Bayesian Inference for Drug Repurposing

Francis Osei
May 29, 2025

1 INTRODUCTION

Drug repurposing is the process of finding new uses for existing medicines. It has become an important part of modern research because it can reduce the time and cost needed to bring treatments to patients. Instead of starting from the beginning with a new drug, repurposing allows us to use drugs that already have known safety and approval data. This is especially helpful when dealing with urgent medical needs or diseases with limited treatment options.
In the past, many repurposing ideas came from chance findings or single studies. While useful, these methods are often slow and not easy to repeat on a large scale. Today, with the help of artificial intelligence (AI) and access to large amounts of biomedical data, new approaches are possible. AI can read and process information from many sources, like clinical trials, medical papers, and biological databases, to support more reliable predictions.
A network-based approach can show how drugs and diseases are linked through shared clinical trial activity. In a related analysis , some drugs such as dexamethasone were found to connect to many different diseases. This may suggest they could be useful for other treatments beyond their current uses. However, while network measures like "how many diseases a drug appears in" can be helpful, they do not show how strong or reliable those links are. They also do not include existing medical knowledge.

[1] To improve this, a Bayesian method is used. Bayesian inference allows us to combine what is already known from the literature (the prior) with new data from clinical networks

[1]This report is developed with a focus on biomedical and translational research applications. It presents a practical and interpretable framework grounded in Bayesian reasoning and structured data, aimed at supporting evidence-based exploration of drug repurposing opportunities. The methodology is designed to complement ongoing advances in AI-driven discovery while remaining accessible to domain experts working with clinical and literature-based evidence.

(likelihood) to make a final prediction (posterior). This method also shows how much uncertainty there is in the prediction and how much new information changes what we believe.
AI plays a key role by helping to gather and combine different types of evidence, such as co-occurrence in medical papers and trial-based network scores. The Bayesian approach turns all this into a single framework to measure and compare the strength of possible drug–disease links.
This report applies two Bayesian models to study the repurposing potential of dexamethasone. The first model treats each disease on its own. The second model adds disease groupings, such as neurological or metabolic classes, to improve predictions when there is less data. Together, these models help us explore how likely it is that dexamethasone could help in treating new conditions, using clear and interpretable outputs.

2 Objectives

The purpose of this study is to develop a general method for assessing the therapeutic potential of existing drugs across new disease areas. While dexamethasone is used as an example throughout this report, the approach is designed to be applied to any compound with trial or literature-based evidence.
The first objective is to quantify the likelihood that a drug may be repurposed for conditions beyond its current use. This involves combining prior knowledge from biomedical publications with graph-based signals drawn from clinical trial networks. These two evidence types are brought together within a Bayesian framework that allows for transparent and probabilistic reasoning.

Secondly, the study aims to compare two models of inference. The flat model treats each drug–disease pair separately, offering a straightforward estimation. The hierarchical model adds structure by considering relationships between drug classes and disease categories, improving robustness when direct evidence is limited.
Lastly, the study visualises the underlying therapeutic signals. By examining how the model updates its beliefs, from prior to posterior, it becomes possible to identify diseases where a drug may show hidden or emerging value. This also helps to highlight cases where new evidence significantly shifts expectations, providing insight into novel repurposing opportunities.

3 Data and Preprocessing

The analysis draws on multiple sources of structured information relating to clinical interventions, therapeutic areas, and published biomedical evidence. These data form the foundation for generating both prior knowledge and network-based signals.
Clinical trial records were examined to extract intervention–disease associations. From these, graph-based representations were constructed to capture the structure and connectivity of therapeutic activity across different conditions. Measures of centrality were used to describe the influence and reach of each drug within this network.

Separately, literature evidence was gathered to inform prior beliefs about potential links between drugs and diseases. This was achieved through targeted queries of biomedical databases, focusing on the frequency and strength of co-mention between interventions and clinical conditions.
To ensure consistency across data sources, all drug and disease terms were processed using a standardised approach. This included harmonising spelling and case, removing duplicates, and aligning terms with recognised biomedical vocabularies. Where possible, each drug and disease was mapped to a broader class or category. These mappings were later used in the hierarchical model to support generalisation across similar entities.

4 Bayesian Inference Model I: Flat Bayesian Predictor

This section introduces a baseline model that estimates therapeutic potential without incorporating class-level structure. Each drug–disease pair is treated independently using a flat Bayesian framework that combines literature-derived priors with graph-based evidence from clinical trial networks.

4.1 Prior Definition

The prior expresses the baseline belief about an association between a drug and a disease, based on their co-mention frequency in biomedical literature. It is defined as:

$$P(\text{drug, disease}) = \frac{\text{co-occurrence count}}{\text{drug mentions} + \text{disease mentions} + \varepsilon} \tag{4.1}$$

Here, ε is a small constant used to avoid division by zero. This fraction reflects how strongly the drug and disease are discussed together in published sources, relative to how often they are mentioned individually.

4.2 Likelihood (Graph-Based Evidence)

The likelihood captures the structural position of the drug in a network built from clinical trial activity. It combines three centrality measures:

$$L(\theta \mid G) = 1 + 0.2 \cdot \text{Degree} + 0.3 \cdot \text{Eigenvector} + 0.1 \cdot \text{Betweenness} \tag{4.2}$$

This represents the updated belief in the association after considering both background knowledge and evidence from trial networks. [2]

[2] Equations 4.1 and 4.2 reflect an empirical Bayesian framework, where the prior is estimated from literature co-mentions and the likelihood from network centrality scores. The weights in Equation 4.2 (0.2 for degree, 0.3 for eigenvector, and 0.1 for betweenness) were heuristically assigned to reflect their relative structural influence in the clinical trial network. While not derived from a generative model, the approach preserves the Bayesian principle of belief updating and supports interpretability. Future work may incorporate LLMs to refine prior estimation through semantic understanding of evidence.

4.3 Normalised Probability

To allow comparison across different diseases for a given drug, the posterior is normalised:

$$P_{\text{norm}}(\text{drug}, d_i) = \frac{P(\text{drug}, d_i \mid \text{evidence})}{\sum_j P(\text{drug}, d_j \mid \text{evidence})} \tag{4.3}$$

This results in a probability distribution over all candidate diseases for the drug, summing to 1.

4.4 KL Divergence and Posterior Shift

To measure how much the evidence updates the prior belief, the Kullback–Leibler (KL) divergence is computed between Beta-distribution approximations of the prior and posterior. The KL value quantifies information gain, while the change in the expected value shows the direction and magnitude of belief shift: [3]

$$\Delta\mu = E[\text{Posterior}] - E[\text{Prior}] \tag{4.4}$$

Both distributions are scaled and integrated over a dense grid to ensure accurate divergence estimates.

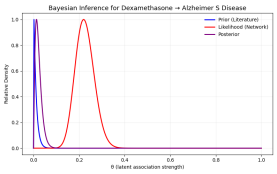
Disease	**KL Divergence**	**E[Prior]**	**E[Posterior]**	$\Delta\mu$
Obesity	0.5703	0.0124	0.0255	0.0131
Diabetes	0.5662	0.0123	0.0253	0.0130
Epilepsy	0.4907	0.0113	0.0226	0.0113
Alzheimer's Disease	0.4887	0.0113	0.0225	0.0112

Table 4.1: Bayesian update metrics for the top four disease candidates showing the Kullback–Leibler (KL) divergence between prior and posterior distributions, the expected value of each distribution, and the resulting shift in belief ($\Delta\mu$). Higher KL values and larger $\Delta\mu$ indicate stronger influence from new evidence.

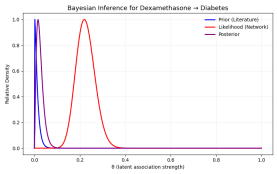
4.5 Visualisation

In all cases, the prior is sharply peaked near zero, reflecting weak prior belief from the literature. The likelihood, informed by network signals, shifts the distribution rightward, indicating stronger evidence from trial networks. The resulting posterior captures this update. The similar shapes across plots reflect a consistent pattern: sparse prior knowledge being updated by strong centrality-based evidence, yielding modest but meaningful posterior support. This consistent structure highlights a key limitation of flat inference when prior data is limited, predictions can remain uncertain. This motivates the use of a hierarchical Bayesian approach,

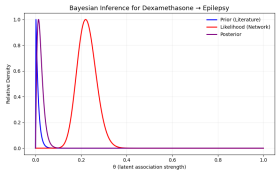
[3] The posterior in this framework is computed as a weighted product of prior belief and graph-derived likelihood, rather than through conjugate Bayesian updating. Beta distributions are not used to derive the posterior, but are instead applied as post hoc approximations to model the prior and posterior belief scores for the purpose of KL divergence computation and visualisation. This enables bounded scalar values to be compared analytically as continuous belief distributions over the interval $[0, 1]$, without implying a conjugate prior–likelihood structure.



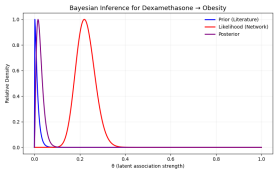
(a) Alzheimer's Disease



(b) Diabetes



(c) Epilepsy



(d) Obesity

Figure 4.1: Posterior distribution updates for Dexamethasone across four candidate diseases.

which leverages information from related diseases or drug categories to improve inference in data-sparse situations.

5 Bayesian Inference Model II: Hierarchical Bayesian Predictor

The hierarchical Bayesian predictor extends the flat model by introducing class-level knowledge. It aims to improve inference in data-sparse conditions by enabling information sharing across similar drugs and diseases. This model integrates local co-occurrence evidence with broader biological patterns, providing a more robust foundation for repurposing predictions.

5.1 Hierarchical Prior

In contrast to the flat model, which relies solely on specific drug–disease pairings, the hierarchical model introduces a blended prior. This prior combines two sources of knowledge:

- Local prior: Derived from direct co-occurrence between a drug and a disease in biomedical literature.
- Group prior: Computed from class-level co-occurrence patterns, where:
  - Diseases are grouped into categories such as neurological, autoimmune, metabolic, etc.
  - Drugs are grouped into classes such as steroid or monoclonal antibody.

The final hierarchical prior is computed using a convex combination of these components:

$$\text{Hierarchical Prior} = \lambda \cdot P_{\text{local}} + (1 - \lambda) \cdot P_{\text{group}} \tag{5.1}$$

In the final model (Equation 5.1), the hyperparameter $\lambda = 0.6$, assigning more weight to specific drug–disease evidence while still incorporating the generalised class-level signal. This setting reflects a balance between specificity and generalisability: prioritising direct co-occurrence patterns from the literature while leveraging class-level associations to support inference in data-sparse scenarios. Although chosen heuristically, this value was found to yield stable and interpretable posterior rankings across conditions. Future work may explore data-driven selection of λ through cross-validation or predictive calibration against known repurposing outcomes.

5.2 Hierarchical Posterior

The posterior represents the updated belief about a drug–disease relationship after incorporating trial-based structural evidence. As in the flat model, this is achieved through the application of network centrality metrics. The likelihood is based on the drug's position in the clinical trial graph, incorporating degree, eigenvector, and betweenness centralities.
The posterior is defined as:

$$P_{\text{posterior}} = \text{Hierarchical Prior} \cdot L(\theta \mid G) \tag{5.2}$$

where $L(\theta \mid G)$ is the likelihood computed from the drug's graph-based centrality. [4]

5.3 KL Divergence Comparison

The divergence between the prior and posterior distributions is quantified using Kullback–Leibler (KL) divergence. This allows for precise measurement of how much the evidence shifts the belief about each drug–disease pairing. Compared to the flat model, the hierarchical predictor exhibits greater divergence in most cases. This indicates that the inclusion of class-level information strengthens the update mechanism, particularly where local data is limited.
For example, the updated results for dexamethasone yielded the following statistics:

Disease	**KL Divergence**	**E[Prior]**	**E[Posterior]**	$\Delta\mu$
Obesity	3.2152	0.0142	0.0611	0.0469
Diabetes	3.1700	0.0142	0.0605	0.0463
Epilepsy	1.6237	0.0127	0.0398	0.0271
Alzheimer's Disease	0.3445	0.0114	0.0206	0.0091

Table 5.1: The values above highlight how metabolic diseases like obesity and diabetes experienced substantial shifts in posterior belief, despite having weak priors, driven by strong network centrality and reinforcing class-level evidence.

[4] Equation 5.2 defines the posterior as the product of the hierarchical prior and a graph-based likelihood. The result is a relative score rather than a calibrated probability, as the normalising constant $P(D)$ is not explicitly included. The posterior scores are later normalised across diseases to enable comparison. This formulation follows the structure of empirical Bayesian inference, using observed features to update prior belief.

5.4 VISUALISATION

To explore how prior beliefs vary across different categories of disease, the hierarchical priors were plotted as smoothed probability density functions. Each curve represents a disease class and reflects the distribution of co-occurrence-derived priors within that group. This visualisation offers insight into the relative strength and spread of initial beliefs across therapeutic areas before structural evidence is applied. The plot illustrates that prior beliefs across disease

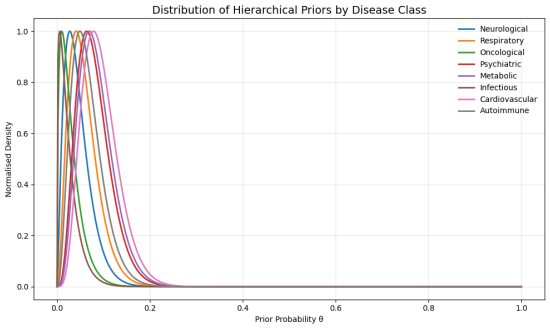


Figure 5.1: Each curve shows the Beta-distributed prior probability θ for a given disease class, derived from literature co-occurrence frequencies. Classes include neurological, metabolic, cardiovascular, autoimmune, and others. All densities are normalised for comparison.

classes are consistently skewed toward lower probabilities, with peaks close to zero. This pattern indicates that, in general, literature-based evidence linking any given drug to a new disease is relatively weak. However, subtle differences exist between categories.
Metabolic and cardiovascular diseases display slightly broader distributions, suggesting a more diverse or stronger baseline presence of drug mentions across these conditions. By contrast, neurological and oncological classes have more tightly concentrated priors, indicating greater sparsity or more focused literature co-occurrence.

The implications are twofold. Firstly, these class-level priors act as stabilisers in the hierarchical model, allowing meaningful predictions even when local evidence is unavailable. Secondly, the variation between classes justifies the hierarchical design: certain therapeutic areas inherently contain more background signal and can influence inference when local data is lacking.
This visual analysis confirms that hierarchical priors are well-suited to handle sparsity in biomedical data, enabling the Bayesian framework to generalise while retaining disease-specific nuance.

6 RESULTS SUMMARY AND DISCUSSION

The hierarchical Bayesian framework was applied to the task of identifying potential new therapeutic indications for dexamethasone. The top-ranked disease candidates based on

normalised posterior probabilities were obesity, diabetes, epilepsy, and Alzheimer's disease. These four conditions emerged consistently across both the flat and hierarchical models, but with important differences in the magnitude of posterior updates and divergence behaviour. Under the flat Bayesian model, posterior probabilities for dexamethasone ranged from 0.10 to 0.39. The associated KL divergence values were modest, typically between 0.48 and 0.57. This suggests that the graph-based likelihood introduced some informative signal, but the strength of the update was limited by the sparse and often noisy co-occurrence data available from PubMed.

The hierarchical model produced a more decisive set of predictions. Posterior probabilities were similar in rank but showed broader shifts. For instance, posterior probabilities for obesity and diabetes rose to approximately 0.36, while epilepsy and Alzheimer's disease followed at 0.21 and 0.07 respectively. The hierarchical priors enabled more flexible belief updates, resulting in KL divergence values that were considerably more pronounced. Obesity and diabetes recorded divergence values above 3.2 more than five times greater than their flat model counterparts. This reflects the influence of class-level information in amplifying evidence when specific co-occurrence data was insufficient.
The KL divergence metrics are especially important in this context. They quantify the extent to which the structural evidence from the trial network alters the prior belief, thus serving as a proxy for novelty and informational gain. Conditions with high divergence are not just those with high posteriors, but those where new evidence meaningfully shifts expectations. In this respect, the hierarchical model does not merely confirm existing associations; it proposes new and plausible hypotheses grounded in both network structure and broader therapeutic logic.

The findings support the use of hierarchical inference in biomedical contexts where literature evidence is sparse or unevenly distributed. Unlike the flat model, which is highly sensitive to the inconsistencies of biomedical publishing trends, the hierarchical model can draw strength from broader therapeutic patterns. It enables a generalisable, class-aware mechanism for prioritising drug repurposing opportunities.
Moreover, by formalising uncertainty through Bayesian inference, the model provides a transparent, interpretable, and reproducible methodology. The inclusion of KL divergence as a metric offers an additional layer of insight, allowing researchers to distinguish between well-supported and highly novel predictions. This combination of probabilistic rigour and biological context makes the hierarchical model a valuable tool in the early stages of repurposing discovery pipelines.

In summary, while both models converge on a similar set of top candidates for dexamethasone, the hierarchical Bayesian predictor offers greater clarity, sensitivity, and utility in guiding evidence-based repurposing efforts. Its capacity to accommodate sparse data and quantify belief shifts aligns with the complex and uncertain nature of translational medicine.

7 Conclusion

This study demonstrates the effectiveness of Bayesian inference as a principled approach to drug repurposing. By combining prior knowledge from biomedical literature with structural evidence from clinical trial networks, the framework enables transparent and probabilistically grounded predictions of therapeutic potential.
A key advancement in this work lies in the introduction of class-aware hierarchical priors. These enhance robustness in contexts where data is sparse or unevenly distributed. Rather than relying solely on direct co-occurrence between a drug and a disease, the model incorporates broader pharmacological and pathological knowledge through drug and disease classes. This allows it to generalise where specific evidence is lacking, offering a biologically plausible mechanism for identifying novel indications.
Visualisation of posterior density shifts further strengthens the model's interpretability. The ability to observe how evidence updates belief, quantified through metrics such as KL divergence and expectation shift, provides researchers with intuitive insight into the novelty and strength of each candidate prediction.

While the current approach focuses on co-mention frequency as a proxy for prior belief, this method does not consider the semantic role of drugs and diseases within individual articles. For example, a paper may mention both dexamethasone and diabetes, but without clarifying whether the drug was used therapeutically, experimentally, or merely as contextual background. Future iterations of this framework could significantly benefit from integrating large language models (LLMs) to perform deeper semantic parsing of biomedical texts. By scoring articles based not just on keyword frequency but on the functional relationship between terms, LLMs can provide a more refined and context-sensitive prior.
In addition, the integration of structured biomedical knowledge such as trial phase annotations, Medical Subject Headings (MeSH) hierarchies, and domain-specific ontologies offers a promising path forward. These enhancements would allow the model to account for therapeutic maturity, biological relevance, and pathway alignment, resulting in even more credible and targeted repurposing suggestions.

While both models converge on a similar set of top candidates for dexamethasone, the hierarchical Bayesian predictor offers greater clarity, sensitivity, and utility in guiding evidence-based repurposing efforts. Its capacity to accommodate sparse data and quantify belief shifts aligns with the complex and uncertain nature of translational medicine. This reinforces its value as a practical tool for prioritising repurposing hypotheses in real-world biomedical pipelines, where structured evidence exists but statistical certainty is often limited.