

# NHANES Sleep Research as a Cautionary Tale: When Big Data Goes Wrong

Ahmed S BaHammam <sup>1,2</sup>

<sup>1</sup>The University Sleep Disorders Center, Department of Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia; <sup>2</sup>King Saud University Medical City, Riyadh, Saudi Arabia

Correspondence: Ahmed S BaHammam, University Sleep Disorders Center, Department of Medicine, College of Medicine, King Saud University, Box 225503, Riyadh, 11324, Saudi Arabia, Email [ashammam2@gmail.com](mailto:ashammam2@gmail.com)

The National Health and Nutrition Examination Survey (NHANES) has emerged as one of the premier datasets for population-based health research, and nowhere is this more evident than in the explosive growth of sleep medicine investigations. We conducted a quick bibliometric analysis using PubMed from inception to October 8, 2025, with the search strategy: “(NHANES OR ‘National Health and Nutrition Examination Survey’) AND (sleep OR ‘sleep disorders’ OR ‘sleep medicine’ OR insomnia OR ‘sleep apnea’ OR ‘sleep duration’ OR ‘sleep quality’)” which yielded 3095 papers. From this dataset, we identified 655 papers mentioning NHANES, 1525 papers related to sleep, and 429 studies specifically using NHANES data to investigate sleep and sleep disorders. This analysis reveals a remarkable 1,244% increase in NHANES sleep studies from 2015 to 2024, with research on sleep duration dominating at 32.0% of all investigations, followed by obstructive sleep apnea studies at 16.0%. However, this unprecedented growth demands critical examination, as recent meta-research has raised significant concerns about the quality and methodological rigor of NHANES-based investigations. [Figure 1](#) shows the publication growth trajectory of NHANES sleep research from 2015 to 2025, demonstrating the exponential increase in publications, particularly after 2021.

## The Formulaic Research Problem

Recently, Suchak et al documented what they termed an “explosion of formulaic research articles” based on NHANES data, including inappropriate study designs and false discoveries.<sup>1</sup> Their systematic analysis revealed concerning patterns: studies consistently failing to account for multifactorial relationships, manuscripts neglecting false discovery rate corrections, and researchers selectively extracting data rather than utilizing comprehensive datasets. Their investigation identified an average of four papers per year between 2014 and 2021, which surged to 190 papers in 2024 alone. Ninety-two percent of recent publications originated from Chinese institutions, compared to just 8% before 2021.

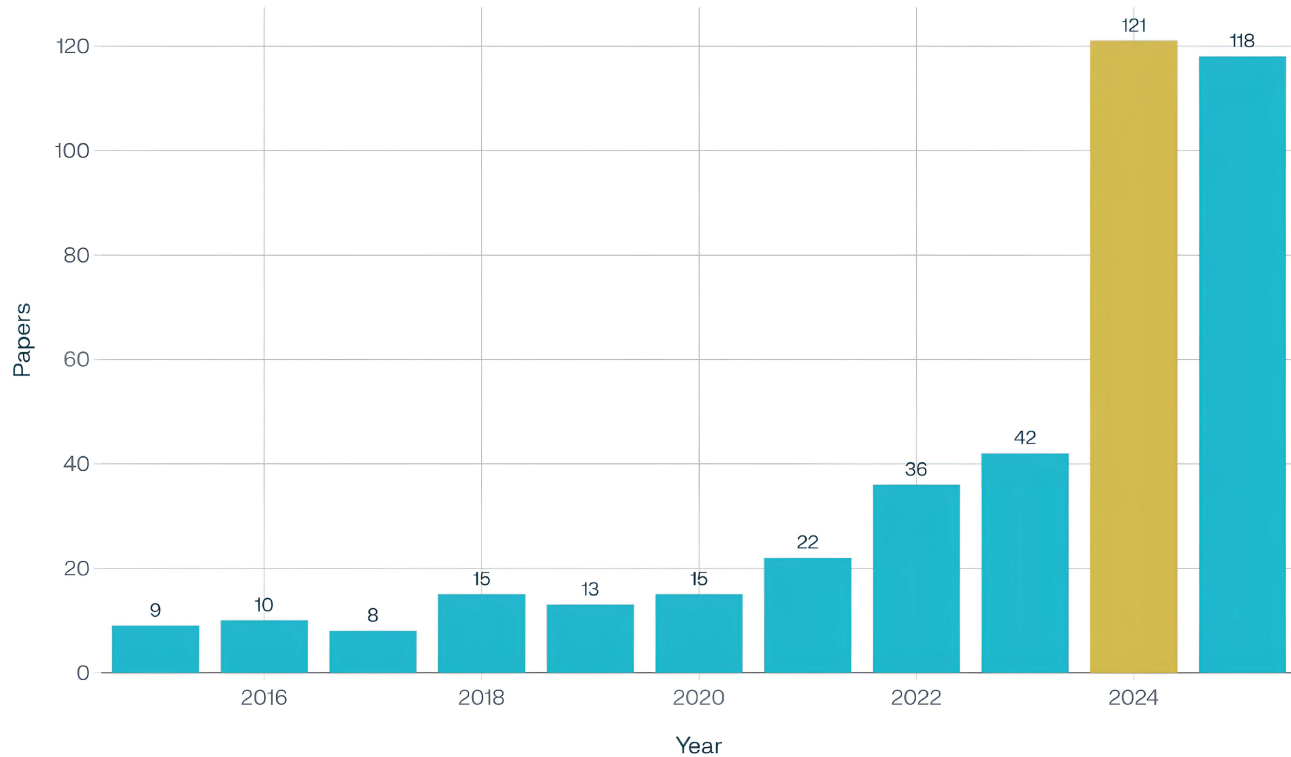
These quality concerns remain largely unaddressed in NHANES sleep research. The implications for sleep medicine research are profound, particularly given that our field has embraced NHANES data with exceptional enthusiasm, producing nearly half of all sleep-related NHANES studies in just the past four years.

The broader scientific publishing landscape has witnessed similar concerning trends across multiple domains. O’Grady reported that, as Northwestern University metascientist Reese Richardson put it, such free data sources allow almost anyone to take a known research method and swap in new variables to create fresh findings in a kind of “research Mad Libs”.<sup>2</sup> This phenomenon extends beyond NHANES to encompass various genetic studies, bibliometric analyses, and other large-scale datasets.<sup>3</sup> The timing coincides suspiciously with the widespread availability of artificial intelligence chatbots that can generate readable text from simple prompts.<sup>4</sup> This could potentially facilitate the rephrasing of template papers with different variable combinations to avoid plagiarism detection.

## Understanding NHANES Appeal and Limitations

The appeal of NHANES data for sleep researchers is understandable. The survey provides nationally representative data with standardized protocols, rich multidimensional variables, and public accessibility that promotes reproducibility. You et al (2024) emphasize that NHANES represents a complex survey design utilizing stratified, multistage probability

## NHANES Sleep Research Growth 2015-2025



**Figure 1** Exponential growth in NHANES sleep research publications (2015–2025). Annual publication counts show a 1,244% increase from 9 papers in 2015 to 121 papers in 2024, with 2024 highlighted in gold to emphasize the publication peak. Data derived from a PubMed bibliometric analysis using search strategy: “(NHANES OR ‘National Health and Nutrition Examination Survey’) AND (sleep OR ‘sleep disorders’ OR ‘sleep medicine’ OR insomnia OR ‘sleep apnea’ OR ‘sleep duration’ OR ‘sleep quality’)”.

sampling to ensure national representativeness across diverse demographic groups, including age, race/ethnicity, and socioeconomic status. The dataset encompasses comprehensive sleep assessments through validated questionnaires, objective measurements like actigraphy in recent cycles, and extensive covariates, including biomarkers, anthropometric measurements, and comorbidity profiles.<sup>5</sup> Sleep duration questionnaires, demographic data, biomarkers, and comorbidity information create an attractive research environment for investigating population-level sleep health relationships. However, it has been cautioned that the apparent simplicity of data access masks substantial methodological complexities that require specialized knowledge of survey weights, variance estimation procedures, and appropriate handling of missing data patterns.<sup>5</sup> The cross-sectional nature of most NHANES components limits causal inference capabilities, while the evolving questionnaire formats across survey cycles demand careful consideration of measurement consistency over time. A particularly concerning trend observed across publicly available big data repositories is that many sleep-related studies are conducted by researchers without specialized sleep medicine training.<sup>3</sup> These investigators may lack the clinical expertise necessary to interpret sleep questionnaire data appropriately, understand the biological plausibility of observed associations, or recognize the limitations of self-reported sleep measures in capturing complex sleep disorders. This knowledge gap often leads to oversimplified interpretations of multifactorial sleep conditions and inadequate consideration of confounding variables specific to sleep pathophysiology. This accessibility has coincided with methodological concerns that threaten the integrity of our scientific literature.

The artificial intelligence revolution has fundamentally transformed how researchers interact with large datasets. Byrne and Stender point out that while AI-ready health datasets have made research more accessible, they can also be misused, leading to a flood of papers that add little genuine value to science.<sup>6</sup> Large language models enable what they describe as “frictionless” manuscript production, where the research process becomes as automated as possible. The

combination of accessible APIs (Application Programming Interfaces) for data extraction and generative AI for manuscript preparation creates unprecedented opportunities for scaling low-value research production.

Our editorial review of recent NHANES sleep submissions reveals recurring patterns that align with broader concerns about formulaic research. Many manuscripts demonstrate a troubling reliance on single-factor analyses for inherently multifactorial sleep disorders. Consider obstructive sleep apnea (OSA), a complex condition influenced by anatomy, obesity, age, genetics, and numerous metabolic factors. Nevertheless, we regularly encounter studies that examine isolated biomarkers or demographic variables as primary predictors, often without appropriate statistical corrections for multiple testing.

The paper mill phenomenon represents perhaps the most insidious threat to scientific integrity when exploiting big data repositories. These covert organizations mass-produce low-quality or fabricated manuscripts for commercial distribution, exploiting the publish-or-perish culture that predicates academic advancement on quantity rather than quality. Mainous describes how paper mills leverage publicly available datasets like NHANES because of their instant credibility, using matrices of variables to generate statistically significant relationships without meaningful hypotheses.<sup>7</sup> These enterprises can overwhelm journals with formulaic submissions that focus on p-values while ignoring biological plausibility or scientific mechanisms. Journal and publishers' policies need systematic approaches to address these challenges, including enhanced transparency requirements for authorship practices, mandatory originality declarations, rigorous multiple testing corrections, and comprehensive AI tool usage disclosure.<sup>8,9</sup> The research community increasingly recognizes that countering paper mills requires coordinated editorial policies that prioritize methodological rigor over publication volume, particularly when evaluating studies utilizing publicly available datasets.

The diagnostic validity issue represents perhaps the most significant methodological challenge. NHANES lacks polysomnographic data, the gold standard for sleep disorder diagnosis. Instead, researchers rely on questionnaire-based assessments that, while useful for screening, have limited specificity for clinical sleep disorders. Researchers commonly employ sleep screening questionnaires without adequate validation or acknowledgment of their limitations. The Berlin questionnaire demonstrates high sensitivity (up to 85%) but poor specificity (as low as 23%), while the STOP-Bang shows higher specificity (up to 68%) but lower sensitivity (as low as 42%) compared to polysomnography.<sup>10</sup> Studies investigating OSA associations often rely on self-reported symptoms that overlap with multiple conditions, creating misclassification bias. This becomes particularly problematic when examining complex relationships without acknowledging that questionnaires may capture symptom clusters rather than confirmed sleep disorders. How many apparent sleep disorder cases actually represent true pathophysiology versus overlapping symptoms?

## Statistical and Design Deficiencies

These diagnostic limitations become particularly problematic when combined with inadequate statistical methodology. Common methodological concerns include the failure to implement false discovery rate corrections, even when testing multiple hypotheses, which dramatically increases the likelihood of spurious associations in large datasets.<sup>1</sup> Studies frequently employ inappropriate reference groups in quartile analyses without statistical justification for cut-point selection, potentially introducing arbitrary categorization bias. Additionally, researchers often neglect comprehensive covariate adjustment that should account for the complex interplay of demographic, metabolic, and environmental factors affecting both sleep exposures and health outcomes.<sup>5</sup> For instance, investigations examining biomarker associations with sleep disorders may fail to consider confounding variables such as age-related changes, dietary patterns, medication use, or comorbid conditions that could explain observed relationships. Strategies for handling missing data and ensuring robust statistical estimates in NHANES analyses require careful consideration of survey design complexities and weighting procedures to maintain population representativeness.<sup>11</sup> The fundamental question remains; how can we ensure that apparent sleep-health associations reflect genuine biological mechanisms rather than methodological artifacts or inadequately controlled confounding?

The selective use of data subsets without clear justification represents another red flag for potentially problematic research. Suchak et al demonstrated that although NHANES diabetes data were available from 1999 to 2020, studies selectively analyzed the periods 2003–2018 and 2017–2020 without a scientific rationale.<sup>1</sup> This pattern suggests data dredging, where researchers systematically explore different variable combinations until statistically significant

associations emerge, a practice that dramatically increases the likelihood of false discoveries. A related concern involves presenting exploratory analyses as confirmatory research, where data-driven findings are reported without acknowledging their hypothesis-generating nature rather than their hypothesis-testing purpose.

The proliferation challenge extends beyond individual methodological issues to broader research integrity concerns in large database studies. Recent analysis has identified troubling patterns where studies employ remarkably similar frameworks, differing only in the specific biomarker or demographic variable examined.<sup>1</sup> This shift toward data mining rather than hypothesis-driven research violates fundamental scientific principles by capitalizing on chance findings instead of testing biologically plausible hypotheses. The accessible nature of comprehensive datasets like NHANES may inadvertently encourage exploratory analyses without appropriate statistical safeguards or theoretical frameworks to guide variable selection.

We need to develop and implement core principles for NHANES data analyses, including understanding dataset limitations, appropriate variable selection, and careful interpretation of cross-sectional findings.<sup>5</sup> However, widespread adoption of these principles remains frustratingly inconsistent across the literature. Far too many studies present associations as though they prove causation, ignore entirely the temporal constraints of cross-sectional data, and fail to situate their findings within the broader context of established sleep medicine knowledge.

## Methodological Challenges in Sleep Research

The widespread availability of artificial intelligence tools has fundamentally altered the landscape of NHANES sleep research, creating both unprecedented opportunities and significant risks for scientific integrity. While AI can accelerate data analysis and manuscript preparation, it may also facilitate the formulaic research patterns observed in NHANES studies.<sup>12</sup> The ease of AI-assisted variable exploration and statistical testing could inadvertently encourage the data dredging behaviors that contribute to spurious associations in large datasets. Instead of merely automating existing approaches, researchers must ensure AI tools enhance methodological rigor rather than expedite the production of low-quality research.

## The Path Forward

At Nature and Science of Sleep, we have become increasingly mindful of these challenges when reviewing papers using NHANES or other large datasets.<sup>13</sup> We encourage our reviewers to watch for red flags like single-factor analyses of complex sleep conditions, missing statistical corrections for multiple testing, and studies that lack clinical relevance or actionable insights. Our goal is simple: support quality science while being more vigilant about studies that present associations as causation or ignore the inherent limitations of cross-sectional data. We believe this balanced approach protects both scientific integrity and genuine researchers who are conducting meaningful work with these valuable datasets. To provide clear guidance for editors and reviewers, we have developed practical safeguards that address the most common methodological pitfalls in NHANES sleep research (Table 1).

Our role as a journal extends beyond gatekeeping to education. We urge investigators to reflect the full complexity of sleep disorders when analyzing NHANES or similar complex surveys, rather than reducing multifactorial conditions to simple exposure-outcome contrasts. Rigor includes principled covariate selection, control of multiple testing (with clearly defined families of hypotheses), and reporting effect sizes with uncertainty, not p-values alone. Claims must be aligned with design; cross-sectional associations should not be framed as causal without explicit assumptions and appropriate methods.

Best practices for NHANES-based studies include: (1) prespecifying variables, time windows, and subgroup plans and justifying exclusions; (2) validating or carefully qualifying sleep-disorder definitions; (3) using survey-aware estimators that incorporate weights, strata, and clusters (including correct re-weighting when combining cycles); (4) harmonizing variables across cycles and documenting assay or coding changes; (5) addressing missing data (eg, multiple imputation) and measurement error; and (6) adopting transparent workflows (code and variable maps shared) and reporting in line with STROBE. Progress here will benefit from coordinated actions by data stewards, publishers, and the research community to enhance guidance and oversight.

**Table 1** Common NHANES Sleep Research Pitfalls and Editorial Safeguards

Methodological Pitfall	Editorial Safeguard
Data mining without a clear hypothesis	<b>Require</b> an explicit a priori question and biological/clinical rationale; if exploratory, label as such and limit claims.
Single-factor analysis of multifactorial conditions	<b>Require</b> an appropriate multivariable framework (confounder control guided by a causal diagram) or revise the scope.
No control for multiplicity	<b>Require</b> a prespecified family of hypotheses and valid multiple-testing control (eg, FDR/FWER) with effect sizes and CIs.
Inadequate covariate strategy	<b>Require</b> theory-informed covariate selection; avoid overadjustment/collider bias; report sensitivity analyses.
Selective dataset/period usage	<b>Require</b> transparent justification for cycle/variable inclusion and exclusion; discourage post-hoc cherry-picking.
Clinical irrelevance	<b>Require</b> clinically interpretable effect sizes (absolute differences, risk ratios, predicted probabilities) and context.
Exploratory work presented as confirmatory	<b>Require</b> clear labeling (hypothesis-generating vs hypothesis-testing) and, when feasible, independent replication.
Ignoring complex survey design (weights/strata/PSUs)	<b>Require</b> survey-aware analyses (design specs in all models); justify any unweighted analyses.
Combining cycles without harmonization/re-weighting	<b>Require</b> documented variable harmonization, attention to assay/code changes, and correct combined-cycle weights.
Misclassification of sleep disorder definitions	<b>Require</b> validation against available clinical indicators or thorough limitation statements, along with sensitivity analyses.
Missing data handled by complete-case only	<b>Require</b> principled handling (eg, multiple imputations, weighting) and reporting of missingness patterns.
Overinterpretation of cross-sectional associations as causal	<b>Require</b> cautious language; explicit assumptions for any causal claims; appropriate methods or reframe as an association.
Small or sparse subgroup analyses	<b>Require</b> minimum cell counts, wide-CI acknowledgment, or defer such analyses to exploratory appendices.
Odds ratios misinterpreted for common outcomes	<b>Require</b> presentation of absolute risks or marginal effects and, when appropriate, risk ratios.

**Abbreviations:** NHANES, National Health and Nutrition Examination Survey; FDR, false discovery rate; FWER, family-wise error rate; CI/CIs, confidence interval[s]; PSU/PSUs, primary sampling unit[s].

The path forward requires collaborative effort across multiple stakeholders. Researchers must embrace higher methodological standards, including comprehensive study design, appropriate statistical analysis, and honest interpretation of limitations. Reviewers need specialized training in both NHANES methodology and sleep medicine principles to effectively evaluate submitted manuscripts. Journals should implement screening procedures that identify potentially problematic submissions while maintaining support for legitimate research.

Educational initiatives represent a critical component of quality improvement. We recommend the development of specialized training programs for big data sleep research, potentially in collaboration with professional societies and other leading journals. These programs should emphasize both the technical aspects of complex survey data analysis and sleep medicine principles necessary for appropriate interpretation.

The remarkable growth in NHANES sleep research over the past decade demonstrates the field's recognition of the value of population-based data for understanding sleep health. However, sustainable growth requires quality alongside quantity. By implementing enhanced methodological standards, appropriate statistical procedures, and careful attention to clinical relevance, we can ensure that NHANES continues to serve as a valuable resource for advancing sleep medicine knowledge while maintaining the scientific integrity that defines our discipline.

## Data Sharing Statement

The dataset obtained from the PubMed search is uploaded as a [supplementary material](#).

## Acknowledgments

The author have no acknowledgments.

## Author Contributions

Ahmed S. BaHamman: Conceptualization, Writing – original draft preparation, Writing – review and editing, and Supervision.

The author gives final approval of the version to be published; has agreed on the journal to which the article has been submitted; and has agreed to be accountable for all aspects of the work.

## Funding

The Strategic Technologies Program of the National Plan for Sciences and Technology and Innovation in the Kingdom of Saudi Arabia, Riyadh, Saudi Arabia (08-MED511-02).

## Disclosure

The author reports no conflicts of interest in this work. Grammarly assisted with grammar correction during the preparation of this manuscript. The figure was prepared with Perplexity AI assistance and author-provided PubMed data, with author validation and oversight.

## References

- Suchak T, Aliu AE, Harrison C, Zwiggelaar R, Geifman N, Spick M. Explosion of formulaic research articles, including inappropriate study designs and false discoveries, based on the NHANES US national health database. *PLoS Biol.* 2025;23(5):e3003152. doi:10.1371/journal.pbio.3003152
- O'Grady C. Low-quality papers surge thanks to public data and AI. *Science.* 2025;388(6749):807–808. doi:10.1126/science.adz1715
- BaHamman AS, Jahrami H. Navigating Mendelian randomization in sleep medicine: challenges, opportunities, and best practices. *Nat Sci Sleep.* 2024;16:1811–1825. doi:10.2147/NSS.S495411
- Bahammam AS, Trabelsi K, Pandi-Perumal SR, Jahrami H. Adapting to the impact of artificial intelligence in scientific writing: balancing benefits and drawbacks while developing policies and regulations. *J Nature Sci Med.* 2023;6(3):152–158. doi:10.4103/jnsm.jnsm\_89\_23
- You Y, Chen Y, Wei M. Leveraging NHANES database for sleep and health-related research: methods and insights. *Front Psychiatry.* 2024;15:1340843. doi:10.3389/fpsy.2024.1340843
- Byrne JA, Stender S. More science friction for less science fiction. *PLoS Biol.* 2025;23(5):e3003167. doi:10.1371/journal.pbio.3003167
- Mainous AG. Papermills as another challenge to research integrity and trust in science. *Front Med.* 2025;12:1557024. doi:10.3389/fmed.2025.1557024
- BaHamman AS. The transparency paradox: why researchers avoid disclosing AI assistance in scientific writing. *Nat Sci Sleep.* 2025;17:2569–2574. doi:10.2147/NSS.S568375

9. Rudan I, Song P, Adeloje D, Campbell H. Journal of Global Health's Guidelines for Reporting Analyses of Big Data Repositories Open to the Public (GRABDROP): preventing 'paper mills', duplicate publications, misuse of statistical inference, and inappropriate use of artificial intelligence. *J Glob Health*. 2025;15:01004. doi:10.7189/jogh.15.01004
10. Hukins C, Duce B. Usefulness of self-administered questionnaires in screening for direct referral for polysomnography without sleep physician review. *J Clin Sleep Med*. 2022;18(5):1405–1412. doi:10.5664/jcsm.9876
11. Pridham G, Rockwood K, Rutenberg A. Strategies for handling missing data that improve frailty index estimation and predictive power: lessons from the NHANES dataset. *Geroscience*. 2022;44(2):897–923. doi:10.1007/s11357-021-00489-w
12. Hoch R, Clarke J. A scientific future shared with AI. *PLoS Biol*. 2025;23(6):e3003274. doi:10.1371/journal.pbio.3003274
13. BaHammam AS, Chee MWL. Publicly available health research datasets: opportunities and responsibilities. *Nat Sci Sleep*. 2022;14:1709–1712. doi:10.2147/NSS.S390292

## Nature and Science of Sleep

### Publish your work in this journal

Nature and Science of Sleep is an international, peer-reviewed, open access journal covering all aspects of sleep science and sleep medicine, including the neurophysiology and functions of sleep, the genetics of sleep, sleep and society, biological rhythms, dreaming, sleep disorders and therapy, and strategies to optimize healthy sleep. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/nature-and-science-of-sleep-journal>

**Dovepress**  
Taylor & Francis Group