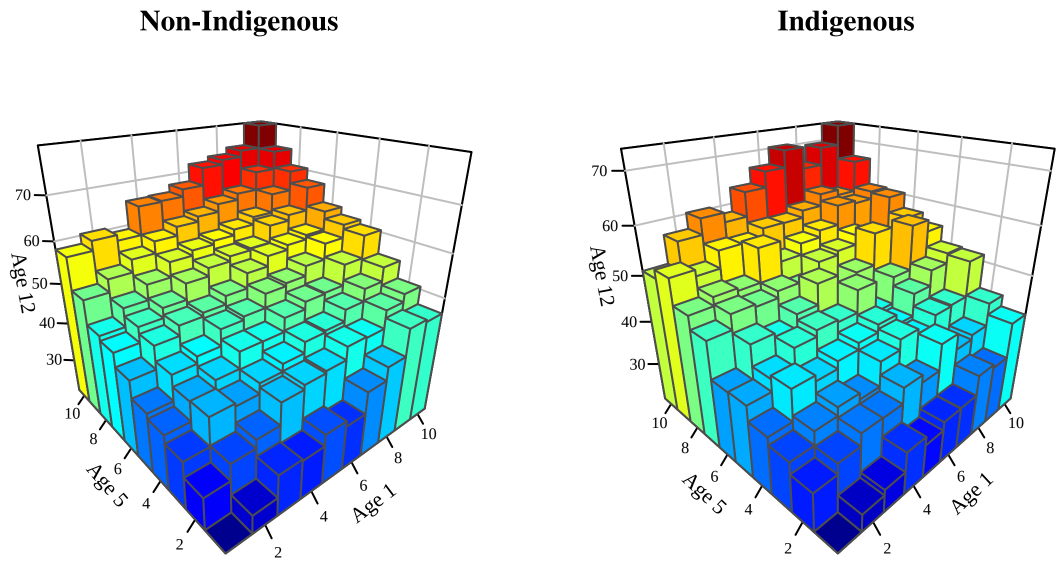
The plot shows the 90-10 and 50-10 percentile differences of the distribution of maternal mental health factor scores. Age represents the age of the child at the time of the survey. The scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health.

Figure A4: Persistence of Maternal Mental Health by Indigenous Status



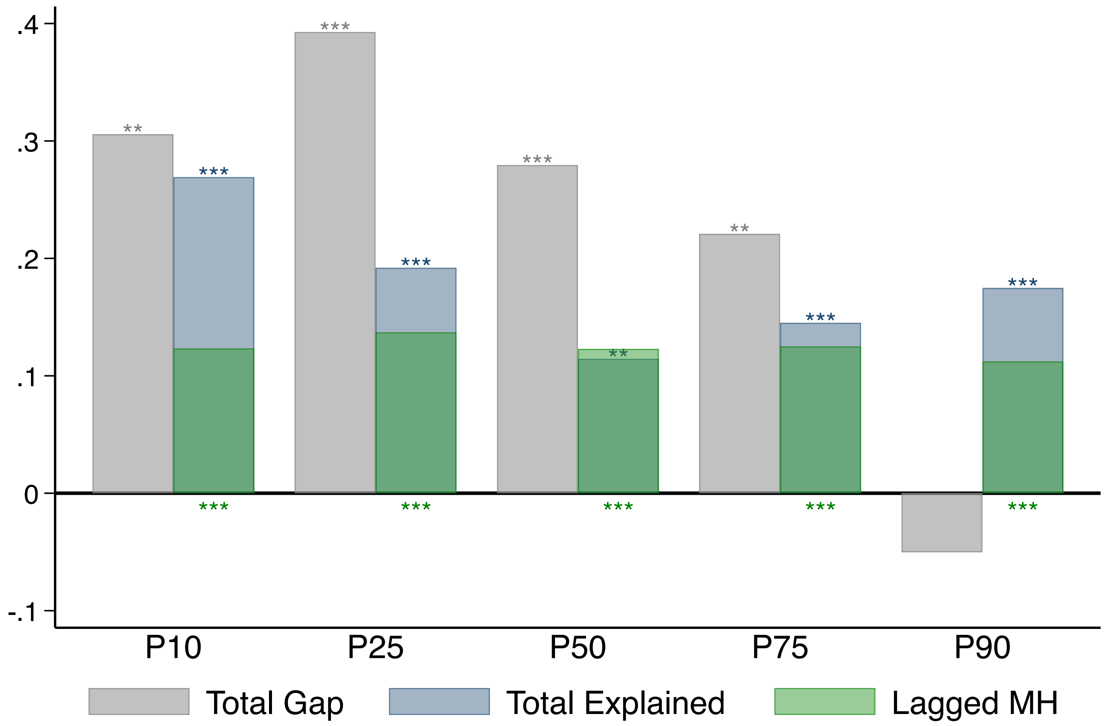
The plot shows the persistence of maternal mental health factor scores by Indigenous status. We standardized latent mental health at age 15 versus standardized latent mental health at age 1, separated by Indigenous status. We plot 100 binned averages (1 percentile), and the lines are linear fits for each group.The scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health.

Figure A5: Joint Distribution of Maternal Mental Health Latent Factor at Ages 1, 5, and 12



The left figure is non-Indigenous mothers mental health, and the right figure is Indigenous mothers mental health. The bottom axis is the *decile* of the mental health distribution at age 1 and 5. The vertical axis is the *percentile* of the mental health distribution at age 12. The height of the bars represents the density of mothers in each cell.

Figure A6: RIF Decomposition of Gaps in Maternal Mental Health (Round 4)



The plot shows the Recentered Influence Function (RIF) decomposition of the gaps in maternal mental health factor scores at the 10th, 25th, 50th, 75th, and 90th percentiles. The total gap is the difference in mental health factor scores between Indigenous and non-Indigenous mothers. The explained component is the part of the gap explained by the covariates, and the unexplained component is the part of the gap not explained by the covariates. The green bar explains the share of the gap at different percentiles explained by lagged maternal mental health. The factor scores are estimated using the Empirical Bayes Modal approach, adjusting the factor scores to align with the model-implied variance-covariance matrix. Higher values represent *better* mental health. *p<0.1, **p<0.05, ***p<0.01.