Proceedings of CONF-SEML 2026 Symposium: Multimodal Data Acquisition: Applications in Physiological and Behavioral Research
DOI: 10.54254/2755-2721/2026.AD32261

Applications of Bayesian Statistics in Medicine

Xiangyu Zha

Reading College, Nanjing University of Information Science and Technology, Nanjing, China
15850335111@163.com

Abstract. In recent years, the increasing complexity of clinical decision-making in modern medicine, together with the inherent uncertainty in diagnosis and prognosis, has required the adoption of more advanced statistical methodologies. Against this background, Bayesian statistics has been progressively introduced into clinical research and practice as a flexible framework for the analysis of diagnostic and therapeutic data. Accumulating evidence indicates that Bayesian methods have evolved into an effective and robust analytical tool for medical research and clinical treatment. Based on the characteristics of medical research, this study first outlines the fundamental concepts and theoretical framework of Bayesian statistics. Subsequently, through literature review, comparative analysis, and data analysis, the study examines the methodological advantages of Bayesian approaches and clarifies their significance for future medical development. Specific applications of Bayesian statistics across different medical domains are further discussed to demonstrate their effectiveness and practical value. The findings suggest that, owing to its distinctive probabilistic framework, Bayesian statistics exhibits notable advantages in multiple medical fields and is particularly well suited to addressing uncertainty and complexity in medical data. With the continuous growth of healthcare data and ongoing advances in computational technology, Bayesian methods are expected to play an increasingly important role in precision medicine and personalized treatment, thereby providing solid theoretical support for both medical research and clinical practice.

Keywords: Bayesian statistics, Medical research, Clinical decision-making, Uncertainty

1. Introduction

In medical research, the selection of appropriate and scientifically sound statistical methodologies is essential. With the substantial increase in global disease incidence in recent years, medical researchers urgently require rigorous statistical approaches to support disease prevention, diagnosis, and treatment. Bayesian statistics, derived from Bayes' theorem, is a statistical inference method that updates posterior distributions by combining prior distributions with observed data. Owing to this mechanism, it has significant application value in medical research.

Bayesian methods offer several advantages, including flexible handling of small-sample data, integration of multi-source information, and explicit quantification of uncertainty. As a result, they have gradually become important tools in clinical trials, disease prediction, and the construction of diagnostic models. Nevertheless, challenges remain, particularly in terms of computational

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AD32261

complexity and data processing requirements. This study focuses on the application of Bayesian methods to the optimization of medical diagnostic models. Markov Chain Monte Carlo (MCMC) and variational inference techniques are employed, and real clinical data are incorporated for comparative analysis. In addition, literature review and synthesis methods are used to support the investigation. The results are expected not only to improve the accuracy of medical decision-making but also to provide methodological references for the application of Bayesian statistics to complex medical problems.

2. Overview of Bayesian models

2.1. Background of Bayes' theorem

Bayes' theorem describes the conditional probability relationship between two random events, A and B. Its basic form is given by: P(A|B)=P(B|A)P(A)/P(B). In statistics, Bayes' theorem provides the theoretical foundation for updating the probability of a hypothesis after new evidence is obtained. The theorem is named after the 18th-century British mathematician Thomas Bayes, and its original ideas were later further developed and widely applied, particularly in modern statistical inference.

By incorporating prior knowledge and observed data, Bayesian methods establish a framework for dynamically updating probability estimates in complex problems. This flexibility enables Bayesian approaches to accommodate diverse practical requirements in medical research. A complete Bayesian analysis typically involves data processing, construction of a probabilistic model, specification of prior information and likelihood functions, and final decision-making [1].

The core principle of Bayesian inference is to combine sample information related to unknown parameters with prior information and to derive the posterior distribution according to Bayes' theorem, thereby enabling further inference regarding the unknown quantities. In a standard Bayesian inference problem, let the parameter to be inferred be denoted by θ and the observed sample by x. The fundamental Bayesian formula is expressed as$\pi(\theta|x) = [\pi(\theta)\ P(x\ |\theta)\ ]/[\int\theta\pi(\theta)P(x|\theta)]$Here, $P(\ x|\theta)$represents the likelihood function based on the sample information, x denotes the observed data,$\pi(\theta)$ is the prior distribution, θis the parameter of interest, and$\pi(\theta|x)$ denotes the posterior distribution [2].

2.2. Overview of Bayesian models

Bayesian models constitute a class of probabilistic models developed on the principles of Bayesian statistics. Their core concept is to conduct statistical inference by integrating prior information with observed data. Within this framework, model parameters are treated as random variables, and probability distributions are used to characterize their uncertainty, thereby enabling the modeling and analysis of complex systems.

In terms of model construction, a Bayesian model generally consists of three key components: the prior distribution, the likelihood function, and the posterior distribution. The prior distribution reflects the researcher's knowledge or assumptions about the parameters before observing the data. The likelihood function describes the data-generating process. The posterior distribution is obtained by combining the prior distribution with the observed data, resulting in an updated estimate of the parameters. Because this framework effectively mitigates the limitations associated with insufficient data, it is particularly suitable for small-sample studies.

With respect to data interpretation, Bayesian models primarily involve two essential elements: the specification of prior distributions and the computational methods used for posterior estimation.

DOI: 10.54254/2755-2721/2026.AD32261

Parameter estimation can be carried out using two main approaches: empirical Bayes and fully Bayesian methods, with the latter generally providing more comprehensive parameter inference. By incorporating structures such as autocorrelation and spatial variance information into the modeling process, Bayesian models have facilitated substantial progress in spatial data analysis.

2.3. Comparison between Bayesian and traditional statistical methods

Traditional statistical methods are typically based on null hypothesis significance testing (NHST). Decisions are made by calculating the p-value to determine whether the null hypothesis (H_0) should be rejected. The p-value is defined as the probability of observing the current data, or more extreme results, under the assumption that H_0 is true. However, p-values are often misinterpreted as the direct probability that H_0 is false, and they do not quantify the strength of evidence in favor of an alternative hypothesis (H_1).

In contrast, Bayesian statistics employs the Bayes factor (BF) to directly compare two competing models (e.g., M_1 versus M_0) through the ratio of their posterior probabilities to prior probabilities [3]. The Bayes factor not only provides quantitative evidence regarding H_0 but also clearly indicates the relative strength of support for one model over another. For example, a BF of 10 indicates that the observed data increase the posterior odds of M_1 relative to M_0 by a factor of ten. Such an interpretation aligns more closely with the intuitive logic of scientific inference.

From a theoretical perspective, traditional methods adopt a frequentist interpretation of probability and emphasize long-run frequency properties under repeated sampling. Bayesian methods, by contrast, are grounded in subjective probability theory and allow prior knowledge to be formally incorporated into the analysis, enabling dynamic updating of posterior beliefs. As a consequence, traditional statistical approaches are generally more suitable for standardized experimental settings, whereas Bayesian methods offer clear advantages in small-sample studies, complex modeling scenarios, and situations requiring the integration of historical data.

In medical research, Bayesian frameworks support adaptive trial designs, allowing sample sizes or intervention strategies to be adjusted dynamically as data accumulate during the trial. For instance, in oncology clinical trials, if interim results indicate a significant therapeutic effect, researchers may expand the sample size or terminate an ineffective control group earlier, thereby improving resource utilization and reducing patient exposure to ineffective treatments.

Conventional p-value–based testing often leads to situations described as "statistically significant but clinically irrelevant" or "statistically non-significant but clinically important." Bayesian methods address this limitation by combining the region of practical equivalence (ROPE) with the highest density interval (HDI) to directly quantify the relationship between parameter estimates and clinically meaningful thresholds. For example, if the 95% HDI of the mean efficacy of an antihypertensive drug lies entirely within the ROPE (e.g., ±2 mmHg), the treatment effect may be considered clinically negligible even when $p < 0.05$, thereby preventing overtreatment.

Furthermore, medical research frequently requires the integration of historical data, real-world evidence, and expert knowledge. Bayesian models enable the systematic incorporation of multi-source information through the specification of appropriate prior distributions. In the development of orphan drugs, for example, researchers can construct weakly informative priors based on published evidence, thereby substantially reducing the required sample size and accelerating the drug approval process.

On the basis of these considerations, Bayesian statistics demonstrates clear methodological advantages over traditional statistical approaches in medical applications.

DOI: 10.54254/2755-2721/2026.AD32261

3. Applications of Bayesian statistics in medicine

3.1. Applications of Bayesian statistics in the treatment of lung cancer

Lung cancer is the second most commonly diagnosed cancer worldwide and remains the leading cause of cancer-related mortality. According to statistical data, in 2020, there were approximately 2.2 million new lung cancer cases and about 1.8 million deaths globally, with China accounting for nearly 37% of newly diagnosed cases. The current situation in China is particularly severe, characterized by both high incidence and high mortality. In addition, marked sex differences are observed in both incidence and mortality, with male mortality substantially exceeding that of females.

To address these challenges, researchers have employed various statistical models to investigate clinical treatment strategies for lung cancer. Among numerous quantitative research approaches, Bayesian statistical methods have been widely adopted due to their methodological advantages. In lung cancer research, different Bayesian models serve distinct analytical purposes. For example, Bayesian spatiotemporal models can effectively analyze temporal and spatial information embedded in epidemiological data, thereby facilitating the investigation of associations between disease occurrence and related factors. Such models provide scientific support for disease prediction and decision analysis, leading to more accurate risk estimation.

Dynamic Bayesian network models represent another class of probabilistic graphical models capable of characterizing time-varying relationships among variables. These models can appropriately handle uncertainty and incomplete data while explicitly representing potential causal relationships. Over the past decade, Bayesian hierarchical spatiotemporal models have been extensively applied in disease mapping and epidemiological analysis. By estimating smoothed risk surfaces after adjusting for covariates, these models reveal spatial patterns of disease and enable further exploration of covariate effects.

In contemporary medical research in China, a variety of Bayesian statistical models have been introduced. Their fundamental principle—updating probability inference through the integration of prior knowledge and current data—has promoted the broader application of Bayesian methods in medical studies. In the context of precision treatment for lung cancer, Bayesian statistics is playing an increasingly important role. Its primary value lies in the dynamic integration of heterogeneous evidence and in providing quantitative probabilistic support for individualized treatment decisions [4].

3.2. Applications of Bayesian statistics in pediatric drug development

In pediatric drug development, a critical scientific and regulatory issue is how to safely and effectively utilize data from adult clinical trials to support efficacy evaluation in pediatric populations. Traditional frequentist statistical methods often encounter limitations in pediatric clinical trials, where sample sizes are typically very small, resulting in insufficient statistical power. In contrast, Bayesian statistics offers a more flexible framework by formally incorporating historical information from adult studies as prior knowledge into the statistical models for current pediatric research [5].

The principal advantage of this approach is that it maintains scientific rigor while substantially reducing the number of pediatric participants required in clinical trials. This is particularly consistent with ethical principles in pediatric research, which emphasize minimizing children' s exposure to unnecessary experimental risks to the greatest extent possible.

DOI: 10.54254/2755-2721/2026.AD32261

In practical applications, researchers may adopt different prior construction strategies to integrate external data. For example, a "power prior" approach can be employed, in which a weight parameter $a_0 \in [0,1]$ is introduced to regulate the degree of influence of historical data on the current analysis [6]. When$a_0=1$, the historical data are fully trusted; when $a_0=0$, the analysis relies exclusively on the current pediatric data, effectively disregarding historical information.

Another commonly used approach is the meta-analytic predictive prior (MAP prior), which synthesizes multiple adult studies through Bayesian random-effects models to generate a predictive prior distribution. This distribution serves as an initial estimate of treatment effects in the target pediatric population. To enhance model robustness and mitigate potential bias arising from differences between adult and pediatric populations, researchers may further employ a robust MAP prior. This approach combines an informative prior with a weakly informative or non-informative prior, thereby ensuring that the analysis is not unduly influenced when historical and current pediatric data are not fully consistent.

Beyond efficacy evaluation, Bayesian methods also demonstrate considerable potential in other aspects of pediatric drug development. During early dose-finding stages, pharmacokinetic (PK) and pharmacodynamic (PD) data from adult studies can be incorporated as prior information to derive more precise initial dosing regimens for pediatric patients. In exposure–response (ER) analyses, multivariable models can be constructed to extrapolate adult exposure–response relationships to pediatric populations while formally assessing their similarity.

A representative example is a pediatric multiple sclerosis study included in the U.S. FDA's Complex Innovative Trial Design (CID) pilot program. In this study, data from six historical trials, involving more than 2,000 participants, were integrated to construct a robust MAP prior. As a result, the projected sample size for the planned pediatric clinical trial was reduced by approximately 30%, substantially alleviating the challenges associated with recruiting pediatric participants for rare diseases.

4. Conclusion

This study systematically examined the application value of Bayesian statistics in modern medical research, with particular emphasis on its methodological advantages and practical implementation in lung cancer treatment and pediatric drug development. The findings indicate that, by flexibly integrating prior knowledge with current data, Bayesian methods effectively address challenges such as small sample sizes, high uncertainty, and multi-source data integration. In doing so, they provide a more dynamic and interpretable probabilistic framework for clinical decision-making.

The analysis demonstrates that Bayesian statistics introduces new perspectives into medical research. In lung cancer studies, it enables more precise investigation of disease patterns; in pediatric drug development, it facilitates ethically sound utilization of adult data and reduces the risks borne by pediatric participants. Nevertheless, certain limitations remain. For instance, the discussion does not sufficiently elaborate on how to determine the appropriate influence of prior information in practical applications. In addition, the study lacks detailed real-world case analyses illustrating how Bayesian methods address complex scenarios encountered in clinical practice.

Future research may proceed along several directions. First, standardized and robust criteria for prior specification should be established to promote broader implementation in clinical settings. Second, further exploration of the integration of Bayesian statistics with emerging technologies, such as artificial intelligence, is warranted to address increasingly complex medical big data. Third, expanded applications in real-world studies and personalized medicine should be pursued to fully realize the potential of Bayesian approaches in precision healthcare.

DOI: 10.54254/2755-2721/2026.AD32261

References

[1] Liu LP, Yuan W. The application of modern Bayes method to accounting and its prospect [J]. Statistical Research, 2002, 2002 (08): 45-49.
[2] Yue JF. Surveys on the application of bayesian method on actuarial study [D]. Jilin: Jilin University, 2009.
[3] Wang Jing, Bao Gui. Comparison Between Bayesian Statistics and Traditional Statistical Methods, 2021.
[4] Yi Jie, Guo Chunchun, Zhuang Jinman, Xu Xin, He Fei, Application of bayesian models in studies related to factors influencing lung cancer incidence and mortality, 2024,
[5] Li Ruobing, Yan Fangrong, Wang Jun. Bayesian Statistics for Extrapolation of Efficacy in Pediatric Medicine, 2022.
[6] Ibrahim J G, Chen M H． Power prior distributions for regression models [J], Stat Sci, 2000, 15(1): 46-60.