Predicting Poor Mental Health Days Using Behavioral and Demographic Data from BRFSS (2022–2023)

Literature Review for I-492 Project

by

Abdullah Siddiqui

Advisor
Dr. Sridhar Ramachandran

Table of Contents

I. Research Aims and Objectives

This literature review examines contemporary research that has used the Behavioral Risk Factor Surveillance System (BRFSS) to study mental health and related behavioral predictors, and work that applies machine learning and predictive modeling to BRFSS or similar large public-health surveys. The aims are to:

1. Summarize recent (post-2022) findings about the relationship between behavioral factors (sleep, exercise, substance use) and self-reported poor mental health days.
2. Review methodological approaches to predicting mental health outcomes using BRFSS or comparable survey data.
3. Identify gaps in the literature that motivate the present thesis.

II. Literature Review

This section consists of a critical evaluation of published literature relating to BRFSS mental health research.

2.1 Theme A

The Behavioral Risk Factor Surveillance System (BRFSS) is the largest ongoing health survey in the world, with ~400,000 U.S. adults surveyed annually. It provides a robust foundation for population-level studies but has limitations in design and methodology.

2.1.1 Subtheme 1

• Centers for Disease Control and Prevention (2022, 2023): Provided documentation of survey design, stratification, and weighting, confirming its representativeness.
• Minnesota Department of Health (2024): Used BRFSS 2023 to report health status among Minnesota adults, demonstrating its state-level applicability.

Comparison: Both studies emphasize representativeness but acknowledge BRFSS's cross-sectional limits. CDC provides technical credibility, while state-level reports show applied utility.

2.1.2 Subtheme 2

• Ahmadi-Montecalvo et al. (2025): Used BRFSS 2022 among veterans (n=15,000). Found unmet social needs and stress significantly associated with poor mental health.

Comparison: Unlike general CDC/state applications, this study focuses on a vulnerable subgroup, showing BRFSS's adaptability across populations.

2.1.3 Subtheme 3

• BRFSS has also been used for monitoring risk factors like obesity, chronic conditions, and oral health outcomes (e.g., Salvi et al., 2025). These applications demonstrate its multidimensional nature.

Comparison: These diverse uses highlight flexibility but also illustrate that mental health applications are less developed relative to other health domains.

2.1.4 Summary for Theme A

The reviewed studies demonstrate that BRFSS is a methodologically strong, nationally representative dataset that has been widely used for both general health monitoring and mental health research. CDC documentation (2022, 2023) establishes its credibility, while state-level applications (Minnesota Department of Health, 2024) show its adaptability for regional insights. Ahmadi-Montecalvo et al. (2025) illustrate its use with vulnerable subpopulations, and other studies confirm its broad utility.

Strengths: Large sample sizes, standardized survey design, and breadth of behavioral measures.
Weaknesses: Reliance on self-report introduces bias, and its cross-sectional design prevents causal inference.

Synthesis: Together, these studies affirm BRFSS as a reliable foundation for mental health research while underscoring the need for innovative analytic methods—such as machine learning—that can maximize its utility despite inherent limitations.

2.2 Theme B

Behavioral and lifestyle factors consistently predict poor mental health days.

• Okeke et al. (2024): Analyzed BRFSS 2022 (~430,000 adults). Short sleep (<7 hrs) strongly predicted frequent mental distress.
• Pearce et al. (2022): Used NHANES cohort (n=120,000). Regular activity reduced depression risk by 25%.
• Kamal et al. (2023): Applied ML to BRFSS. Found smoking and alcohol use were leading predictors of poor mental health.

Comparison: Okeke and Pearce highlight protective factors (adequate sleep, activity), while Kamal focuses on risk factors. Despite dataset differences, all confirm that modifiable behaviors shape mental health.

Behavioral predictors such as sleep, physical activity, and substance use consistently emerge across studies, supporting their role as key drivers of mental health outcomes. Strengths: Consistency across datasets strengthens confidence in behavioral predictors. Weaknesses: Most studies are correlational, limiting causal claims. Synthesis: These findings suggest behavioral predictors are strong inputs for predictive modeling, though methodological innovation is needed to improve generalizability.

2.3 Theme C: Machine Learning and Predictive Modeling

Overview
Machine learning (ML) has expanded BRFSS applications beyond traditional regression, though methodological gaps remain.

2.3.1 Subtheme 1: Predictive Models Using BRFSS

- Kamal et al. (2023): **ML predicted poor mental health days (~70% accuracy) but ignored weighting.**
- Salvi et al. (2025): **Applied explainable AI (SHAP) to BRFSS 2022 tooth loss classification, improving transparency.**

Comparison: **Kamal demonstrates predictive potential; Salvi shows interpretability can be achieved. Together, they highlight opportunities to merge prediction with transparency.**

2.3.2 Subtheme 2: Other BRFSS Applications

- **Stroke Risk Prediction (2023): ML achieved high accuracy predicting stroke risk from BRFSS 2022.**
- **Weaver (2024): Predicted diabetes risk using ensemble ML on BRFSS 2022.**

Comparison: Both expand BRFSS beyond mental health, confirming its versatility for multiple health outcomes.

2.3.3 Subtheme 3: Longitudinal and Youth Models

- **Siordia et al. (2025): Used ML + poststratification to track mental health trends (1993–2023).**
- **Ding et al. (2025): Built ML models predicting youth mental health risks, showing generalization to younger populations.**

Comparison: Siordia emphasizes trends across time, while Ding highlights youth risks, together showing BRFSS + ML can capture both temporal and developmental perspectives.

2.3.4 Summary for Theme C

ML enhances BRFSS applications but is often limited by lack of weighting and weak interpretability.

Strengths: Extends BRFSS utility across conditions; shows high predictive power.
Weaknesses: Models often fail to account for survey design, limiting external validity.
Synthesis: ML can transform BRFSS into a predictive tool for public health, but methodological refinement is required to ensure fairness and interpretability.

2.5 Critical Evaluation

- **Contributions: Literature shows BRFSS's power for behavioral and ML studies.**
- **Strengths: Large representative datasets; versatile applications.**

- **Weaknesses: Self-report bias, cross-sectional design, underuse of survey design in ML.**
- **Gaps: Limited 2023 BRFSS use for poor mental health days; need for survey-weighted, interpretable ML models.**
- **Next Steps: This thesis will address these by applying survey-aware, interpretable ML (e.g., SHAP, LIME) to BRFSS 2022/2023 to predict poor mental health days.**

III. Conclusion: An evaluation/critique of the existing literature

The reviewed literature confirms BRFSS's value in mental health research and highlights consistent behavioral predictors (sleep, activity, substance use). ML studies show predictive potential but lack methodological rigor.

Gap Identified: No existing study applies interpretable, survey-weighted ML to BRFSS 2022/2023 specifically targeting poor mental health days.

Thesis Contribution: This project will fill that gap by integrating behavioral predictors with interpretable ML while respecting survey design. This contribution is both theoretical (advancing methodological rigor in ML + survey data) and practical (informing behavioral interventions for mental health).

List of References

1. **Centers for Disease Control and Prevention.** (2022). *2022 BRFSS Survey Data and Documentation.* U.S. Department of Health and Human Services. Retrieved from https://www.cdc.gov/brfss/annual_data/annual_2022.html
2. **Centers for Disease Control and Prevention.** (2023). *2023 BRFSS Survey Data and Documentation.* U.S. Department of Health and Human Services. Retrieved from https://www.cdc.gov/brfss/annual_data/annual_2023.html
3. **Bruss, K. V., Seth, P., Zhao, G., ...** (2024). Loneliness, lack of social and emotional support, and mental health issues — United States, 2022. *MMWR Morb Mortal Wkly Rep, 73*(24), 539-545. https://doi.org/10.15585/mmwr.mm7324a1
4. Mann, S., Schuler, M. S., Paulson, A., Dunbar, M. S., & colleagues. (2025). The mental health age gradient by gender identity. *Social Psychiatry and Psychiatric Epidemiology*. Advance online publication. https://pmc.ncbi.nlm.nih.gov/articles/PMC12378824/
5. **Faruque, F., Shah, G. H., & Bohler, R. M.** (2025). The association between social determinants of health and mental health status in the United States: Analysis of 2023 BRFSS. *European Journal of Investigation in Health, Psychology and Education, 15*(5), 87. https://doi.org/10.3390/ejihpe15050087
6. **Ahmadi-Montecalvo, H., White, Z., Castillo, Y., Beattie-Lopez, S., Sanders, K., & Daus, M.** (2025). Examining health-related social needs and their association with stress and mental health among US Veterans using 2022 BRFSS data. *Discover Public Health, 22*(353). https://doi.org/10.1186/s12982-025-00746-9
7. **Minnesota Department of Health.** (2024). *Health Status Among Minnesota Adults: BRFSS 2023.* St. Paul, MN. https://www.health.state.mn.us/data/mchs/pubs/healthstatus-brfss-2023.pdf
8. Pearce, M., Garcia, L., Abbas, A., Strain, T., Schuch, F. B., Golubic, R., Kelly, P., Khan, S., Utukuri, M., Laird, Y., Mok, A., Smith, A., Tainio, M., Brage, S., & Woodcock, J. (2022). *Association Between Physical Activity and Risk of Depression: A Systematic Review and Meta-analysis.* JAMA Psychiatry, 79(6), 550-559. https://doi.org/10.1001/jamapsychiatry.2022.0609
9. **Kamal, S., Alharbi, F., & Kumar, M.** (2023). Using machine learning to predict poor mental health from BRFSS. *The American Journal of Geriatric Psychiatry.* https://doi.org/10.1016/j.jagp.2023.03.011

10. **MacNell, N., Feinstein, L., Wilkerson, J., ...** (2023). Implementing machine learning methods with complex survey data: Impacts of accounting for sampling weights in gradient boosting. *PLOS ONE, 18*(9), e0280387. https://doi.org/10.1371/journal.pone.0280387
11. **Wadekar, A. S., & Reiter, J. P.** (2023). Evaluating binary outcome classifiers estimated from survey data. *arXiv preprint arXiv:2311.00596.* https://doi.org/10.48550/arXiv.2311.00596
12. **Salvi, S., Roy, A., & Kalagnanam, J.** (2025). Classifying tooth loss in the United States using BRFSS 2022 and explainable AI. *Electronics, 14*(17), 3559. https://doi.org/10.3390/electronics14173559
13. **Ding, H., Li, N., Li, L., Xu, Z., & Xia, W.** (2025). Machine learning-enabled mental health risk prediction for youths with stressful life events: A modelling study. *Journal of Affective Disorders, 368,* 537-546. https://doi.org/10.1016/j.jad.2024.09.111
14. **Siordia, C., Dwyer-Lindgren, L., & Mokdad, A. H.** (2025). Mental health temporal trends from 1993 to 2023: Supervised machine learning and poststratification with BRFSS and ACS. *Harvard Dataverse Preprint.* https://doi.org/10.7910/DVN/3CWJJX
15. **Weaver, A.** (2024). Machine learning and evaluation of diabetes risk using 2022 BRFSS data (Master's thesis). *Western Michigan University.* https://scholarworks.wmich.edu/masters_theses/5447

Link to the References

Google Doc folder link:

https://docs.google.com/document/d/1FReT1GlA9o9T3MTGsNI4uckBdxV0Ebi0Dq43JGOQDso/edit?usp=sharing