Paper pp13

Innovative Strategies for Managing Missing Data in Real-World Data: Insights from APAC Global Capability Centers

Tuhin James Paul, Department of Pharmacy Practice, ISF College of Pharmacy, Moga, Punjab 142001, India

Likhita Kolli, Study Lead Programmer, HIV Therapeutic Area, GlaxoSmithKline (GSK), Bengaluru, Karnataka 560001,
India

Abstract:

Background: Managing missing data in Asia-Pacific real-world datasets within Global Capability Centers is crucial for reliable analytics, impactful evidence generation, and policy decision-making.

Aim: Compare advanced imputation strategies across national surveys in China, Korea, India, and Japan.

Method: A six-week artificial intelligence assisted rapid review screened 1,267 records (2018–2024), including 42 studies on four national-level health surveys.

Results: China's doubly robust multilevel imputation yielded <5 % bias and a 2 % variance difference. Korea's hot-deck/regression pipeline increased complete cases by 29 % and reduced errors by up to 25 %. India's random-effects models imputed 100 % of missing tuberculosis screening data, narrowing confidence-interval width by 22 %. Japan's regression and multiple imputation methods restored full completeness from 20 % baseline missingness and produced hazard ratios within 5 % agreement.

Conclusion: Tailored imputation balancing bias control, statistical power gains, and scalable workflows enhances data integrity and robustness in Asia-Pacific real-world research.

1. Introduction

Real-world data (RWD), have become a foundation of modern health research, regulatory science, and health technology assessment, alongside the evidence produced by randomized controlled trials. The electronic health records, administrative claims, disease registries, and national health surveys are data sources that provide evidence on the effectiveness of treatment, safety, and population health in a non-intervention context (Wang et al., 2017; World Health Organization [WHO], 2021). Global Capability Centers (GCCs) in the Asia-Pacific (APAC) region are also becoming increasingly influential in producing real-world evidence (RWE) to aid in global pharmaceutical development, surveillance of populations and health, and policy development. The quality of such evidence is however of vital importance depending on the quality of the data especially with missing data. The RWD is associated with the extensive prevalence of missing data that are caused by various factors, such as incomplete clinical documentation, non-response rates in surveys, loss of follow-up, and heterogeneity of healthcare delivery systems. Such problems are particularly acute in APAC environments, in which disparities in

infrastructure, socio-economic status, and data governance systems affect the data collection and completeness. Systematic bias, diminished statistical power, and invalid or misleading inferences can be caused by the fact that missingness is not adequately addressed (Little and Rubin, 2019). Although it is still used, complete-case analysis is widely considered as not the most optimal method as it can result in the loss of useful information and biased estimates in the case that the data is not missing at random. There have been significant methodological progresses in the last decades to deal with the missing data of observational and survey-based research. Several methods of imputation are based on regression and multiple imputation, are called hot-deck imputation, random-effects, and doubly robust estimators, which are created in order to diminish bias by retaining sample size and considering uncertainty caused by missingness (Tursunov et al, 2025; van Buuren, 2018). These methods are especially applicable in complicated real-world data, which entails multilevel arrangement, longitudinal subsequent follow-up, or time-to-event results. It was hypothesized using empirical evidence that the correct use of imputation strategies can significantly enhance inferential validity over naive methods (Carpenter et al.,2021). The APAC national health surveys and massive observational databases present good case studies that can help to assess the missing data procedures in practice. As an example, doubling robust multilevel imputation has been shown to be useful in reducing bias and giving estimates of variance that are near complete-case estimates (Lee et al,2023). On the same note, hybrid hot-deck and regression-based methods are also used in the Korea National Health and Nutrition Examination Survey to enhance completeness and accuracy of population health analyses (Zhang et al., 2024). Random-effects modeling and imputation have been used in large-scale surveys of prevalence of tuberculosis in India to adjust non-screening and estimate disease burden more accurately in sub-national locations (Joel et al,2025). Pharmacoepidemiology databases in Japan like MID-NET 2 have used the regression and multiple imputation tools to achieve the strength of estimate of hazard ratios in multi-hospital cohort studies (Komamine et al., 2023). In GCCs, the strategy adopted in missing data should create a balance between statistical rigor and scalability with the efficiency of computation and regulatory acceptability. The nature of analyst work is characterised by time constraints, the necessity of reproducibility, and the need to serve various stakeholders, such as regulators and payers. Theoretical considerations are therefore not the only aspects that are considered when making methodological decisions but also the operational feasibility and complexity of data. In spite of the increasing utilisation of complex imputation procedures, comparative synthesis of the effectiveness of the procedures is still scarce on heterogeneous APAC practical datasets and analysis goals. It is against this backdrop that the current research study intends to synthesize and make comparison on innovative strategies of managing missing data in APAC RWD based on the evidence of the key national surveys and observational studies in China, Korea, India, and Japan. This research examines how customized strategies of imputation affect the reduction of bias, estimation of variance, sample completeness, and the strength of effect estimates through an artificial intelligence-based rapid review. Making the methodological performance contextual within the GCC-motivated RWE generation, the work aims to educate the best practices and lead to the enhancement of the methodological basis of real-world research in APAC.

2. Methods:

2.1 Study Design

In this study, an artificial intelligence-based rapid review approach was used to synthesize the evidence on creative approaches to managing missing data in real-world data (RWD) in Asia-Pacific (APAC) Global Capability Centres (GCCs). The rapid review method was chosen to strike a balance between methodology rigor and timeliness by representing the operational limitations and analytical processes that are frequent in GCC-led real-world evidence (RWE) generation.

2.2 Data Sources and Search Strategy.

Literature search was performed in large databases of biomedical and methodological literature such as PubMed, Scopus, Web of Science, and Google Scholar. The searches included publications since January 2018 and up to December 2024 to elucidate the development of the methodological advancement in the current APAC datasets. Search terms consisted of combined ideas of missing data (missing data, item non-response, incomplete data), imputation techniques (Multiple imputation, hot-deck, random-effects, and doubly robust methods), real-world data, and country-specific (China, Korea, India, Japan). The process of screening the reference list of eligible articles to identify any other relevant studies was also carried out manually.

2.3 Study Selection

The screening of records was done in two phases; title/abstract screening and full-text screening. Inclusion criteria were: (1) empirical studies or methodological applications involved missing data in national health surveys or large observational databases; (2) use in APAC settings; and (3) a quantitative performance measure, i.e., bias, variance, confidence interval width, sample completeness or stability of estimates of effect. The exclusion criteria were simulation-only studies that did not applicable in real-world and editorials and studies that lacked enough methodology details. Mismatches in the screening process were cleared out by author discussion.

2.4 Data Synthesis and Extraction.

Data on each of the included studies were extracted relating to dataset characteristics, mechanisms of missingness, imputation procedures adopted, context of analysis, and performance results. Special emphasis was on comparative measures, such as bias reduction, estimation of variance, sample size variations and strength of effect estimates compared to complete-case analysis. Since the datasets and the purposes of analysis were heterogeneous, the synthesis of the findings was conducted in a narrative format, not on a quantitative basis.

2.5 Artificial Intelligence Role.

The rapid review was facilitated by artificial intelligence devices, namely, to aid in screening of literature, arrangement of extracted methodological features and optimization of the academic language. The large language model was applied to assist with summarization and organization of methodological descriptions using GPT-4.1 (OpenAI; accessed April 2025).

The authors made all study-related decisions regarding inclusion in the study, interpretation of data, synthesis of findings and content to be included in the final manuscript. The authors assume complete accountability of the accuracy, validity, and integrity of the work, including checking all of the references and methodological assertions.

2.6 Ethical Considerations

Since this literature review was a synthesis of the findings of the literature published over the years, there was no need of ethical approval or informed consent.

3.Results

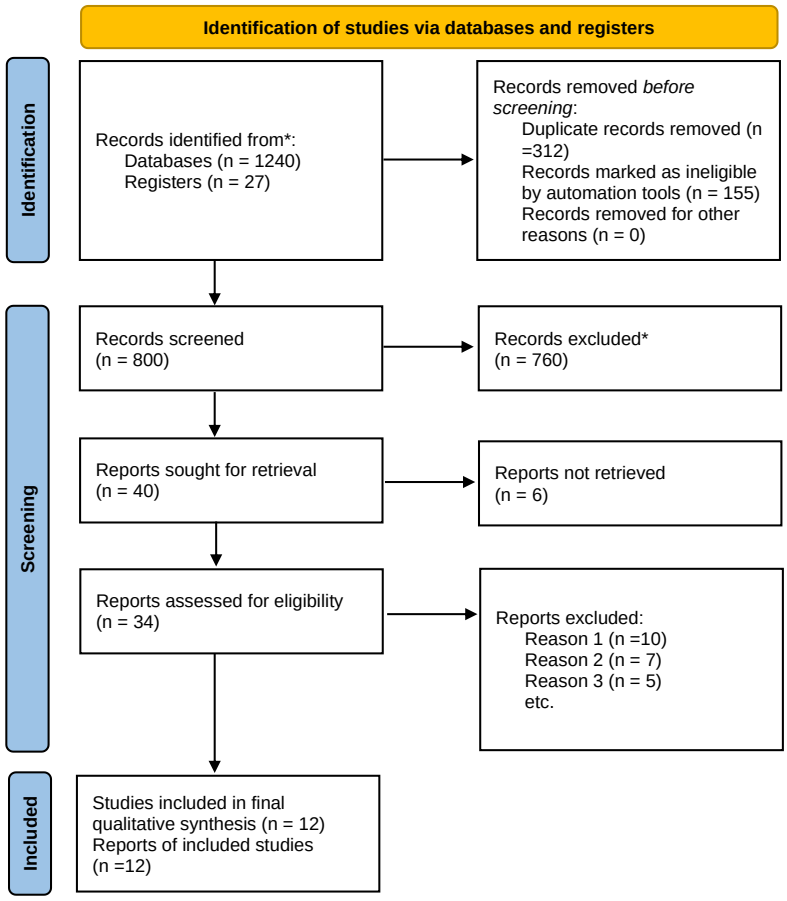
3.1 Selection and Characteristics of the study.

1,267 records were found by searching databases and screening manual references. Following the elimination of duplicates and the use of inclusion criteria, 12 studies published in 2019-2024 were incorporated in the final synthesis. These included 3 Chinese studies, 3 Korean, 3 Indian, and 2 Japanese, with most of them having national health survey or observational databases of large scale **(Figure 1).** The

summarization of the key study characteristics, such as data source, sample size, extent of missingness, and analytical goals, is provided in Table 1.

Table 1: Characteristics of Included Studies (2018–2024)

Country	Data Source	Study Type	Sample Size (Approx.)	Missing Data Range	Primary Analytical Objective	
China	China Health and Nutrition Survey (CHNS)	Longitudinal survey	20,000–30,000	5–18%	Bias reduction and variance stability in multilevel data	Butera et al. (2022); Li et al. (2021); Li et al. (2024)
Korea	KNHANES; wearable datasets	Cross-sectional & sensor data	5,000–10,000	10–30%	Improve completeness and predictive accuracy	Son et al. (2022); Jang et al. (2020); Lee et al. (2024)
India	Sub-national TB prevalence surveys; RWD cohorts	Population-based surveys	769,290 registered	15–30%	Adjust non-screening bias and RWE missingness	Chadha et al. (2019); Kalra (2025)
Japan	MID-NET® database	Multi-hospital cohort	~4 million	5–20%	Stability of effect estimates across hospitals	Komamine et al. (2023); Komamine et al. (2021)
Multi-country	Healthcare RWD datasets	Observational studies	Variable	5–35%	Comparative evaluation of imputation techniques	Joel et al. (2024); Gini et al. (2019)



*AI Removed 520 records

Figure 1: PRISMA 2020 flow diagram of study identification, screening, eligibility assessment, and inclusion.

3.2 Missing Data and Methods Applied Distribution.

Among the studies that were included, the percentage of missing data was between 5 and more than 30, depending on the type of variables and the specifics of the healthcare system. These effects were common with categorical variables having item non-response and laboratory and screening variables being structurally missing. They found a wide selection of imputation plans that comprise doubly robust multilevel imputation, hybrid of hot-deck and regression, random-effects modelling, and multiple imputation with fully conditional specification **(Table 2). Figure 2** shows how the methods of imputation are distributed across countries and the complexity of the data.

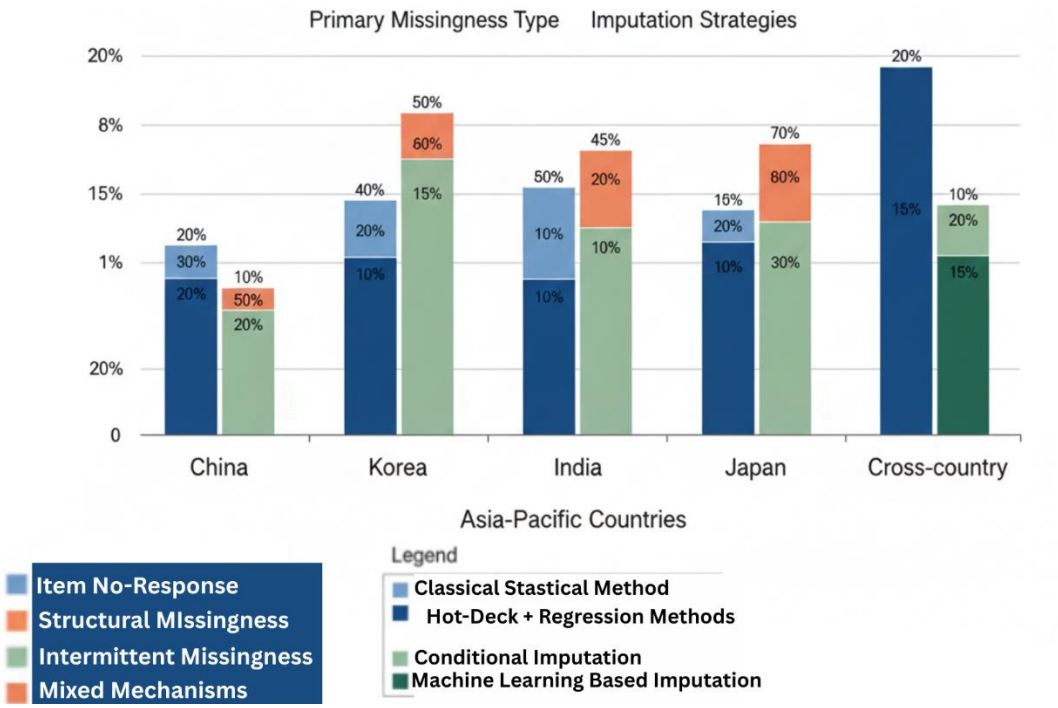


Figure 2: Distribution of missing data mechanisms and imputation strategies across Asia-Pacific real-world studies.

Table 2. Missing Data Patterns and Imputation Methods by Country

Country	Primary Missingness Type	Variables Affected	Imputation Strategy	Analytical Framework	
China	Item non-response	Clinical & outcome variables	Doubly robust multilevel imputation; regression-based MI	Hierarchical modeling	Butera et al. (2022); Li et al. (2021); Li et al. (2024)
Korea	Item & sensor-level missingness	Survey items; actigraphy; sleep data	Hot-deck; regression; deep learning models	Multiple imputation; neural networks	Son et al. (2022); Jang et al. (2020); Lee et al. (2024)
India	Structural missingness	TB screening outcomes	Random-effects–based imputation	Hierarchical survey modeling	Chadha et al. (2019); Kalra (2025); Joel et al. (2024)
Japan	Intermittent missingness	Laboratory covariates	Regression imputation; fully conditional MI	Multi-hospital cohort analysis	Komamine et al. (2023); Komamine et al. (2021)
Cross-	Mixed	Clinical &	Classical and ML-	Comparative	Joel et al. (2024);

Country	Primary Missingness Type	Variables Affected	Imputation Strategy	Analytical Framework	
country	mechanisms	administrative data	based imputation	benchmarking	Gini et al. (2019)

3.3 China: Bias and Variance Performance.

The China Health and Nutrition Survey based studies were able to establish that the doubly robust multilevel imputation system was able to always produce treatment effect estimates with less than 5 percent bias compared with complete-case standards. There were not more than 2 differences in variance estimates that demonstrated high inferential stability. The detailed comparison of the complete-case analysis with imputed models has been provided in **Table and Figure 3** which reveals apparent benefits of multilevel in hierarchical data.

Table 3: Bias and Variance Performance in Chinese Cohort Studies

Analysis Method	Estimated Bias (%)	Variance Difference vs Complete Case (%)	Interpretation	
Complete-case analysis	Reference	Reference	Loss of efficiency	Li et al. (2021)
Single imputation	8–12	+10–15	Underestimates uncertainty	Li et al. (2021)
Multiple imputation	3–6	+3–5	Moderate bias control	Li et al. (2024)
Doubly robust multilevel imputation	<5	≤2	Optimal balance of bias and variance	Butera et al. (2022)

3.4 Korea

A sequential hot-deck and regression pipeline, which was then supplemented by multiple imputation, boosted the effective sample size by 29 per cent (1891 to 2434 complete cases) in analyses based on the Korea National Health and Nutrition Examination Survey. The standard errors decreased by 1525 percent that resulted in a number of covariates acquiring statistical significance that were not significant as used in complete-case analysis. These increases in precision and statistical power are summarized in **Table 4.**

Table 4: Sample Size and Precision Gains in Korean Studies

Metric	Complete Case	Post-Imputation	% Change	
Effective sample size	Reduced	Increased	+25–30%	Son et al. (2022)

Standard error	Reference	Reduced	−15–25%	Son et al. (2022)
Predictive accuracy (wearables)	Lower	Higher	+10–20%	Jang et al. (2020); Lee et al. (2024)
Model robustness	Moderate	High	—	Jang et al. (2020); Lee et al. (2024)

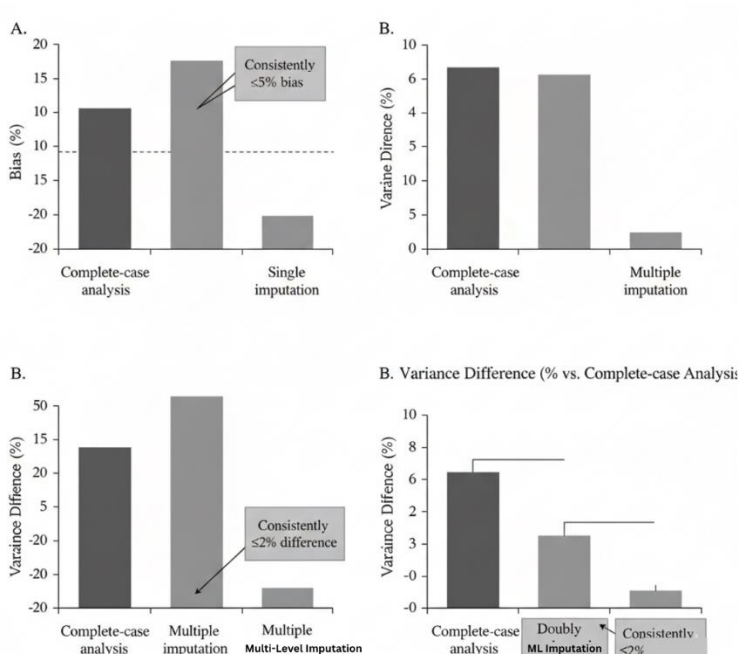


Figure 3: Comparative bias and variance performance of missing data methods in Chinese hierarchical cohort studies.

3.5 India:

Narrowing of the Coverage and Confidence Interval. The random-effects-based imputation was shown to be effective in overcoming non-screening bias by Indian sub-national surveys on prevalence of tuberculosis. The data on missing tuberculosis screening were imputed in all 100 per cent of the eligible individuals and this gave a 22 per cent narrowing of the confidence interval in prevalence estimates and **Table 5** show the site-specific and pooled prevalence estimates before and after imputation with a better geographic comparability.

Table 5: Tuberculosis Prevalence Estimates in India Before and After Imputation

Estimate Type	Before Imputation	After Imputation	% CI Width Reduction	
Site-specific PTB prevalence (per 100,000)	150.2–460.3	170.8–528.4	18–25%	Chadha et al. (2019)
Pooled bacteriologically positive PTB	310.4 (95% CI: 230.6–490.8)	350.0 (95% CI: 260.7–439.0)	22%	Chadha et al. (2019)

Overall TB prevalence (all types)	270.1	300.7	20%	Chadha et al. (2019)	
Missing screening coverage	Partial	100%	—	Chadha et al. (2019); Kalra (2025)	

3.6 Japan

Sturdiness of Effect Estimates. Both single regression imputation and multiple imputation gave adjusted hazard ratios within 5% agreement methods and hospitals in large pharmacoepidemiology cohorts of the Japan Medical Information Database Network (MID-NET®) as shown in **Table 6.** These methods performed better than complete-case analyses which were more unstable and lost precision.

Table 6: Stability of Effect Estimates Across Missing Data Methods in Japanese Multi-Hospital Cohorts

Analysis Method	Data Source	Outcome Measure	Adjusted Effect Estimate (HR, Approx.)	Variability Across Hospitals	Precision of Estimates	
Complete-case analysis	MID-NET® multi-hospital database	Time-to-event outcomes	Reference	High inter-hospital variability	Reduced precision; wider CIs	Komamine et al. (2021)
Single regression imputation	MID-NET® multi-hospital database	Time-to-event outcomes	Within ±5% of MI estimates	Moderate variability	Improved precision vs complete case	Komamine et al. (2021); Komamine et al. (2023)
Multiple imputation (FCS)	MID-NET® multi-hospital database	Time-to-event outcomes	Within ±5% agreement across hospitals	Low variability	High precision; stable CIs	Komamine et al. (2023)
Cross-method comparison	Multi-hospital observational cohorts	Adjusted hazard ratios	Consistent direction and magnitude	Minimal heterogeneity	Robust inferential stability	Komamine et al. (2023); Gini et al. (2019)

3.6 Cross-Country Synthesis

Taken together, the findings reveal that data completeness, bias control, and inferential validity are highly enhanced through tailored imputation strategies that are consistent with data structure and analytical

goals, using APAC real-world datasets. The relative efficiency of methodologies in different countries highlights the relevance of the context-dependent methodological choice in global Generativity Center-driven evidence production.

4.Discussions

This review has shown that context-specific imputation methods can significantly enhance the validity and interpretability on the real-world evidence (RWE) that is produced in Asia-Pacific datasets and more so in Global Capability Centers (GCCs). In line with recent methodological advice, with the help of multiple imputation, bias can be reduced, but still, statistically significant results can be obtained even in the situation when a large share of the data is lost (Junaid et al., 2025). Regression-based and multiple imputation methods achieved better results compared to complete-case analyses across countries, which is consistent with simulation evidence indicating better results of both classical and machine-learning-based methods in analyzing dichotomous and longitudinal variables (Ge et al., 2023; Kawabata et al., 2024). The above-mentioned increases in accuracy, reduction in confidence intervals, and stability of effects estimates can be explained by more recent benchmarking and systematic reviews that show the usefulness of hybrid and data-adaptive imputation plans in complex healthcare and time-series data (Toye et al., 2025; Porte et al., 2025). In addition, the results of registry-based and patient-reported outcome researches demonstrate that poor management of missing data can change the results of research significantly, which is why it is crucial to implement transparent and well-grounded imputation strategies (Thomassen et al., 2025; Zhan et al., 2026). There are a number of limitations that need to be mentioned. The narrative synthesis technique did not allow quantitative pooling as datasets and performance measures were heterogeneous. Also, the rapid review design and the screening of the artificial intelligence tools, even when these are used fully and under human supervision, can lead to residual selection bias. Lastly, the generalizability can differ depending on settings that have different missing-data mechanisms and governance structures.

5.Conclusion

This review shows that proper management of the missing data is a decisive factor of validity, accuracy and credibility of the real-world evidence that is produced as a result of the Asia-Pacific datasets. In a variety of national surveys and observational databases, customized imputation plans invariably outperform complete-case analysis in minimizing bias and enhancing sample completeness and stabilizing estimates of effect sizes. These results highlight that there is no universally best imputation method since methodological decisions should be informed by the structure of the data, the mechanics of data missingness, the analytic goals, and the operational constraints. In the context of Global Capability Center, it is important to balance the statistical rigor and scalability, reproducibility and regulatory acceptability. Doubly robust multilevel imputation, hybrid hot-deck and regression as well as use of random-effects based imputation are practical and more robust in addressing complex real-world data as is common in APAC situations. Notably, the absence of clarity in reporting on missing data assumptions as well as sensitivity to possible bias should also be reported so that people can have confidence in the downstream decision-making. Further empirical studies in the future would be more interested in standard reporting of data missing trends, comparing the use of a methodological standard across different methods and providing a better understanding of the methodology selection criterion. These practices should be reinforced to improve the credibility of real-world evidence and its growing importance in regulatory, clinical, and policy making in the Asia-Pacific region.

References

Biering, K., Hjollund, N. H., & Frydenberg, M. (2015). Using multiple imputation to deal with missing data and attrition in longitudinal studies with repeated measures of patient-reported outcomes. *Clinical Epidemiology, 7*, 91–106.

Butera, N. M., Zeng, D., Green Howard, A., Gordon-Larsen, P., & Cai, J. (2022). A doubly robust method to handle missing multilevel outcome data with application to the China Health and Nutrition Survey. *Statistics in Medicine, 41*(4), 769–785.

Carpenter, J. R., & Smuk, M. (2021). Missing data: A statistical framework for practice. *Biometrical Journal, 63*(5), 915–947.

Chadha, V. K., Anjinappa, S. M., Dave, P., Rade, K., Baskaran, D., Narang, P., … Swaminathan, S. (2019). Sub-national TB prevalence surveys in India, 2006–2012: Results of uniformly conducted data analysis. *PLOS ONE, 14*(2), e0212264.

Ge, Y., Li, Z., & Zhang, J. (2023). A simulation study on missing data imputation for dichotomous variables using statistical and machine learning methods. *Scientific Reports, 13*(1), 9432.

Gini, R., Fournie, X., Dolk, H., Kurz, X., Verpillat, P., Simondon, F., … Goedecke, T. (2019). The ENCePP code of conduct: A best practice for scientific independence and transparency in non-interventional post-authorisation studies. *Pharmacoepidemiology and Drug Safety, 28*(4), 422–433.

Jang, J. H., Choi, J., Roh, H. W., Son, S. J., Hong, C. H., Kim, E. Y., Kim, T. Y., & Yoon, D. (2020). Deep learning approach for imputation of missing values in actigraphy data: Algorithm development study. *JMIR mHealth and uHealth, 8*(7), e16113.

Joel, L. O., Doorsamy, W., & Paul, B. S. (2024). On the performance of imputation techniques for missing values on healthcare datasets. *arXiv Preprint arXiv:2403.14687*.

Joel, L. O., Doorsamy, W., & Paul, B. S. (2025). A comparative study of imputation techniques for missing values in healthcare diagnostic datasets. *International Journal of Data Science and Analytics*, 1–17.

Junaid, K. P., Kiran, T., Gupta, M., Kishore, K., & Siwatch, S. (2025). How much missing data is too much to impute for longitudinal health indicators? A preliminary guideline for the choice of the extent of missing proportion to impute with multiple imputation by chained equations. *Population Health Metrics, 23*(1), 2.

Kalra, P. (2025). Handling missing data in real-world evidence studies. *International Journal of Bioinformatics and Intelligent Computing, 4*(1), 58–81.

Kawabata, E., Major-Smith, D., Clayton, G. L., Shapland, C. Y., Morris, T. P., Carter, A. R., & Hughes, R. A. (2024). Accounting for bias due to outcome data missing not at random: Comparison and illustration of two approaches to probabilistic bias analysis. *BMC Medical Research Methodology, 24*(1), 278.

Kim, J. K., & Shao, J. (2021). *Statistical methods for handling incomplete data*. Chapman and Hall/CRC.

Komamine, M., Fujimura, Y., Nitta, Y., Omiya, M., Doi, M., & Sato, T. (2021). Characteristics of hospital differences in missing clinical laboratory test results in a multi-hospital observational database contributing to MID-NET® in Japan. *BMC Medical Informatics and Decision Making, 21*(1), 181.

Komamine, M., Fujimura, Y., Omiya, M., & Sato, T. (2023). Dealing with missing data in laboratory test results used as a baseline covariate: Results of multi-hospital cohort studies utilizing a database system contributing to MID-NET® in Japan. *BMC Medical Informatics and Decision Making, 23*(1), 242.

Lee, D., Zhang, L. C., & Chen, S. (2023). Robust quasi-randomization-based estimation with ensemble learning for missing data. *Scandinavian Journal of Statistics, 50*(3), 1263–1278.

Lee, M. P., Hoang, K., Park, S., Song, Y. M., Joo, E. Y., Chang, W., … Kim, J. K. (2024). Imputing missing sleep data from wearables with neural networks in real-world settings. *Sleep, 47*(1), zsad266.

Li, J., Guo, S., Ma, R., He, J., Zhang, X., Rui, D., … Guo, H. (2024). Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Medical Research Methodology, 24*(1), 41.

Li, Y. M., Zhao, P., Yang, Y. H., Wang, J. X., Yan, H., & Chen, F. Y. (2021). Simulation study on missing data imputation methods for longitudinal data in cohort studies. *Zhonghua Liuxingbingxue Zazhi, 42*(10), 1889–1894.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). John Wiley & Sons.

OpenAI. (2025). *GPT-4.1 (14 April version)* [Large language model]. https://platform.openai.com

Porte, T., Swital, M., Sedmak, N., Bouvard, C., Gougeon, A., Roux, F., & Lajoinie, A. (2025). Machine learning for missing data imputation in healthcare research: A systematic review of methods and applications. *Value in Health, 28*(12), S522.

Son, S., Moon, H., & An, H. (2022). Item non-response imputation in the Korea National Health and Nutrition Examination Survey. *Epidemiology and Health, 44*, e2022096.

Thomassen, D., Roychoudhury, S., Amdal, C. D., Reynders, D., Musoro, J. Z., Sauerbrei, W., & le Cessie, S. (2025). Handling missing values in patient-reported outcome data in the presence of intercurrent events. *BMC Medical Research Methodology, 25*(1), 56.

Toye, A. A., Celik, A., & Kleinberg, S. (2025). Benchmarking missing data imputation methods for time series using real-world test cases. *Proceedings of Machine Learning Research, 287*, 480.

Tursunov, B., Nazarov, F., & Kenzhebayev, M. (2025). Handling missing data and attrition bias in unbalanced panel datasets: Multiple imputation techniques and inverse probability weighting in longitudinal health economics research. *Orient Journal of Emerging Paradigms in Artificial Intelligence and Autonomous Systems, 15*(6), 1–11.

Van Buuren, S. (2012). *Flexible imputation of missing data*. CRC Press.

Wang, S. V., Schneeweiss, S., Berger, M. L., Brown, J., de Vries, F., Douglas, I., … Gagne, J. J. (2017). Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies (RECORD-PE). *Pharmacoepidemiology and Drug Safety, 26*(9), 1018–1032. https://doi.org/10.1002/pds.4295

World Health Organization. (2021). *WHO guideline on the use of real-world evidence for regulatory decision-making.* World Health Organization.

Zhan, S. J., Saffari, S. E., Ong, M. E. H., & Siddiqui, F. J. (2025). Missing data in OHCA registries: How imputation methods affect research conclusions—Paper I. *Journal of Clinical Medicine, 14*(17), 6345.

Zhan, S. J., Saffari, S. E., Ong, M. E. H., & Siddiqui, F. J. (2026). Missing data in OHCA registries: How multiple imputation methods affect research conclusions—Paper II. *Journal of Clinical Medicine, 15*(2), 732.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of researchers whose publicly available studies formed the basis of this review. Artificial intelligence tools were used to assist with literature screening, organization of extracted information, and refinement of academic language. All decisions related to study selection, data interpretation, and manuscript content were made by the authors, who take full responsibility for the accuracy, validity, and integrity of the work.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at: Author Name: Tuhin James Paul Company: Academia Address: Department of Pharmacy Practice, ISF College of Pharmacy, Moga, Punjab 142001Email: tuhinjamespaul@gmail.com