Proceedings of CONF-CIAP 2026 Symposium: International Conference on Atomic Magnetometer and Applications
DOI: 10.54254/2753-8818/2026.HZ27847

Bayesian and Frequentist Approaches in Utilizing Historical Data

Yuehan Huang

Beijing National Day School, Beijing, China
huangyuehan.spring@gmail.com

Abstract. Incorporating historical data in current studies becomes a crucial method due to the need to increase statistical power and reduce costs. Yet, historical data present challenges such as limited sample sizes and a lack of guidelines for method suitability based on their types. This paper provides a comprehensive review of Bayesian, Frequentist, and empirical Bayes approaches for integrating historical data into research. The Bayesian approach incorporates prior knowledge through formal distributions, offering flexibility but introducing subjectivity. The Frequentist approach treats historical data as additional samples, prioritizing objectivity but requiring larger samples and homogeneity assumptions. Empirical Bayes, serving as a hybrid approach, estimates priors from data to strike a balance between flexibility and objectivity. The review compares these approaches in terms of philosophical foundations, data requirements, flexibility, computational demand, and application domains, concluding that method selection should be guided by data characteristics, prior knowledge availability, and practical constraints. Furthermore, this review proposes practical guidelines for researchers choosing among the three approaches.

Keywords: Bayesian, Frequentist, Empirical Bayes, historical data.

1. Introduction

Data is the basis of almost all the application-related research. Since data collection is essential but costly, the utilization of historical data is a critical strategy in order to reduce cost and maintain validity. In clinical research, historical data offer control groups that could improve trial efficiency. In oncology trials, incorporating historical data can reduce patient enrollment needs while maintaining statistical validity. In economics, historical datasets allow researchers to analyze without new large-scale surveys. In ancient history and archaeology, where new data collection is impossible, historical data are not just beneficial but irreplaceable. In an era where data-driven decision-making dominates fields from healthcare to social sciences, the efficient use of existing datasets has become a methodological imperative. However, compared to modern experimental data, historical data sometimes suffers from a relatively limited sample size, incomplete records, and measurement procedures that are different from those used now. Failure to address these issues may lead to biased estimates, reduced statistical power, or even spurious conclusions. Therefore, it's crucial for researchers to remind themselves that they need to consider the effects that those

© 2025 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: 10.54254/2753-8818/2026.HZ27847

characteristics have on the results of their study, which raises methodological challenges for effectively incorporating historical data into their study.

With the evolution of methodology, there exist two paradigms, the Bayesian and the Frequentist approaches, that address these challenges. This dichotomy reflects a fundamental philosophical divide in interpreting probability, where the Bayesian paradigm embraces the idea that probability is a degree of belief, while the Frequentist paradigms adhere to the idea that probability is the proportion of an event's occurrence when repeated infinitely.

Which approach is better while utilizing historical data, Bayesian or Frequentist? Are there other alternatives? The choice between these paradigms is far from trivial. The idea of combining them is also worth noting. This paper aims to offer a general review of the methods that Bayesian and Frequentist use while incorporating historical data, compare the advantages and disadvantages between Bayesian and Frequentist, and evaluate them through multiple aspects.

Additionally, despite the theoretical and operational differences, Bayesian and Frequentist approaches are not completely mutually exclusive. In fact, there is a hybrid approach called Empirical Bayesian. It aims to combine Bayesian and Frequentist approaches. In other words, it tries to find the balance between the flexibility of Bayesian and the objectivity of Frequentist. This approach will also be considered in this paper.

This review seeks to synthesize research on the use of Bayesian and Frequentist approaches in historical data analysis, compare the differences between them in terms of applicability and limitations, and discuss the potential of a hybrid approach. By organizing this relevant review, this paper hopes to provide some guidance for researchers to choose suitable approaches while analyzing historical data.

The paper is organized as follows. The theoretical background of Bayesian and Frequentist is introduced in Section 2, along with a literature review of the main methods used in each paradigm. Subsection 2.1 will detail Bayesian techniques, while Subsection 2.2 will cover Frequentist methods. In Section 3, there is the methodology used when writing this review. The results of the comparison within the three approaches are in Section 4. Finally, discussions and conclusions are provided in Section 5, including recommendations for method selection and future research directions.

2. Theoretical background

A comprehensive understanding of the paper's theme requires a mention of the conception of probability, since the way historical data are incorporated is deeply tied to the underlying interpretation of probability. There are mainly two perspectives that offer reasonable interpretations about probability: the Bayesian paradigm and the Frequentist paradigm.

2.1. The Bayesian approach

The Bayesian perspective views probability as a degree of belief, given the knowledge one has. Since the acquisition of knowledge varies among people and over time, probability is therefore considered to be personal and subjective from a Bayesian perspective. The phrase "personal" is relatively trivial to understand, but it seems radical to say that probability is subjective. As explained by Myles Lavan, the uncertainty in a random process is caused by the epistemic uncertainty about deterministic systems [1]. In other words, what appears random is often the result of processes governed by fixed laws but exhibiting deep sensitivity to initial conditions, rendering precise long-term prediction impossible without perfect knowledge. For instance, the results of a coin toss, from

DOI: 10.54254/2753-8818/2026.HZ27847

a Bayesian's perspective, are determined by air resistance, surface texture, and many other microscopic factors that are not measured or controlled. Bayesian probability is used to formalize the knowledge-dependent uncertainty. As for the Frequentist's perspective, it implicitly relies on subjective assumptions about the generating mechanism (e.g., presuming that the coin is "fair"). It is ultimately based on the subjective judgement of the observer and therefore can't be called objective.

The foundation of Bayesian is Bayes' Theorem:

$$p(\theta \mid y) \propto p(y \mid \theta) \cdot p(\theta) \tag{1}$$

for which $p(\theta \mid y)$ is the posterior probability of parameters θ given data y; $p(y \mid \theta)$ is the likelihood of observing data y under parameter θ; and $p(\theta)$ is the prior probability.

The Bayesian approach incorporates prior distribution in the inference process, dealing with historical data as existing knowledge. This is implemented through formal prior elicitation, where historical data are transformed into probability distributions that constrain or inform parameter estimates.

Furthermore, different models are used when dealing with different data.

Bayesian prevalence estimation, which incorporates uncertainty in diagnostic test accuracy, outperformed traditional methods by avoiding truncation issues and providing well-calibrated credible intervals.

Bayesian hierarchical modeling provides a robust framework for multiplicity adjustment in trials with multiple endpoints by leveraging shrinkage effects to improve power while maintaining error control at the same time [2, 3].

2.1.1. Power prior

In the comprehensive review published in the Journal of Data Science, Chen et al. systematically trace the development and applications of power prior [4]. Power prior adjusts the weight of historical data by raising the likelihood function of historical data to a power parameter a0($0 \leq a_0 \leq 1$):

$$\pi(\theta \mid D_0, a_0) \propto L(\theta \mid D_0)^{a_0} \pi_0(\theta) \tag{2}$$

where D_0 is historical data, and $\pi_0(\theta)$ is the initial prior. When a_0=0, the historical data is completely disregarded; when a_0=1, the historical data is fully used.

2.1.2. Commensurate prior

Hobbs et al. introduced the commensurate prior in their seminal Biometrics paper. The commensurate prior accounts for potential conflicts with current trial data, which is a critical challenge in leveraging historical data [5]. (Traditional power priors rely on a fixed discounting parameter (a_0) to downweigh historical data, but this approach lacks flexibility when the degree of compatibility between historical and current datasets is uncertain.)

The commensurate prior resolves this by introducing a commensurability parameter τ, which dynamically quantifies the consistency between historical parameter θ_0 and current parameter θ. In its simplest form, the prior for θ is centered at θ_0 with precision τ:

$$\pi\left(\theta \mid \theta_0, \tau\right) \propto N\left(\theta \mid \theta_0, \tau^{-1}\right) \tag{3}$$

DOI: 10.54254/2753-8818/2026.HZ27847

As $\tau \to 0$, the prior variance of θ diverges, effectively ignoring historical data when evidence of incompatibility is strong. On the other hand, as $\tau \to \infty$, the prior enforces $\theta \approx \theta_0$, fully borrowing historical information when datasets are highly consistent.

2.1.3. Bayesian Hierarchical Model (BHM)

The BHM introduces two critical parameters to model heterogeneity: prognostic heterogeneity and predictive heterogeneity [5,6].

Prognostic heterogeneity (baseline differences between trials): Captured by a random effect $r \sim N(0, \tau^2)$ where τ controls the degree of borrowing. Smaller τ assumes higher similarity in baseline responses.

Predictive heterogeneity (treatment effect differences): Modeled via a scaling parameter $a \sim N_{[b,1/b]}(1, \eta^2)$, which adjusts the historical treatment effects to align with the current trial. A truncated normal prior ensures realistic bounds.

This hierarchical structure allows the model to adaptively downweigh historical data when discrepancies exist.

2.2. The frequentist approach

How Frequentists understand probability is relatively trivial to understand. Probability is understood objectively, as the relative frequency with which an event occurs when an experiment is repeated under identical conditions. Probability in this case is viewed as a property. Grounding in this idea, Frequentists' approaches treat historical data as additional samples and integrate them into current datasets using variance, weighting schemes, meta-analytic techniques, and so on. These methods often rely on homogeneity testing or propensity score matching to ensure compatibility between historical and contemporary data.

Davidov and Rudas proposed a Frequentist framework for incorporating historical estimates into current analyses, addressing uncertainty through asymptotic theory and optimal study design [7].

2.2.1. Meta-analysis

Being utilized by Jelena Ivetić et al., the meta-analysis derives a pooled estimate of the treatment effect [8]. Based on different assumptions about the heterogeneity across studies, two models are derived: the fixed-effect model (FEM) and the random-effect model (REM).

The FEM assumes that all studies included in the meta-analysis are estimating a single, common true effect size. In this case, any observed variability between the results of the studies is attributed exclusively to sampling error. Consequently, the weight w_i assigned to each study in the pooling process is determined solely by the precision of its individual estimate, which is a function of its within-study variance σ_i^2.

The pooled effect estimate $\hat{\theta}_{FE}$ is calculated as the variance-weighted average of the individual study effects θ_i:

$$\hat{\theta}_{FE} = \frac{\sum w_i \theta_i}{\sum w_i}, \ where \ w_i = \frac{1}{\sigma_i^2} \tag{4}$$

The REM acknowledges that the true treatment effect may vary across studies. It assumes that the included studies are a random sample from a larger population of potential studies, and their individual true effects are distributed around an average overall effect. Thus, the observed variance

DOI: 10.54254/2753-8818/2026.HZ27847

combines two components: the within-study variance σ_i^2 and the between-study variance τ^2, which represents the heterogeneity.

The weight assigned to each study w_i in a REM is modified to incorporate this additional source of variation, and the pooled effect estimate, $\hat{\theta}_{RE}$, is then computed using these adjusted weights:

$$\hat{\theta}_{RE} = \frac{\sum w_i \theta_i}{\sum w_i}, \ where \ w_i = \frac{1}{\sigma_i^2 + \tau^2} \tag{5}$$

2.2.2. Weighted likelihood

The core idea of weighted likelihood is to assign a weight w_i to each observed data point i, modifying the standard likelihood function to reflect the varying importance or inclusion probability of data points. The general form of the weighted likelihood function can be expressed as:

$$L_w(\theta \mid D) = \prod_{i=1}^{n} f(y_i|x_i, \theta)^{w_i} \tag{6}$$

where D is the observed data, f is a probability density or mass function, and w_i is a pre-specified or estimated weight for the i-th observation. When all weights $w_i = 1$, the method reduces to standard maximum likelihood estimation. (The introduction of weights w_i allows the model to adapt to complex sampling, missing data, or treatment assignment mechanisms, thereby yielding more robust and less biased estimates of population parameters.)

2.2.3. Propensity Score Matching (PSM)

The core concept of PSM aims to approximate the conditions of a randomized trial by constructing a matched analytical sample where treated and control units share similar probabilities of receiving the treatment [9]. This probability, the propensity score denoted as e(X), is the conditional probability of treatment assignment given a set of observed covariates X, typically estimated using a model like logistic regression.

In practice, PSM employs a matching algorithm to select for each treated individual one or more "statistical twins" from the pool of control units. As detailed in evaluations such as that by Loiseau et al., a rigorous implementation involves setting a caliper to discard poor-quality matches where propensity scores are too dissimilar [10]. Unmatched subjects are excluded from subsequent efficacy analysis. This process seeks to balance the distribution of observed covariates between the treatment and control groups in the matched sample, relying critically on the assumptions of conditional ignorability (no unmeasured confounders) and positivity.

2.3. Hybrid approach (Empirical Bayes)

Empirical Bayes is a hybrid approach that combines Frequentist and Bayesian philosophies. Its core principle is to use the observed data itself to estimate the hyperparameters of the prior distribution in a Bayesian model, rather than relying on subjective specification by the analyst. Empirical Bayes provides a pragmatic and data-driven solution by "using the data to select the prior" [11].

In practice, the analyst first specifies the form of the likelihood and the prior family, but not the hyperparameters.

Data Likelihood: $L(\theta \mid x)$

Parameterized Prior: $\pi(\theta \mid \eta)$

The hyperparameters η are learned from the data by maximizing the marginal distribution.

DOI: 10.54254/2753-8818/2026.HZ27847

$$\hat{\eta}_{MLE} = arg\max_{\eta} m(\mathrm{X}|\eta) = arg\max_{\eta} \int L\left(\theta \mathrm{X}\right) \cdot \pi(\theta|\eta) d\theta \tag{7}$$

The estimated hyperparameters $\hat{\eta}_{\mathrm{MLE}}$ are then treated as known constants to compute the final posterior distribution for inference. Empirical Bayes Posterior:

$$\pi\left(\theta \mid \mathrm{X}, \hat{\eta}_{MLE}\right) = \frac{L\left(\theta|\mathrm{X}\right) \cdot \pi\left(\theta|\hat{\eta}_{MLE}\right)}{m\left(\mathrm{X}|\hat{\eta}_{MLE}\right)} \tag{8}$$

3. Methodology and results

Table 1. Comparison of Bayesian, frequentist, and empirical bayes approaches in utilizing historical data

Aspects	Bayesian	Frequentist	Empirical Bayes
Philosophical Foundation	Probability represents the degree of belief, given knowledge. It will be updated with new evidence using Bayes' Theorem.	Probability is defined as the long-term frequency of an event when it's repeated in identical trials.	Blend data-driven objectivity with a Bayesian framework.
Data Requirements	Less reliant on large sample sizes; can deal with heterogeneity.	Requires larger sample sizes; relies on assumptions of homogeneity or quantifiable heterogeneity.	Have a sample size requirement that is between Bayesian and Frequentist; can be affected when very limited data or severe incompatibility exists.
Flexibility	High: Incorporates historical knowledge or data through informative prior distributions.	Low: Treats historical data as additional fixed samples; integration relies on weighting or matching schemes.	Medium: Provides flexibility through the prior structure and reduces subjectivity by estimating parameters from the data.
Computational Demand	High: Typically requires Markov Chain Monte Carlo (MCMC) sampling for posterior inference.	Low: Mostly rely on closed-form solutions or optimization algorithms.	Medium: Requires estimation of hyperparameters, which is more demanding than standard Frequentist methods but less than full Bayesian analysis.
Application Domains	Clinical Trials, Epidemiology, Biostatistics, etc.	Social Sciences, Economics, Observational Studies, etc.	Genetics, Ecology, Certain Clinical Settings, etc.
Representative Methods	Power Prior, Commensurate Prior, Bayesian Hierarchical Models	Meta-Analysis (FEM/REM), Weighted Likelihood, Propensity Score Matching	Uses marginal likelihood maximization to estimate priors

This paper adopts a scoping review methodology, as it evaluates Bayesian and Frequentist approaches applied to different subjects and identifies research gaps in this field. The review follows the five-stage framework proposed by Arksey and O'Malley: (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) charting the data, and (5) collating, summarizing, and reporting the results [12].

The primary research question guiding this review is to explore the characteristics, effectiveness, and adaptability of Bayesian and Frequentist approaches when dealing with historical data, along with sub-questions concerning applicable subjects, philosophical differences, suitable data types, and so on. Relevant studies were identified through keywords such as "historical data", "Empirical

DOI: 10.54254/2753-8818/2026.HZ27847

Bayes", "Bayesian modeling", and "Frequentist". Except for those whose primary aim was to trace the origin of a formula or to provide a conceptual definition, papers published between 2019 and 2025 that provided either empirical validation or methodological innovation were included. Studies without data, lacking theoretical grounding, or without methodological evaluations were excluded. Then, a data extraction form was developed to capture study characteristics, incorporation details, and key outcomes. Finally, the findings are synthesized thematically and presented in tables summarizing study characteristics and key outcomes.

Table 1 shows that Bayesian methods excel in small-sample contexts due to informative priors, while Frequentist methods remain preferable when assumptions of homogeneity hold. Empirical Bayes consistently performs well in balancing bias and variance.

The Bayesian approach offers high flexibility for incorporating historical knowledge through informative priors, whereas the Frequentist approach exhibits low flexibility as it treats historical data as fixed and relies on weighting or matching. Empirical Bayes strikes a medium balance, leveraging prior structures while estimating parameters from data to reduce subjectivity.

From the aspect of computational demand, the Bayesian approach, often using MCMC sampling, is costly. The frequentist approach is generally efficient since it often uses closed-form solutions or optimization. Empirical Bayes lies in between, needing hyperparameter estimation but remaining less intensive than the full Bayesian approach.

4. Discussion

This review has synthesized and compared the Bayesian, Frequentist, and Empirical Bayes approaches for integrating historical data into research. This paper analyzes them through five key aspects—philosophical foundation, data requirements, flexibility, computational demand, and application domains—revealing that no approach is universally superior. Instead, each of them has its own characteristics, leading to distinct advantages and disadvantages.

The Bayesian approach's strength lies in its flexible mechanism for incorporating prior knowledge through prior distributions. Methods like the power prior and commensurate prior provide researchers with sophisticated tools to dynamically weight historical data based on its perceived relevance. However, this strength is also its weakness, since the subjectivity in prior selection may cause bias if not conducted with seriousness.

In contrast, the Frequentist approach offers objectivity and computational simplicity by treating historical data as additional samples. Nevertheless, this advantage results in rigidity, as it suffers from integrating uncertain prior knowledge and often relies on assumptions of homogeneity. In clinical settings, the choice between Bayesian and Frequentist methods often depends on regulatory acceptance and the availability of reliable prior information.

Empirical Bayes emerges as a pragmatic bridge between these two paradigms. By using the data to estimate prior hyperparameters, it mitigates the subjectivity of pure Bayesian analysis while retaining its adaptive borrowing capabilities. However, it introduces new layers of complexity since it relies on the quantity and quality of the current data for stable hyperparameter estimation. In genomic studies, Empirical Bayes is preferred as it automatically estimates priors from the data—overcoming the infeasibility of specifying subjective priors for each gene of Bayesian methods while avoiding the multiple testing burden of Frequentist methods.

The computational demand of each method is also an essential and practical consideration. While Frequentist methods are relatively fast, advanced Bayesian models requiring MCMC sampling can be computationally intensive. However, modern advanced technologies may lower this barrier.

DOI: 10.54254/2753-8818/2026.HZ27847

The optimal method choice depends on the actual conditions of the application. Researchers need to balance between prior knowledge and risk from subjectivity, evaluate the quantity and quality of historical data, and consider the computational resources available.

5. Conclusion

This paper provides an overview of how Bayesian, Frequentist, and Empirical Bayes approaches utilize historical data. The review begins with the philosophical understanding that separates these paradigms and then details the main methods used in each paradigm. Lastly, these three approaches were compared and evaluated from different aspects.

The central conclusion is that the effective use of historical data lies in balancing. It requires balancing the desire to increase efficiency and leverage existing knowledge with the risk of making biased conclusions. The Bayesian approach offers the most direct path to incorporation but demands careful and justified prior specification. Fields such as clinical trials and biostatistics, where prior knowledge or historical evidence is available and can be formally incorporated through informative priors to improve inference and decision-making, should pay more focus on the Bayesian approach. The Frequentist approach is more robust and simpler, but its performance in formal knowledge integration is relatively low. The Frequentist approach in the future may find broader application in social sciences and economics, where transparency and replicability are emphasized, and where prior information is either unavailable or intentionally excluded to avoid subjectivity. Empirical Bayes makes a compromise between the two paradigms, but it also has its own complexity that requires consideration. Future work should focus on hybrid frameworks that integrate Bayesian adaptivity with Frequentist robustness, offering scalable tools for fields like genetics, where borrowing strength is wanted but specifying a subjective prior is difficult.

Therefore, researchers should consider the characteristics while choosing a suitable method. In addition, they should think about whether Empirical Bayes can serve as an eclectic solution for the problem, despite its ambivalence at the philosophical level.

This review also reveals a promising direction for future study: to evaluate the suitability of different approaches for different types of historical data. Rather than seeking the best approach among the three, a more practical goal is to establish clear guidelines on which method works most effectively with which specific kinds of historical data based on their characteristics.

References

[1] Lavan, M. (2019). Epistemic Uncertainty, Subjective Probability, and Ancient History. Journal of Interdisciplinary History, 50(1), 91–111. https: //doi.org/10.1162/jinh_a_01377

[2] Flor, M., Weiß, M., Selhorst, T., Müller-Graf, C., & Greiner, M. (2020). Comparison of Bayesian and Frequentist Methods for Prevalence Estimation Under Misclassification. BMC Public Health, 20, 1135. https: //doi.org/10.1186/s12889-020-09177-4

[3] Lee, S. Y. (2024). Using Bayesian Statistics in Confirmatory Clinical Trials in the Regulatory Setting: A Tutorial Review. BMC Medical Research Methodology, 24, 110. https: //doi.org/10.1186/s12874-024-02235-0

[4] Chen, M.-H., Guan, Z., Lin, M., & Sun, M. (2025). Power Priors for Leveraging Historical Data: Looking Back and Looking Forward. Journal of Data Science, 23, 1–30. https: //doi.org/10.6339/24-JDS1161

[5] B. P. Hobbs, et al., "Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials, " *Biometrics*, vol. 67, no. 3, pp. 1047–1056, 2011, doi: 10.1111/j.1541-0420.2011.01564.x.

[6] Han, L., Deng, Q., He, Z., Fleischer, F., & Yu, F. (2024). Bayesian Hierarchical Model for Dose-Finding Trial Incorporating Historical Data. Journal of Biopharmaceutical Statistics, 34(5), 646–660. https: //doi.org/10.1080/10543406.2024.2304062

DOI: 10.54254/2753-8818/2026.HZ27847

[7] Davidov, O., & Rudas, T. (2024). On the Use of Historical Estimates. Statistical Papers, 65(1), 203–236. https: //doi.org/10.1007/s00362-022-01375-z
[8] Ivetić, J., Dedeić, J., Milićević, S., Vidojević, K., & Delić, M. (2025). Comparative Meta-Analysis of Survival, Risk, and Treatment Efficacy in Immunotherapy for Metastatic Melanoma Using Random-, Fixed-, and Mixed-Effects Models. Journal of Clinical Medicine, 14(14), 5017. https: //doi.org/10.3390/jcm14145017
[9] Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70(1), 41–55. https: //doi.org/10.1093/biomet/70.1.41
[10] Loiseau, N., Trichelair, P., He, M., Andreux, M., Zaslavskiy, M., Wainrib, G., & Blum, M. G. B. (2022). External Control Arm Analysis: An Evaluation of Propensity Score Approaches, G-Computation, and Doubly Debiased Machine Learning. BMC Medical Research Methodology, 22, 335. https: //doi.org/10.1186/s12874-022-01799-z
[11] Casella, G. (1985). An Introduction to Empirical Bayes Data Analysis. The American Statistician, 39(2), 83–87. https: //doi.org/10.2307/2682801
[12] Arksey, H., & O'Malley, L. (2005). Scoping Studies: Towards a Methodological Framework. International Journal of Social Research Methodology, 8(1), 19–32. https: //doi.org/10.1080/1364557032000119616