BACHPAN STUDY

Ethical Guidelines for Secondary Data Use

UNC Gillings School of Global Public Health, USA | Pakistan Institute of Living and Learning (PILL), Pakistan

Preamble

The Bachpan cohort study is a pregnancy-birth cohort with an embedded cRCT investigating the effects of perinatal depression on the cognitive and socioemotional development of children in rural Rawalpindi, Pakistan. Enrolled between October 2014 and February 2016, the study follows 1,154 mother-child dyads, encompassing longitudinal measures of maternal mental health, child development, and other contextual data.

To ensure the responsible stewardship of this detailed longitudinal dataset, all secondary use must be conducted in accordance with the ethical principles set forth in this document. These guidelines are consistent with the original study approvals, international research ethics standards (including the Declaration of Helsinki), and Pakistani national research ethics frameworks.

Study Profile	
Study Name	Bachpan Study
Location	Kallar Syedan sub-district, Rawalpindi, Pakistan
Participants	1,154 pregnant women (570 prenatally depressed; 584 non-depressed) and their children
Follow-up Waves (as of March 2026)	Baseline (antenatal), 3, 6, 12, 24, 36, 48, 72, 84, 96 months postnatal
Funding	National Institute of Health (NIH)
Ethics Contacts	jmaselko@unc.edu | sihamsikander@gmail.com

1. Scope and Applicability

These guidelines apply to all individuals, institutions, and research teams who seek to access, analyze, publish from, or otherwise use data collected as part of the Bachpan Study. This includes:

- Secondary analyses of quantitative survey data

- Record linkage with external datasets
- Methodological or validation studies using Bachpan instruments or measures
- Development, training, testing, and validation of machine learning models, artificial intelligence systems, and predictive algorithms

These guidelines do not supersede the original study's ethics approvals but supplement them for secondary use contexts. Any secondary use not explicitly covered here requires written consultation with the Bachpan principal investigators.

2. Foundational Ethical Principles

2.1 Respect for Persons

Participants provided informed consent for the original study. Secondary users must honor the spirit of that consent. Secondary analyses should be limited to purposes reasonably foreseeable to participants at the time of enrolment. Broadly, this includes research into maternal mental health, child development, and related social determinants of health in low-resource settings.

2.2 Beneficence and Non-Maleficence

Secondary research must be designed to generate knowledge that can benefit populations similar to the study cohort. Analysis that could stigmatize communities, pathologize poverty, or produce findings that could be weaponized against the interests of women in rural Pakistan is prohibited.

2.3 Justice

Researchers using Bachpan data should actively consider how findings will be communicated to and benefit the communities from which data were drawn. Efforts should be made to include researchers from Pakistan and other LMICs in secondary study teams.

2.4 Scientific Integrity

Analyses should be pre-registered where possible, transparent in their methods, and honest about limitations. Researchers should clearly distinguish between pre-planned confirmatory analyses and post hoc exploratory analyses to prevent data dredging or "p-hacking." We strongly encourage the publication of all rigorously conducted research, including null or negative findings, to mitigate publication bias. To ensure reproducibility, researchers are expected to make their analytic code and statistical scripts openly available upon publication.

3. Data Access and Governance

Data from pregnancy through the 3-year-old wave is free and accessible through the Harvard Dataverse (https://doi.org/10.7910/DVN/IJE2PC) as of the creation of this document.

If an investigator is interested in data beyond the 3-year-old wave, please contact the Principal Investigators (Drs. Joanna Maselko or Siham Sikander; see Section 7. Contact Information).

4. Participant Protection and Confidentiality

4.1 Anonymization and De-identification

All data shared for secondary is de-identified in accordance with the study's original anonymization protocols. Direct identifiers (names, addresses, National Identity Card numbers) are removed before any data release. Secondary users must not:

- Attempt to re-identify participants through data linkage, inference, or any other means
- Store data on unsecured devices, personal computers, or non-institutional cloud platforms

Given the small, geographically concentrated population (Kallar Syedan, population ~190,000–250,000), researchers should be particularly vigilant about the risk of indirect re-identification through combinations of demographic variables such as village, age, parity, and occupation.

4.2 Sensitive Data Categories

The following data domains require heightened protection and are subject to additional restrictions on use and publication:

- Intimate partner violence (IPV) disclosures (physical, psychological, and sexual)
- History of mental illness and substance use in the family
- Adverse childhood experiences of enrolled women
- Specific household financial data (debt, economic shocks)

Findings derived from these domains should not be reported at a level of granularity that could identify individuals, villages, or sub-communities. Cell sizes below 10 must be suppressed or combined in published tables.

4.3 Children's Data

Children in the Bachpan cohort were enrolled as minors and cannot themselves provide informed consent. Secondary users should treat child data with the highest standard of care. Cognitive, socioemotional, and developmental scores must not be used to make any characterization of individual children. All child-level data publications must comply with relevant child protection frameworks in both Pakistan and the researcher's home country.

5. Scientific Conduct and Analytical Integrity

5.1 Handling the Study Design

The Bachpan cohort intentionally oversampled prenatally depressed women at a 1:1 ratio with non-depressed women. This design means the unweighted sample does not represent the population distribution of depression. If a researcher wants to constitute a sample that is representative of the underlying population of pregnant women at the time of the study, they should:

- Apply cluster-specific survey weights when reporting population-representative estimates
- Clearly describe the sampling design in all methods sections
- Distinguish between internal group comparisons (no weighting required) and population-level prevalence estimates (weighting required)

5.2 Cultural and Contextual Interpretation

Findings should be interpreted within the cultural, economic, and social context of rural Pakistan. Analysts are strongly encouraged to consult with the Bachpan Team, Pakistani researchers, or community representatives when interpreting results. Factors to be considered include, but are not limited to:

- Gender-based risks: IPV, son preference, role restrictions
- Extended family household structures
- The role of community health workers (Lady Health Workers) as trusted intermediaries
- Cultural practices such as chilla (40-day post-birth confinement period)

5.3 Reporting of Null or Complex Findings

Researchers using Bachpan Study data should report non-significant, null, or unexpected findings transparently and responsibly. Null results carry scientific value and should not be suppressed or framed in ways that imply failure of psychosocial interventions or cast doubt on the mental health needs of the study population. Specifically, researchers should:

- Contextualize null findings within the broader LMIC evidence base, acknowledging structural, economic, and sociocultural factors that may moderate intervention effects or mask associations in the data
- Avoid language or framing that inadvertently stigmatizes the Bachpan Study community, pathologizes caregiving practices, or implies that psychosocial care is ineffective without addressing the contextual constraints under which it was delivered
- Distinguish clearly between absence of evidence and evidence of absence, particularly when sample sizes, follow-up periods, or measurement tools may limit statistical power
- Discuss plausible explanations for unexpected findings, including study design considerations, data collection challenges in field settings, or population heterogeneity

Where findings are complex or mixed, researchers are encouraged to consult with the Bachpan Team prior to publication to ensure accurate interpretation of the data.

5.4 Algorithmic Fairness and Machine Learning (AI/ML)

Researchers applying artificial intelligence (AI) or machine learning (ML) methods to Bachpan Study data should ensure that their analytical approaches account for the study's survey design and are documented with sufficient transparency for methodological review. Deliberate efforts should be made to identify and mitigate algorithmic bias throughout model development, validation, and reporting. All AI/ML-based analyses should be reported in accordance with the most current reporting guidelines for AI/ML research at the time of submission for publication.

6. Publication, Dissemination, and Attribution

6.1 Data Citation

All publications must cite the original cohort profile paper:

- Sikander, S., Ahmad, I., Bates, L. M., Gallis, J., Hagaman, A., O'Donnell, K., Turner, E. L., Zaidi, A., Rahman, A., & Maselko, J. (2019). Cohort Profile: Perinatal depression and child socioemotional development ; the Bachpan cohort study from rural Pakistan. *BMJ Open, 9*(5), e025644. https://doi.org/10.1136/bmjopen-2018-025644

6.2 Tracking Publications

Researchers are requested to share any published manuscripts or reports using Bachpan data with the Bachpan Team via the study website (https://www.bachpanstudy.com/) or by emailing the Principal Investigators (see Section 7 for contact information).

6.3 Commercialization and Digital Applications

In the event that secondary data is used for digitalization, commercialization, or the development of digital tools and predictive applications (e.g., apps related to depression, IPV, adverse childhood experiences, or child outcomes), users are expected to act responsibly and in adherence to the ethical principles outlined in this document. Any such commercial, digital, or predictive products must explicitly acknowledge the Bachpan Study as the underlying data source.

7. Contact Information

PI Contact	Dr. Joanna Maselko, Department of Epidemiology, UNC-Chapel Hill — jmaselko@unc.edu
PI Contact	Dr. Siham Sikander, Pakistan Institute of Living and Learning, Karachi — sihamsikander@gmail.com

8. References and Bibliography

The following sources informed the development of these ethical guidelines. Secondary users are encouraged to consult these documents when designing studies and interpreting findings.

Primary Study Reference

1. Sikander S, Ahmad I, Bates LM, Gallis J, Hagaman A, O'Donnell K, Turner EL, Zaidi A, Rahman A, Maselko J. Cohort profile: perinatal depression and child socioemotional development—the Bachpan cohort study from rural Pakistan. *BMJ Open*. 2019;9:e025644. doi:10.1136/bmjopen-2018-025644.

International Research Ethics Frameworks

2. World Medical Association. *Declaration of Helsinki: ethical principles for medical research involving human subjects*. Fortaleza: WMA General Assembly; 2024. https://www.wma.net/policies-post/wma-declaration-of-helsinki/
3. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: ethical principles and guidelines for the protection of human subjects of research*. Washington (DC): U.S. Department of Health, Education and Welfare; 1979. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/
4. Council for International Organizations of Medical Sciences (CIOMS), World Health Organization. *International Ethical Guidelines for Health-Related Research Involving Humans*. 4th ed. Geneva: CIOMS; 2016. https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/
5. World Health Organization. *Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants*. Geneva: World Health Organization; 2011. https://www.who.int/publications/i/item/9789241502948

Data Protection and Governance

6. European Parliament and Council of the European Union. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*. 2016;L119:1–88. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679
7. Ministry of IT & Telecom, Government of Pakistan. *Personal Data Protection Bill (Draft)*. Islamabad: National Assembly of Pakistan; 2023. https://www.icnl.org/resources/library/personal-data-protection-bill
8. National Institutes of Health (NIH). *NIH Genomic Data Sharing Policy*. https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/gds
9. Snapes E, Astrin JJ, Krüger NB, Grossman GH, Hendrickson E, Miller N, et al. Updating International Society for Biological and Environmental Repositories best practices, fifth edition: a new process for relevance in an evolving landscape. *Biopreservation and Biobanking*. 2023:537-546. doi:10.1089/bio.2023.0140

LMIC and Global Health Research Ethics

10. Nuffield Council on Bioethics. *The Ethics of Research Related to Healthcare in Developing Countries*. London: Nuffield Council on Bioethics; 2002. https://www.nuffieldbioethics.org/publication/the-ethics-of-research-related-to-healthcare-in-developing-countries/

11. Singer PA, Benatar SR. Beyond Helsinki: a vision for global health ethics. *BMJ*. 2001;322(7289):747–748. doi:10.1136/bmj.322.7289.747

12. Rahman A, Fisher J, Bower P, et al. Interventions for common perinatal mental disorders in women in low- and middle-income countries: a systematic review and meta-analysis. *Bulletin of the World Health Organization*. 2013;91:593–601. doi:10.2471/BLT.12.109819

Children's Rights and Protection

13. United Nations. *Convention on the Rights of the Child*. New York: United Nations; 1989. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

14. Wendler D, Varma S. Minimal risk in pediatric research. *Journal of Pediatrics*. 2006;149(6):855–861.

Secondary Data Use and Open Science

15. Fernando B, King M, Sumathipala A. Advancing good governance in data sharing and biobanking - international aspects. *Wellcome Open Research*. 2019;4:184. doi:10.12688/wellcomeopenres.15540.1

16. Hrynaszkiewicz I, Altman DG. Towards agreement on best practice for publishing raw clinical trial data. *Trials*. 2009;10:17. doi:10.1186/1745-6215-10-17

These guidelines are effective as of the date of signing of any Data Use Agreement and supersede any prior informal arrangements.

Questions regarding the interpretation of these guidelines should be directed to the Bachpan Team.