Unmasking and Rethinking Hierarchical Inefficiency in Healthcare Systems

Abstract

This study introduces a hierarchical Bayesian framework to disentangle inherited and self-generated inefficiency in publicly funded hospital systems. Traditional stochastic frontier models typically assume that inefficiency arises solely at the provider level. This approach challenges that orthodoxy by allowing inefficiency to propagate through interdependent administrative tiers—from provinces and regions down to individual hospitals—capturing the structural nature of performance constraints.

Using 942 hospital-level observations across Alberta, Nova Scotia, and Ontario (2015–2019), the model decomposes persistent input-oriented technical inefficiency into three latent components: provincial inherited inefficiency, regional inefficiency conditional on the province, and hospital-level self-generated inefficiency. These components are specified using log-normal distributions within a shrinkage-based hierarchical structure to reflect the nested governance of health systems.

Findings reveal that inherited inefficiency accounts for between 72.8% and 76.1% of total persistent inefficiency across provinces—72.8% in Alberta, 76.1% in Nova Scotia, and 74.7% in Ontario. These results challenge the policy orthodoxy of targeting hospitals in isolation: even hospitals with low internal inefficiency remain far from the input distance frontier due to systemic constraints imposed by higher tiers.

This framework introduces the first empirical decomposition of persistent technical inefficiency across an entire national health system. While implemented in Canada, the approach is applicable to any universal healthcare system grappling with multilevel governance. It offers policymakers a diagnostic tool to benchmark efficiency accurately and design system-level reforms—realigning funding flows, simplifying administrative structures, and reducing institutional bottlenecks—to achieve meaningful gains in technical efficiency.

Unmasking and Rethinking Hierarchical Inefficiency in Healthcare Systems

1. Introduction

Healthcare systems worldwide are under mounting pressure to enhance efficiency amid rising demand, ageing populations, and fiscal constraints. Persistent inefficiencies undermine service delivery, inflate costs, and erode health outcomes, even in high-spending countries (Organization of Economic Co-operation and Development, 2017; World Health Organization et al., 2018). International assessments suggest that a substantial portion of healthcare expenditures—estimated at up to 20–40%—could be attributed to inefficiency (Organization of Economic Co-operation and Development, 2017; World Health Organization, 2010)Yet, efforts to improve system performance often focus narrowly on the provider or hospital level, overlooking the structural origins of waste embedded across layers of governance.

Conventional inefficiency measurement typically centres on technical efficiency at the point of service delivery (Andrews & Emvalomatis, 2024; Hollingsworth & Peacock, 2008; Jacobs et al., 2006). While these approaches are valuable, they often fail to capture the broader, hierarchical nature of inefficiency that arises across multiple administrative levels (Colombi et al., 2017; Eklom & Callander, 2020; Herr, 2008; Rosko & Mutter, 2008; Worthington, 2004). From national and regional decision-making to frontline hospital operations, inefficiency is not merely a localised issue but a cascading phenomenon rooted in governance structures, resource allocation mechanisms, and bureaucratic bottlenecks(Organization of Economic Co-operation and Development, 2017; World Health Organization et al., 2018). These inefficiencies accumulate as healthcare resources pass through administrative hierarchies, creating substantial waste before funds and services ultimately reach patients. Decisions made by provincial ministries, regional health authorities, and hospital management are deeply interconnected, creating a system in which inefficiency is not only generated locally but also inherited, compounded, and propagated through governance tiers. Despite growing recognition of inefficiencies within healthcare systems, existing studies have yet to conceptualise inefficiency as a structural phenomenon cascading across administrative layers, where each tier both generates and inherits inefficiency.

This study advances the understanding of healthcare inefficiency by introducing a hierarchical model of inherited and self-generated inefficiency. The framework explicitly decomposes observed inefficiency at each level—provincial, regional, and hospital—into two components: inherited inefficiency, reflecting upstream governance and policy constraints passed down from higher levels; and self-generated inefficiency, arising from local operational failures within each administrative unit. By modelling inefficiency as a layered, interdependent process rather than a static attribute of individual hospitals, this approach reframes the locus of inefficiency within the entire health system architecture.

The cascading nature of inefficiency can be illustrated through a stylised example: if a provincial authority allocates a $100 budget but loses 4% to bureaucratic friction, only $96 reaches regional authorities. Further regional inefficiencies may reduce the effective amount to $94 before reaching hospitals. After hospital-level inefficiencies, perhaps only $90 is converted into patient care. Thus, despite no gross mismanagement at the hospital, systemic inefficiencies upstream cause cumulative resource erosion. Such effects remain invisible under traditional models focused exclusively on the frontline.

Canada's decentralised healthcare system provides an ideal setting to empirically implement and test the hierarchical inefficiency framework. The division of responsibilities among federal, provincial, regional, and hospital levels (Canadian Institute for Health Information, 2014; Marchildon et al., 2020) creates natural layers through which inefficiency can flow. Provinces control health budgets and set regulatory frameworks, while regional authorities coordinate service delivery, and hospitals operate under regional and provincial constraints (Canadian Institute for Health Information, 2014) . Despite universal coverage, the Canadian system experiences significant disparities in access, delayed funding flows, and variable efficiency across regions (Marchildon G, 2013; Sutherland & Crump, 2013) - patterns consistent with inefficiency propagation across administrative levels.

Existing literature has increasingly acknowledged the complexity of health system inefficiencies (Lipsitz, 2012; Renger et al., 2017). Cascading failures in healthcare delivery, such as administrative bottlenecks, delays in patient transitions, and preventable adverse events, often originate upstream and intensify at the point of care (Baker et al., 2004; Bruce et al., 2013). Yet few studies have formally distinguished between inefficiencies generated at a given level versus those inherited from above. This study fills that gap by estimating distinct inefficiency components at the provincial, regional, and hospital tiers within a single integrated framework.

The hierarchical model is implemented using a hierarchical Bayesian estimation strategy, allowing for flexible modelling of inefficiency transmission between tiers while accounting for uncertainty (Gelman et al., 2013, pp. 101 - 132). This approach not only quantifies inefficiency at each administrative layer but also measures the proportion of inefficiency that is locally generated versus inherited. Empirically, it leverages detailed hospital-level panel data from three Canadian provinces: Alberta, Nova Scotia, and Ontario. These provinces provide diverse administrative contexts, ranging from Alberta's centralised health authority to Ontario's recently restructured oversight system (Marchildon G, 2013; Sutherland & Crump, 2013).

By disentangling inherited and self-generated inefficiencies, the hierarchical framework provides clearer diagnostic insights for policy interventions. Strategies that focus exclusively on hospital management improvements risk addressing only a fraction of the problem. Instead, meaningful efficiency gains require tackling inefficiencies embedded in higher-level governance structures. In the Canadian context, this suggests that reforms targeting provincial funding mechanisms, administrative workflows, and regulatory coordination may yield larger efficiency dividends than hospital-level initiatives alone.

Although the empirical application focuses on Canada, the framework is broadly applicable to any multi-tiered health system, including those in the United States, Australia, Germany, and other federalised or decentralised contexts. In these systems, inefficiency cascades are likely pervasive, yet undermeasured. The hierarchical model offers a generalisable tool for identifying how and where waste is generated within complex healthcare structures.

Recognising inefficiency as a cascading, systemic phenomenon represents a significant departure from standard efficiency evaluation in health economics. It highlights the need for governance-level reforms alongside operational improvements and offers a path toward more effective, system-aligned interventions aimed at reducing waste and improving patient care. In the following sections, the hierarchical modelling approach is developed in detail, and its empirical application to Canadian hospital data is presented to illustrate how this framework captures the complex dynamics of inherited and self-generated inefficiencies.

2. Canadian Health System: A Natural Platform for Studying Inherited Inefficiency

Canada's healthcare system, characterised by universal public coverage and decentralised governance, offers an ideal context for studying how inefficiencies propagate through multiple administrative tiers. Healthcare services are primarily funded by federal, provincial, and territorial taxes, with provinces responsible for organising and delivering care. While the federal government contributes funding through mechanisms like the Canada Health Transfer, its direct role in service delivery is limited to specific populations (Health Canada, 2023a).

The foundational principles of the Canada Health Act (1984)—public administration, comprehensiveness, universality, portability, and accessibility—shape the structure of provincial insurance plans, which must ensure medically necessary services are provided free at the point of care ("Canada Health Act" 1984). Yet, significant discretion remains at the provincial and territorial levels, enabling wide variation in healthcare organisation and performance (Boychuk, 2008).

Beneath provincial governments, over 100 health regions, typically administered through Regional Health Authorities (RHAs), coordinate service delivery within defined geographic areas. While some provinces, like Alberta, have consolidated healthcare delivery under a single province-wide authority (Alberta Health Services, 2024), others, such as Nova Scotia and Ontario, maintain more fragmented systems (Nova Scotia Health, 2023; Ontario Health, 2023). This structure, though designed to encourage local responsiveness, has also introduced variability in access, performance, and governance transparency (Marchildon et al., 2020)

Despite relatively high spending levels, Canada's healthcare system underperforms on several efficiency metrics compared to other high-income countries. Systemic inefficiencies—ranging from administrative fragmentation, funding delays, and poor care coordination—contribute to an estimated 12,600 to 24,500 preventable deaths annually (Allin Sara et al., 2015; Canadian Institute for Health Information, 2014). Variations in hospital efficiency across provinces and regions are shaped not merely by local management but also by upstream decisions on funding formulas, staffing allocations, and regulatory frameworks (Allin Sara et al., 2015; Sutherland & Crump, 2013).

Particularly relevant for this study is the observation that structural inefficiencies often emerge from governance layers beyond the hospital itself. Health regions frequently encounter workforce shortages and infrastructural constraints that are rooted in provincial funding policies (Canadian Institute for Health Information, 2025). Likewise, federal and provincial-level administrative bottlenecks delay service innovation and impair hospital resource flow (Marchildon et al., 2020) .

Canada's decentralised but layered healthcare system creates natural tiers—federal, provincial, regional, and hospital—through which inefficiencies can be transmitted, inherited, and compounded. These features make it an exemplary empirical setting for developing and testing a hierarchical model of inherited and self-generated inefficiency. The next section introduces the model specification and empirical strategy used to disentangle inefficiency propagation across governance levels.

3. Model Specification and Empirical Methodology

Estimating inefficiency using a stochastic frontier model requires specifying an appropriate production relationship. This study employs an input distance function (IDF) to analyse how inefficiently hospitals utilise inputs relative to the most efficient use, given a fixed level of outputs. Moreover, the IDF does not require explicit knowledge of input or output prices and is particularly well-suited for contexts involving multiple inputs and outputs.

To estimate this relationship, a translog functional form is utilised for its flexibility and ability to approximate complex production functions. Originally introduced by Christensen, Jorgenson, and Lau (1973), the translog input distance function provides a second-order approximation of an unknown production function, allowing it to capture diverse input substitution possibilities without imposing restrictive assumptions. This flexibility is especially advantageous in the hospital sector, where inputs such as labour, materials, and capital interact in complex ways to deliver healthcare services.

Furthermore, the translog input distance function offers several practical benefits. First, it accommodates the simultaneous use of multiple inputs and outputs, particularly suited to healthcare systems where hospitals deploy a wide range of resources to deliver diverse services. Second, it enables the estimation of input elasticities and returns to scale, providing critical insights into how hospitals adjust their input utilisation in response to variations in output levels or technological progress. These features make it a powerful tool for analysing efficiency in complex production environments like healthcare (Christensen et al., 1973).

For a longitudinal dataset with P inputs and Q outputs, the translog input distance function for hospital j at time t is given by:

$$\ln D_{jt}^{I} = \alpha + \alpha_j + \alpha_k + \alpha_l \sum_{q=1}^{Q} \beta_q \ln y_{qjt} + \frac{1}{2} \sum_{q=1}^{Q} \sum_{r=1}^{Q} \beta_{qr} \ln y_{qjt} \ln y_{rjt} + \sum_{p=1}^{P} \delta_p \ln x_{pjt}$$

$$+ \frac{1}{2} \sum_{p=1}^{P} \sum_{s=1}^{P} \delta_{ps} \ln x_{pjt} \ln x_{sjt} + \sum_{p=1}^{P} \sum_{q=1}^{Q} \gamma_{pq} \ln x_{pjt} \ln y_{qjt} + v_{jt} \quad (1)$$

In the above equation, $\ln D_{jt}^{I}$ represents the natural logarithm of the input distance function for the hospital j at time t. The term α denotes the global intercept, capturing baseline inefficiency across all levels. The term v_{jt} is the random noise component, assumed to be normally distributed with a mean of zero, capturing random shocks to the production system. Additionally, α_j, α_k and α_l represent hospital-specific, regional, and provincial intercepts, respectively.

Each intercept represents a distinct layer of variation within the healthcare system. The hospital-specific intercept, α_j, reflects factors such as patient demographics, hospital size, and available services, which vary significantly across hospitals. In contrast, the regional intercept, α_k, captures differences arising from region-specific factors, such as the allocation of resources, coordination of care, and regional policy implementations. These differences are particularly relevant in Canadian healthcare, where regional healthcare authorities manage key aspects of service provision, as highlighted by the Canadian Institute for Health Information (2014). The provincial intercept, α_l, on the other hand, accounts for province-level factors, reflecting variations in health policies, funding models, and access to specialised services (Canadian Institute for Health Information, 2014; Marchildon G, 2013).

On the input side, x_{pjt} denotes the $p-$th input for hospital j at time t with P representing the total number of inputs. These inputs include resources such as labour, medical equipment, and supplies. The $\sum_{p=1}^{P} \delta_p \ln x_{pjt}$ represents the elasticities of the inputs. Interaction terms $\frac{1}{2}\sum_{p=1}^{P}\sum_{s=1}^{P} \delta_{ps} \ln x_{pjt} \ln x_{sjt}$ capture the relationships between different inputs, where s represents another input. The coefficient δ_{ps} shows whether two inputs are complementary or substitutes, for example, whether increasing staffing levels reduces the need for additional medical equipment. Additionally, cross-product terms $\sum_{p=1}^{P}\sum_{q=1}^{Q} \gamma_{pq} \ln x_{pjt} \ln y_{qjt}$ show how specific inputs influence the production of specific outputs, with γ_{pq} capturing the joint contribution of inputs and outputs.

The estimable form of the translog input distance function (IDF) for hospital j at time t can be rewritten and extended to account for persistent inefficiency at various hierarchical levels—namely, the provincial, regional, and hospital levels. This approach captures heterogeneity at each level while distinguishing persistent inefficiency, or inefficiency that does not change over time, at each hierarchical tier. This yields the following form:

$$-\ln x_{Pjt} = \alpha + \alpha_j + \alpha_k + \alpha_l - u_j - u_k - u_l + \sum_{q=1}^{Q} \beta_q \ln(y_{qjt}) + \frac{1}{2}\sum_{q=1}^{Q}\sum_{r=1}^{Q} \beta_{qr} \ln(y_{qjt}) \ln(y_{rjt})$$

$$+ \sum_{p=1}^{P-1} \delta_p \ln(x^*_{pjt}) + \frac{1}{2}\sum_{p=1}^{P-1}\sum_{s=1}^{P-1} \delta_{ps} \ln(x^*_{pjt}) \ln(x^*_{sjt}) + \sum_{p=1}^{P-1}\sum_{q=1}^{Q} \gamma_{pq} \ln(x^*_{pjt}) \ln(y_{qjt})$$

$$+ v_{jt} \quad (2)$$

where $x^*_{pjt} = \frac{x_{pjt}}{x_{Pjt}}$, with x_{Pjt} is the normalising input.

In addition to heterogeneity, the Eq. (2) model incorporates persistent inefficiency terms at each hierarchical level (u_j for hospitals, u_k for regions, and u_l for provinces) to capture inefficiency that remains fixed over time. Persistent inefficiency at the hospital level (u_j) might reflect structural inefficiencies rooted in hospital practices, infrastructure limitations, or entrenched operational processes that do not change significantly year over year (Colombi et al., 2017; Mutter et al., 2008). At the regional level, persistent inefficiency (u_k) could represent inefficiencies inherent to regional healthcare management or consistent issues in resource distribution that are difficult to alter over time (Allin Sara et al., 2016; Jacobs et al., 2006, p. 33). In contrast, the persistent inefficiency at the provincial level (u_l) captures inefficiency arising from provincial healthcare policies or systemic issues in funding and regulation that impact the entire healthcare system within the province (Canadian Institute for Health Information, 2014; Marchildon G, 2013). These time-invariant inefficiencies reflect structural and policy-related factors that persist at each level.

The estimated persistent inefficiency terms ($u_{(.)}$) at various levels in Eq. (2) represents input-oriented inefficiency, which assesses the extent to which healthcare providers use more inputs than necessary to deliver a given level of services. This measure reflects the proportional excess in input utilisation relative to the efficient input frontier, indicating how much input usage could be reduced while maintaining the same output level.

To incorporate the hierarchical effects, the model in Eq. (2) can compactly be written as:

$$z_{jt} = \boldsymbol{p}'_{jt}\beta + v_{jt} + \psi_j + \psi_k + \psi_l \quad (3)$$

Here, $z_{jt} = -\ln x_{P_{jt}}$ represents the dependent variable for the hospital j at time t, and $\boldsymbol{p}'_{jt}$ is a row vector of independent variables, including the global intercept. The parameter vector β contains the coefficients to be estimated. The term v_{jt} represents random noise, assumed to follow a normal distribution with mean zero and standard deviation of σ_v.

The hierarchical effects in Eq. (3) are represented by convolution terms that capture the interplay between time-invariant heterogeneity and persistent inefficiency at different levels. At the hospital level, the Hospital-Level Convolution Effect ($\psi_j = \alpha_j - u_j$) combines the hospital-specific intercept (α_j), which reflects unobserved, time-invariant heterogeneity such as structural or socio-economic factors beyond hospital management's control, with persistent inefficiency (u_j). This inefficiency arises from entrenched operational shortcomings, infrastructure limitations, or systemic challenges specific to the hospital. Similarly, the Regional-Level Convolution Effect ($\psi_k = \alpha_k - u_k$) captures the regional intercept (α_k), representing stable regional characteristics such as population density or resource allocation policies, along with persistent inefficiency (u_k), which reflects inefficiencies in regional healthcare management or systemic resource distribution issues. At the provincial level, the Provincial-Level Convolution Effect ($\psi_l = \alpha_l - u_l$) incorporates the province-specific intercept (α_l), reflecting broad, stable factors unique to each province, alongside persistent inefficiency (u_l), which arises from inefficiencies linked to healthcare policies, regulatory frameworks, or systemic funding mechanisms that affect the healthcare system at the provincial scale. To ensure strictly positive values, all inefficiencies (u_j, u_k, and u_l) are modelled hierarchically on the logarithmic scale, following log-normal distributions derived from normal distributions.

In Eq. (3), the hierarchical effects—hospital, regional, and provincial—are modelled as independent, implying no interaction or dependency between the levels. However, both theoretical and practical considerations suggest that this assumption is unlikely to hold, as inefficiencies at one level often influence and are influenced by factors at other levels. For instance, provincial healthcare policies can shape regional resource allocation, affecting hospital-level operations. This interdependence is a fundamental characteristic of healthcare systems, where decisions and inefficiencies cascade across the hierarchy. Consequently, the model in Eq. (3), referred to as the Naïve Hierarchical Effects Model (Naïve), can be described as naïve because it does not account for these dependencies and interactions across levels.

To address this limitation and incorporate hierarchical dependency, the model in Eq. (3) can be reformulated compactly as:

$$z_{jt} = \boldsymbol{p}'_{jt}\beta + v_{jt} + \psi_j \quad (4)$$

In Eq. (4), the regional (ψ_k) and provincial (ψ_l) effects are not directly included but instead propagate through the hierarchical structure to influence hospital-level effects, encapsulated in ψ_j. By embedding higher-level influences within ψ_j, the model aligns with the organisational structure of healthcare systems, where provincial and regional factors converge at the hospital level. This reformulation captures the cascading nature of hierarchical dependencies, ensuring that inefficiencies at broader levels of the system are reflected in hospital-level effects.

The hierarchical framework models unobserved heterogeneity across levels, beginning with the provincial level. At this level, the provincial intercept α_l is assumed to follow a normal distribution with a mean of zero and a standard deviation of σ_{α_l}, representing province-specific variability. At the regional level, the regional intercept α_k is modelled as dependent on the provincial intercept α_l, with additional variability captured by σ_{α_k}. Similarly, at the hospital level, the hospital-specific intercept α_j depends on the regional intercept α_k, with variability scaled by σ_{α_j}. These relationships are expressed mathematically as:

$$\alpha_l \sim Normal(0, \sigma_{\alpha_l}) \quad (5)$$

$$\alpha_k \sim Normal(\alpha_l, \sigma_{\alpha_k}) \quad (6)$$

$$\alpha_j \sim Normal\left(\alpha_k, \sigma_{\alpha_j}\right) \quad (7)$$

In this hierarchical framework, inefficiency is modelled on the logarithmic scale, where the mean inefficiency at the preceding level directly influences each level's inefficiency. This approach, termed the Hierarchical Inefficiency Model (HIM), captures the nested dependency structure across provincial, regional, and hospital levels, aligning with the natural organisation of healthcare systems.

At the provincial level, inefficiency (u_l) is modelled as a log-normal distribution derived from a normal distribution with mean μ_{u_l} and standard deviation σ_{u_l}, representing province-level inefficiency. At the regional level, inefficiency (u_k) is also modelled as log-normal, where the mean is determined directly by the logarithm of provincial inefficiency ($\log(u_l)$), reflecting the hierarchical relationship between these levels. Similarly, at the hospital level, inefficiency (u_j) follows a log-normal distribution, with its mean determined by the logarithm of regional inefficiency ($\log(u_k)$).

To properly reflect the hierarchical structure in the dispersion of inefficiency, we allow regional-level variability $\sigma_{u_{k[l]}}$ to depend on provinces and hospital-level variability $\sigma_{u_{j[k]}}$ to depend on regions. In healthcare, regional inefficiency is shaped by province-wide policies, resource allocation, and infrastructure, leading to province-specific differences in regional inefficiency variation. Similarly, hospital inefficiency varies within regions due to differences in funding, management, and patient demographics. By allowing regional dispersion to vary by province and hospital dispersion to vary by region, the model aligns inefficiency variability with the nested structure of healthcare systems, ensuring a more realistic representation of inefficiency propagation.

The hierarchical relationships are defined as:

$$u_l \sim \exp\left[Normal\left(\log(\mu_{u_l}), \sigma_{u_l}\right)\right] \quad (7)$$

$$u_k \sim \exp\left[Normal\left(\log(u_l), \sigma_{u_{k[l]}}\right)\right] \quad (8)$$

$$u_j \sim \exp\left[Normal\left(\log(u_k), \sigma_{u_{j[k]}}\right)\right] \quad (9)$$

Building on the HIM framework, the model is extended to incorporate Influence Factors that explicitly model the proportional transmission of inefficiency between hierarchical levels. This extended framework termed the Hierarchical Inefficiency Model with Influence Factors (HIM-IF), allows for the partial inheritance of inefficiency while accounting for localised adjustments at each level. By introducing these influence factors, the model better captures the interplay between provincial, regional, and hospital inefficiencies in healthcare systems.

In the HIM-IF, inefficiency is still modelled hierarchically on the logarithmic scale to ensure strictly positive values. At the provincial level, inefficiency (u_l) is modelled as a log-normal distribution, derived from a normal distribution with mean μ_{u_l} and standard deviation σ_{u_l}, representing province-level heterogeneity in inefficiency. At the regional level, inefficiency (u_k) is also modelled as log-normal, but its mean is now determined by the logarithm of provincial inefficiency ($\log(u_l)$) scaled by a Provincial Influence Factor (PIF), denoted as ϕ_l. This factor quantifies the proportional impact of provincial inefficiency on regional inefficiency while $\sigma_{u_{k[l]}}$ captures the variability around this scaled mean, representing regional-level inefficiency heterogeneity. Similarly, at the hospital level, inefficiency (u_j) follows a log-normal distribution, with its mean depending on the logarithm of regional inefficiency ($\log(u_k)$) scaled by a Regional Influence Factor (RIF), denoted as ϕ_k. Hospital-specific variability around this scaled mean is captured by $\sigma_{u_{j[k]}}$, reflecting inefficiency heterogeneity among hospitals. The hierarchical relationships incorporating influence factors are expressed as:

$$u_l \sim \exp\left[Normal\left(\log(\mu_{u_l}), \sigma_{u_l}\right)\right] \qquad (10)$$

$$u_k \sim \exp\left[Normal\left(\phi_l . \log(u_l), \sigma_{u_{k[l]}}\right)\right] \qquad (11)$$

$$u_j \sim \exp\left[Normal\left(\phi_k . \log(u_k), \sigma_{u_{j[k]}}\right)\right] \qquad (12)$$

The parameters ϕ_l and ϕ_k, known as the Provincial Influence Factor (PIF) and Regional Influence Factor (RIF), respectively, measure the proportional transfer of inefficiency between hierarchical levels. These factors are constrained between 0 and 1, enabling partial inheritance of inefficiency while allowing for localised adjustments. A higher value of ϕ_l implies that regional inefficiency closely reflects the logarithmic magnitude of provincial inefficiency, with minimal attenuation. Conversely, lower values suggest greater regional autonomy or mitigating factors that reduce the influence of provincial inefficiency. Similarly, ϕ_k, quantifies the degree to which hospital inefficiency aligns with the logarithmic magnitude of regional inefficiency, with higher values reflecting a stronger influence and lower values allowing for greater hospital-specific variability.

It is important to note that a high value of ϕ_l does not imply that most of the inefficiency observed at the regional level is directly transferred from the provincial level. Instead, the PIF acts as a scaling mechanism, quantifying the proportion of provincial inefficiency that influences regional inefficiency. The observed inefficiency at each level combines scaled contributions from higher levels and level-specific variability, reflecting both systemic influences and localised adjustments.

This interpretation of ϕ_l and ϕ_k underscores their roles as proportional factors rather than sole determinants of total inefficiency. By capturing both inherited inefficiencies and localised heterogeneity, these parameters highlight the dynamic relationship between systemic influences and site-specific adjustments, providing a robust framework for understanding inefficiency propagation within hierarchical systems.

In passing, it is important to emphasise the distinction in how total log inefficiency is conceptualised between the HIM and HIM-IF models. For the HIM model, the total log inefficiency at the hospital level is represented by u_j , which inherently captures inefficiencies at all levels (province, region, and hospital). This is because, in HIM, inefficiency flows directly through the logarithmic means from one level to the next, with each successive level absorbing and reflecting inefficiencies from higher levels. Thus, the inefficiency observed at the hospital level (u_j) is an aggregate representation of inefficiencies from all preceding levels.

Conversely, in the HIM-IF model, the total log inefficiency is computed as the sum of inefficiencies across all levels, i.e., $u_l + u_k + u_j$. This distinction arises because, in HIM-IF, inefficiencies at each level (u_l , u_k and u_j) are modelled separately with influence factors (ϕ_l and ϕ_k) to capture the proportional transfer of inefficiency between hierarchical levels. While these influence factors allow inefficiencies to flow to lower levels, they do not fully aggregate into a single value at the hospital level, as in HIM. Instead, each level retains its distinct contribution to the total inefficiency. The summing approach in HIM-IF ensures that the total inefficiency reflects all sources of inefficiency—province, region, and hospital—that collectively impact hospitals.

Bayesian estimation of HIM-IF model requires the specification of complete data likelihood that incorporates observed and latent variables. Let $\boldsymbol{Z} = \{z_{jt}\}$ denote the observed dependent variables for all hospitals and time periods. The complete data likelihood incorporates the observed data Z, the latent variables for unobserved heterogeneity $(\alpha_j, \alpha_k, \alpha_l)$ and inefficiencies (u_j, u_k, u_l) at each level. The parameter vector θ includes the following unknown parameters: β (coefficients of the independent variables p'_{jt}) ϕ_k (Regional Influence Factor, RIF), ϕ_l (Provincial Influence Factor, PIF), σ_v (standard deviation of the random noise in z_{jt}), σ_{α_j}, σ_{α_k}, σ_{α_l}

(standard deviations of unobserved heterogeneity at the hospital, regional, and provincial levels), $\sigma_{u_{j[k]}}, \sigma_{u_{k[l]}}, \sigma_{u_l}$ (standard deviations of log inefficiency at the hospital, regional, and provincial levels), and μ_{u_l} (mean of the natural logarithm of inefficiency at the provincial level).

The complete data likelihood can thus be expressed as follows:

$$P(\boldsymbol{Z}, u_j, u_k, u_l, \alpha_j, \alpha_k, \alpha_l \middle| \boldsymbol{\theta})$$

$$= P(Z \middle| \alpha_j, u_j, \beta, \sigma_v, p'_{jt}) \times P\left(\alpha_j \middle| \alpha_k, \sigma_{\alpha_j}\right) \times P(\alpha_k \middle| \alpha_l, \sigma_{\alpha_k}) \times P(\alpha_l \middle| \sigma_{\alpha_l})$$

$$\times P\left(u_j \middle| u_k, \phi_k, \sigma_{u_{j[k]}}\right) \times P\left(u_k \middle| u_l, \phi_l, \sigma_{u_{k[l]}}\right) \times P(u_l \middle| \mu_{u_l}, \sigma_{u_l}) \quad (13)$$

Expanding the likelihood in its parametric form, it becomes:

$$P(\boldsymbol{Z}, \alpha_j, \alpha_k, \alpha_l, u_j, u_k, u_l, \middle| \boldsymbol{\theta})$$

$$= \left[\prod_{j=1}^{J}\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma_v^2}}\, exp\left(-\frac{\left(z_{jt} - \boldsymbol{p}'_{jt}\beta - \alpha_j + u_j\right)^2}{2\sigma_v^2}\right)\right] \times \left[\prod_{j=1}^{J}\frac{1}{\sqrt{2\pi\sigma_{\alpha_j}^2}}\exp\left(-\frac{\left(\alpha_j - \alpha_k\right)^2}{2\sigma_{\alpha_j}^2}\right)\right]$$

$$\times \left[\prod_{k=1}^{K}\frac{1}{\sqrt{2\pi\sigma_{\alpha_k}^2}}\exp\left(-\frac{(\alpha_k - \alpha_l)^2}{2\sigma_{\alpha_k}^2}\right)\right] \times \left[\prod_{l=1}^{L}\frac{1}{\sqrt{2\pi\sigma_{\alpha_l}^2}}\exp\left(-\frac{\alpha_l^2}{2\sigma_{\alpha_l}^2}\right)\right]$$

$$\times \left[\prod_{j=1}^{J}\frac{1}{u_j\sigma_{u_j}\sqrt{2\pi}}\exp\left(-\frac{\left(\log(u_j) - \log(\phi_k \times \log(u_k))\right)^2}{2\sigma_{u_{j[k]}}^2}\right)\right]$$

$$\times \left[\prod_{k=1}^{K}\frac{1}{u_k\sigma_{u_k}\sqrt{2\pi}}\exp\left(-\frac{(\log(u_k) - \log(\phi_l \times \log(u_l)))^2}{2\sigma_{u_{k[l]}}^2}\right)\right]$$

$$\times \left[\prod_{l=1}^{L}\frac{1}{u_l\sigma_{u_l}\sqrt{2\pi}}\exp\left(-\frac{\left(\log(u_l) - \mu_{u_l}\right)^2}{2\sigma_{u_l}^2}\right)\right] \quad (14)$$

Along with the data likelihood, prior distributions are specified for each parameter in the model. All scale parameters, including those for residual noise and hierarchical heterogeneity, are modelled using half-normal distributions. Half-normal priors are particularly suitable for scale parameters because they enforce positivity while avoiding boundary issues, as Gelman (2006) recommended. These priors allow for concentration near plausible values without being overly restrictive, balancing informativeness and flexibility.

The prior for the residual noise standard deviation (σ_v) is specified as a half-normal distribution to enforce positivity while constraining the values to plausible magnitudes. The given prior $\sigma_v \sim HalfNormal(0.01, 0.30)$, ensures concentration on small deviations, consistent with the log-transformed nature of the dependent variable, where large residual deviations are unlikely. Further, the standard deviation of the residual noise is restricted to a range between zero and the standard deviation of the observed data. This restriction helps maintain a balance between realism by grounding the parameter within plausible values derived from the data and flexibility by allowing the model to capture variability within this range.

The standard deviation of persistent inefficiency parameters is modelled on the logarithmic scale, reflecting relative variability in inefficiency at different hierarchical levels of the healthcare system. These standard deviations increase progressively from provinces to regions to hospitals, aligning with the hierarchical structure and the expectation of greater variability at more granular levels. At the provincial level, where only three units are in the data, inefficiency variability is limited due to centralised policies and standardised practices. This is reflected in the prior $\sigma_{u_l} \sim HalfNormal(0, 0.40)$, which imposes a smaller expected standard deviation. At the regional level, the number of regions slightly exceeds the number of provinces, and local management introduces additional variability. This is captured by the prior $\sigma_{u_{k[l]}} \sim HalfNormal(0, 0.55)$, allowing for moderate variability increases. Hospitals represent the largest group with the most significant variability due to infrastructure operational processes and resource allocation differences This is reflected in the prior $\sigma_{u_{j[k]}} \sim HalfNormal(0, 0.70)$, which accommodates the highest level of inefficiency variability in the hierarchical system.

The priors for the scale of the intercept terms at the provincial, regional, and hospital levels are modelled as half-normal distributions: $\sigma_{\alpha_l} \sim HalfNormal(0, 0.10)$, $\sigma_{\alpha_k} \sim HalfNormal(0, 0.15)$ and $\sigma_{\alpha_j} \sim HalfNormal(0, 0.20)$. These priors capture the hierarchical structure of the healthcare system, where heterogeneity is smallest at the provincial level due to centralised policies and fewer units, increases at the regional level with localised factors and slightly more units, and is largest at the hospital level due to the high number of units and diverse operational and structural differences. The half-normal priors are weakly informative, enforcing positivity while allowing flexibility for the data to influence the posterior. If the actual heterogeneity at any level exceeds the prior expectation, the half-normal prior's broad support enables the model to adjust accordingly, accurately reflecting the variability observed in the data.

Using the findings from Colombi et al. (2017) as a guide, where persistent inefficiency at the hospital level was found to have a mode of approximately 0.16, the prior for the province-level inefficiency log mean ($log\mu_{u_l}$) is specified as $Normal(-1.30, 0.50)$. This choice reflects the hierarchical nature of the healthcare system, where inefficiency at the provincial level encompasses broader systemic influences compared to hospital-level inefficiency. The hierarchical framework of this study estimates inefficiency across multiple layers—provincial, regional, and hospital—making it likely that inefficiency values are higher at broader levels due to aggregated systemic effects. Under this prior, the mean inefficiency at the provincial level is modelled flexibly to account for potential variability. Specifically: The sampled mean is approximately $e^{-1.30-0.5\times0.5^2} \approx 0.27$ and a mode of $e^{-1.30} \approx 0.27$. This specification allows the model to balance prior knowledge with flexibility, ensuring that provincial inefficiency is realistically captured while accommodating the complexity of inefficiency propagation across hierarchical levels. The wider standard deviation (0.50) enables the prior to reflect the variability associated with provincial-level inefficiency while remaining informed by empirical observations. This approach provides a robust basis for estimating inefficiency at the provincial level, which is likely to exhibit greater values than hospital-level inefficiency due to its systemic nature.

While Colombi et al. (2017) measured inefficiency at the hospital level, capturing both systemic and localised factors, the inefficiency modelled at the provincial level reflects broader systemic inefficiencies tied to centralised policies and resource allocation. Provincial inefficiency is expected to be lower than hospital-level inefficiency, as it forms the genesis of inefficiency that propagates through the hierarchy—regions and hospitals—each adding level-specific inefficiencies. This hierarchical propagation aligns with the model's structure. It justifies the prior assumption, ensuring it remains realistic, interpretable, and flexible to adapt to the observed data, even without direct empirical evidence.

Uniform priors are used for both ϕ_k and ϕ_l, assuming that all values within the range [0, 1] are equally likely. This choice is consistent with a non-informative prior belief, providing no preference for any specific level of influence and allowing the data to inform the posterior distribution.

The No-U-Turn Sampler (NUTS[1]), an efficient extension of the Hamiltonian Monte Carlo (HMC) algorithm, is used to estimate the posterior distribution. Sampling is conducted in CmdStanR[2.] with five independent Markov chains, each generating 4000 samples, of which the first 2000 are used for warmup. The full model code is included in the supplementary materials. However, due to data-sharing restrictions imposed by the Canadian Institute for Health Information (CIHI), the underlying dataset cannot be made publicly available.

4. Data & Summary statistics

The dataset used in this study was obtained from the Canadian Institute for Health Information (CIHI) through a formal data request and spans the fiscal years 2011/2012 to 2017/2018. It includes hospital-level data from Alberta, Ontario, and Nova Scotia, forming an unbalanced panel consisting of 83 hospitals in Alberta, 13 in Nova Scotia, and 123 in Ontario. While some hospitals are observed throughout the entire study period, others are only observed for part of this timeframe, reflecting the typical structure of administrative health data.

The dataset captures essential inputs and outputs relevant to hospital input-oriented inefficiency. Labour inputs are classified into three categories: Medical Personnel (MED), Management and Operational Support (MOS), and Unit Producing Personnel (UPP) compensation. These categories align with best practices outlined by Allin Sara et al. (2016) , emphasising the importance of including distinct labour input categories for understanding efficiency variations. MED compensation covers gross salaries, professional fees, and benefit contributions for medical personnel. MOS compensation reflects payments to administrative and operational staff managing hospital functions, while UPP compensation includes wages paid to direct service delivery staff such as nurses and technicians.

Following McConnell et al. (2018, pp. 526 - 528), UPP compensation was adjusted relative to MOS compensation

1 For detailed explanation of NUTS algorithm refer to Hoffman and Gelman (2014).

2 CmdStanR (Command Stan R) is a lightweight interface to Stan for R users. For more information, refer Gabry and Cesnovar (2023). *Getting started with CmdStanR* [Online]. Available: https://mc-stan.org/cmdstanr/articles/cmdstanr.html [Accessed].

based on hourly wage differentials, producing a weighted aggregated measure (W_MOSUPP_EXP). This method reflects the relative productivity contributions of each labour input, ensuring alignment with the marginal productivity theory of resource allocation.

Operational expenditures are also critical for assessing hospital inefficiency. Following the framework proposed by Cantor Victor John M and Poh Kim Leng (2018), the dataset includes for supplies, drugs, sundry expenses, equipment, and contracted-out services. They are expressed on a per-bed basis to standardise these costs and make them comparable across hospitals of different sizes. The price per bed for each category, such as supplies (SUP), drugs (DRG), contracted services (CON), and sundry expenses (SUN), is calculated by dividing the real expenditure by the number of Beds Staffed and In Operation (BEDS).

A weighted average price per bed was calculated for supplies and drugs, as well as for contracted services and sundry expenses. The weighted aggregated values for supplies and drugs were derived using the weighted average price per bed multiplied by the total expenditure for each category, resulting in the weighted aggregated supplies and drugs expenditure (W_SUPDRG_EXP). Similarly, the weighted aggregated contracted services and sundry expenditures (W_CONSUN_EXP) were derived for contracted-out services and sundry expenses. These weighted variables will be used in the analysis to capture the relative importance of each category while accounting for hospital size. All expenditure figures are deflated using province-specific deflators to adjust for inflation. This adjustment ensures that comparisons over time reflect real changes in hospital expenditures rather than nominal increases due to inflation.

Capital in healthcare typically refers to the physical assets such as buildings, equipment, and grounds. In production economics, capital is ideally measured as a flow, reflecting the ongoing use of these assets over time. However, due to the nature of healthcare services, it is not easy to measure capital flow directly. In this study, Following Colombi et al. (2017) and Jacobs et al. (2006, p. 31).Beds Staffed and In Operation (BEDS) are used as a proxy for capital, representing the number of beds available and staffed at the beginning of each fiscal year. This measure includes bassinets set up outside the nursery for infants other than newborns. Although it provides a convenient proxy for hospital capital, it is acknowledged as a simplified measure, representing capacity rather than the dynamic use of assets (Jacobs et al., 2006, p. 31).

The outputs in this study are measured using Resource Intensity Weight values, which capture the resources used for inpatient and outpatient care. The Total Acute Resource Use Intensity (ARU) is calculated by summing the Resource Intensity Weight values for all valid inpatient cases, providing a comprehensive measure of resource demand for acute care services. Similarly, the Total Outpatient Resource Use Intensity (ORU) is determined by summing the Resource Intensity Weight values for all valid outpatient cases, reflecting the resource utilisation for outpatient services. These Resource Intensity Weight values provide a standardised approach to assessing care intensity across cases, taking into account each patient's complexity and resource needs. This approach accounts for patient heterogeneity and resource demands, as emphasized by Tsionas (2006) and Jacobs et al. (2006, pp. 21-27) ensuring that output measures reflect both the volume and complexity of care. Descriptive summary statistics for all key input and output variables, pooled across provinces and years, are presented in Table 1.

Table 1

Variable	Observation	Min	1st Qu	Median	Mean	Std.Dev	3rd Qu	Max
MED	942	0.06	0.93	3.42	16.43	31.95	19.22	249.30
UPP	942	2.74	9.62	21.86	92.82	146.75	111.06	762.49
MOS	942	0.48	1.64	4.27	18.16	30.71	19.75	197.72
SUP	942	0.31	1.55	4.57	21.19	37.24	22.96	217.39
DRG	942	0.01	0.22	0.91	8.67	22.82	5.63	259.04
CON	942	0.01	0.24	1.09	5.10	11.12	4.60	83.96
SUN	942	0.13	0.62	1.66	12.89	53.80	7.22	704.36
BEDS	942	15.00	38.00	84.50	225.12	306.66	281.00	1510.00
ARU	942	97.58	816.73	2244.98	11620.76	19007.29	13696.43	90815.70
ORU	942	58.47	637.88	1403.72	4304.84	6099.92	6172.36	43620.94

All expenditure variables in Table 1 are expressed in millions of Canadian dollars. Capital is measured as the number of staffed hospital beds, while output variables (ARU and ORU) are reported in standardised Resource Intensity Weight (RIW) scores, which adjust for both patient complexity and service volume. The substantial variation across hospitals and provinces in both inputs and outputs underscores the suitability of this dataset for hierarchical Bayesian modelling. The unbalanced panel structure, with repeated observations nested within hospitals and provinces (and regions where applicable), enables the identification of multi-level sources of inefficiency. This structure supports the decomposition into inherited and self-generated inefficiency and allows a robust investigation of performance variation across Canada's federated health system.

5. Empirical Results

The selection of the most appropriate model was guided by a comprehensive validation strategy incorporating two distinct evaluation approaches. The first approach involved Leave-One-Entity-Out Cross-Validation with a 10-fold structure, which assessed how well models generalised to previously unseen hospitals (Aritz et al., 2024). The second approach utilised full-sample model selection criteria, including the Watanabe-Akaike Information Criterion (WAIC), the Bayesian Information Criterion (BIC), and Leave-One-Out Cross-Validation (LOO-CV)(Gelman et al., 2014; Kohavi, 1995; Stone, 1974; Vehtari et al., 2017). These full-sample methods provided a deeper understanding of model complexity, predictive accuracy, and generalisation across all observations. While the 10-fold cross-validation strategy was designed to test the predictive accuracy of each model on holdout hospitals, WAIC, BIC, and LOO-CV assessed how well models fit the full dataset while penalising excessive complexity to ensure out-of-sample reliability.

The 10-fold Leave-One-Entity-Out Cross-Validation framework systematically excluded approximately 17 hospitals per iteration, ensuring that each model was evaluated on entirely unseen hospitals. The choice of 10 folds balances computational efficiency with reliability, maintaining sufficiently large samples for training and testing while keeping processing times manageable (Liu & Rue, 2022). Across all three models, predictive accuracy remained stable, with minimal variation in Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE), suggesting that each model captured overall inefficiency patterns with similar precision. Specifically, the Naïve model recorded an RMSE of 3.3683, an MSE of 11.4471, and an MAE of 2.6695, while the Hierarchical Inefficiency Model (HIM) produced an RMSE of 3.3682, an MSE of 11.4470, and an MAE of 2.6695. Similarly, the Hierarchical Inefficiency Model with Influence Factors (HIM-IF) yielded an RMSE of 3.3683, an MSE of 11.4471, and an MAE of 2.6695.

Although these values are extremely close, they confirm that differences in inefficiency modelling do not negatively impact overall predictive accuracy. However, this raises an important question: if predictive accuracy remains largely unchanged, what is the benefit of structuring inefficiency hierarchically? The key advantage of models like HIM and HIM-IF lies in their ability to explicitly capture inefficiency propagation across hierarchical levels—a feature that is not reflected in traditional predictive statistics but is crucial for understanding how inefficiencies at provincial, regional, and hospital levels interact. While 10-fold cross-validation evaluated predictive accuracy across different hospital subsets, it did not incorporate model complexity penalties or hierarchical inefficiency transmission. This made the WAIC, BIC, and LOO-CV criteria essential for the final model selection, as they provide a more refined assessment of model fit and complexity beyond simple predictive performance.

However, despite the similarity in predictive accuracy, log-likelihood values revealed a key distinction in model fit. The Naïve model recorded the highest log-likelihood (17,510.62), outperforming both HIM (16,983.0) and HIM-IF (17,000.89). This suggests that the Naïve model better fits the overall dependent variable in an unpenalized sense. However, this higher likelihood is largely attributable to the Naïve model's simpler structure, which does not impose hierarchical dependency constraints between provinces, regions, and hospitals. The HIM and HIM-IF models integrate hierarchical inefficiency propagation, redistributing variance across multiple levels, which inherently reduces log-likelihood due to additional model constraints. This reduction in log-likelihood is expected as more structured models introduce constraints that prevent overfitting to idiosyncratic patterns in the data.

While log-likelihood provides an unpenalized measure of model fit, WAIC, BIC, and LOO-CV offer more refined evaluations by incorporating model complexity adjustments to avoid overfitting. Unlike the 10-fold cross-validation approach, which validates models on held-out hospital data, WAIC, BIC, and LOO-CV were calculated using the full dataset, providing a global assessment of model performance across all observations. WAIC balances predictive accuracy with model complexity by penalising models that fit the data too closely, preventing overfitting. Under WAIC, the HIM-IF model exhibited the lowest penalised deviance, outperforming both the Naïve and HIM models, suggesting it was the best-performing model in terms of balancing accuracy and complexity. However, in calculating WAIC, the estimation of the effective number of parameters revealed that some were relatively large, indicating the presence of highly influential observations. In such cases, WAIC may not be fully reliable, and Leave-One-Out Cross-Validation (LOO-CV), a more robust alternative, was also considered.

BIC, which applies a stronger penalty for model complexity than WAIC, favoured the HIM-IF model, which recorded the lowest BIC value (-48,829.43), outperforming the Naïve model (-19,999.60) and the HIM model (-24,283.68). The superior performance of HIM-IF under BIC suggests that it strikes an optimal balance between flexibility and explanatory power, whereas the HIM model incurs a slightly higher penalty due to its additional complexity. Since BIC applies a more substantial complexity penalty than WAIC, this result indicates that HIM-IF achieves the most efficient trade-off between model parsimony and fit.

Leave-One-Out Cross-Validation (LOO-CV) estimates how well a model is expected to predict unseen data by systematically omitting one observation at a time, refitting the model to the remaining data, and then predicting the left-out observation. This process provides a robust measure of generalizability while preventing overfitting. The LOO-CV results are summarised using the expected log pointwise predictive density (ELPD), which measures how well a model predicts new data points while penalising overly flexible models. Among the three models, the HIM had the highest ELPD (set as the reference with an elpd difference of 0), indicating the best overall generalisation to new data. The HIM-IF followed closely (elpd difference = -27.3), while the Naïve model performed the worst (elpd difference = -23.5). A more negative elpd difference indicates slightly weaker predictive performance. The close performance between HIM and HIM-IF suggests that both models generalise well, with HIM demonstrating a slight advantage in stability. However, high Pareto k diagnostic values revealed that some observations had a high influence on the results, meaning that a small subset of extreme data points contributed disproportionately to the model's performance estimates. This indicates that while LOO-CV favours HIM slightly over HIM-IF, these results should be interpreted cautiously, particularly in cases where high-leverage observations impact model rankings.

While the HIM model performed best under LOO-CV, the HIM-IF model remains the preferred choice due to its superior performance under WAIC and BIC, both of which account for model complexity and parsimony. Unlike the Naïve model, which treats inefficiency as a hospital-level residual, and the HIM, which assumes full inefficiency inheritance across levels, the HIM-IF model provides a more balanced approach, capturing both inherited inefficiency and localised variation.

Even though HIM optimises predictive performance, HIM-IF provides a more interpretable and theoretically robust framework for modelling inefficiency propagation. By aligning inefficiency transmission with real-world healthcare structures, HIM-IF is the most policy-relevant model, offering deeper insights into inefficiency sources at hospital, regional, and provincial levels. Thus, the HIM-IF model is selected as the final model for inefficiency estimation, as it provides the best balance between predictive performance and model complexity. While HIM achieves slightly better predictive generalisation, HIM-IF offers a more interpretable structure by explicitly capturing inefficiency propagation across levels, making it the most theoretically sound and policy-relevant model. However, for transparency, results from both HIM and HIM-IF models are presented in Table 2, allowing for a comparative evaluation of their estimates. This ensures that policymakers and researchers can assess the trade-offs between model complexity and interpretability when analysing hospital inefficiency.

In this study, input variables were normalised by their geometric mean and expressed in logarithmic form. This normalisation ensures that the first-order coefficients of the translog input distance function can be directly interpreted as elasticities at the mean of the data. Such an approach is particularly valuable in the translog framework, which accounts for nonlinear relationships and interactions through second-degree and cross-product terms. By adopting this normalisation strategy, the model provides a more flexible representation of hospital production technology while maintaining interpretability. A key feature of this study is the choice of numeraire input, where the weighted aggregated contracted services and sundry expenditure (W_CONSUN_EXP) is used to impose the homogeneity restriction. This ensures that all other input variables are expressed relative to W_CONSUN_EXP, allowing for a meaningful comparison of elasticities while maintaining the theoretical properties of the input distance function.

Table 2. Posterior means, standard deviations and 95 credible intervals of the translog HIM-IF & HIM Model

Variables	Posterior Mean (SD) [95 credible interval]	
	HIM_IF	HIM
α (global intercept)	0.300 (0.113) [0.120, 0.478]	0.342 (0.115) [0.157, 0.520]
$\beta_{\log(MED\_EXP)}$	0.023 (0.013) [0.002, 0.045]	0.023 (0.013) [0.001, 0.045]
$\beta_{log(W\_MOSUPP\_EXP)}$	0.360 (0.031) [0.310, 0.411]	0.360 (0.031) [0.309, 0.412]
$\beta_{log(W\_SUPDRG\_EXP)}$	0.075 (0.015) [0.050, 0.099]	0.075 (0.015) [0.050, 0.099]
$\beta_{log(BEDS)}$	0.529 (0.023) [0.4941, 0567]	0.529 (0.024) [0.491, 0.567]
$\beta_{log(ORU)}$	-0.253 (0.024) [-0.291, -0.214]	-0.253 (0.024) [-0.292, -0.214]
$\beta_{log(ARU)}$	-0.452 (0.026) [-0.492, -0.408]	-0.452 (0.025) [-0.491, -0.410]
β_{trend}	0.001 (0.001) [-0.001, 0.003]	0.001 (0.001) [-0.001, 0.003]
$\beta_{log(MED\_EXP)^2}$	0.007 (0.004) [0.001, 0.014]	0.007 (0.004) [0.001, 0.013]
$\beta_{log(W\_MOSUPP\_EXP)^2}$	0.060 (0.044) [-0.011, 0.135]	0.060 (0.045) [-0.013, 0.134]
$\beta_{log(W\_SUPDRG\_EXP)^2}$	-0.033 (0.011) [-0.051, -0.015]	-0.034 (0.011) [-0.052, -0.015]
$\beta_{\log(BEDS)^2}$	-0.079, (0.025) [-0.121, -0.037]	-0.079 (0.026) [-0.122, -0.037]
$\beta_{\log(ORU)^2}$	-0.080 (0.014) [-0.104, -0.057]	-0.080 (0.014) [-0.104, -0.057]
$\beta_{\log(ARU)^2}$	-0.156 (0.014) [-0.178, -0.133]	-0.155 (0.014) [-0.178, -0.133]
$\beta_{(trend)^2}$	-0.001 (0.001) [-0.003, 0.000]	-0.001 (0.001) [-0.003, 0.000]
$\beta_{\log(MED\_EXP \times W\_MOSUPP\_EXP)}$	-0.068 (0.026) [-0.110, -0.025]	-0.068 (0.025) [-0.109, -0.026]
$\beta_{\log(MED\_EXP \times W\_SUPDRG\_EXP)}$	0.040 (0.011) [0.022, 0.058]	0.041 (0.011) [0.023, 0.058]
$\beta_{\log(MED\_EXP \times BEDS)}$	0.016 (0.017) [-0.012, 0.043]	0.016 (0.017) [-0.012, 0.044]

$\boldsymbol{\beta}_{\log(\boldsymbol{MED\_EXP}\times\boldsymbol{ORU})}$	-0.004 (0.013) [-0.026, 0.017]	-0.004 (0.013) [-0.026, 0.017]
$\boldsymbol{\beta}_{\log(\boldsymbol{MED\_EXP}\times\boldsymbol{ARU})}$	-0.028 (0.011) [-0.046, -0.010]	-0.028 (0.011) [0.023, 0.058]
$\boldsymbol{\beta}_{\log(\boldsymbol{MED\_EXP}\times\boldsymbol{trend})}$	0.005 (0.002) [0.002, 0.008]	0.005 (0.002) [0.002, 0.008]
$\boldsymbol{\beta}_{\log(\boldsymbol{W\_MOSUPP\_EXP}\times\boldsymbol{W\_SUPDRG\_EXP})}$	-0.081 (0.035) [-0.139, -0.023]	-0.080 (0.036) [-0.139, -0.021]
$\boldsymbol{\beta}_{\log(\boldsymbol{W\_MOSUPP\_EXP}\times\boldsymbol{BEDS})}$	0.031 (0.061) [-0.069, 0.131]	0.032 (0.062) [-0.070, 0.135]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{W\_MOSUPP\_EXP}\times\boldsymbol{ORU})}$	0.111 (0.039) [0.047, 0.176]	0.111 (0.040) [0.046, 0.177]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{W\_MOSUPP\_EXP}\times\boldsymbol{ARU})}$	-0.039 (0.035) [-0.097, 0.019]	-0.039 (0.037) [-0.099, 0.021]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{W\_MOSUPP\_EXP}\times\boldsymbol{trend})}$	-0.006 (0.004) [-0.013, 0.001]	-0.006 (0.005) [-0.014, 0.001]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{W\_SUPDRG\_EXP}\times\boldsymbol{BEDS})}$	0.105 (0.022) [0.068, 0.142]	0.105 (0.022) [0.068, 0.141]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{W\_SUPDRG\_EXP}\times\boldsymbol{ORU})}$	0.018 (0.020) [-0.015, 0.050]	0.018 (0.020) [-0.015, 0.050]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{W\_SUPDRG\_EXP}\times\boldsymbol{ARU})}$	0.017 (0.018) [-0.013, 0.047]	0.017 (0.018) [-0.013, 0.047]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{W\_SUPDRG\_EXP}\times\boldsymbol{trend})}$	0.002 (0.003) [-0.002, 0.007]	0.002 (0.003) [-0.002, 0.007]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{BEDS}\times\boldsymbol{ORU})}$	-0.121 (0.030) [-0.172, -0.072]	-0.121 (0.031) [-0.172, -0.071]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{BEDS}\times\boldsymbol{ARU})}$	0.051 (0.029) [0.004, 0.099]	0.051 (0.030) [0.003, 0.100]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{BEDS}\times\boldsymbol{trend})}$	-0.004 (0.003) [-0.009, 0.001]	-0.004 (0.003) [-0.009, 0.002]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{ORU}\times\boldsymbol{ARU})}$	0.162 (0.025) [0.121, 0.202]	0.162 (0.024) [0.122, 0.202]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{ORU}\times\boldsymbol{trend})}$	-0.005 (0.003) [-0.010, -0.001]	-0.005 (0.003) [-0.010, -0.001]
$\boldsymbol{\beta}_{\boldsymbol{log}(\boldsymbol{ARU}\times\boldsymbol{trend})}$	0.004 (0.002) [0.000, 0.008]	0.004 (0.002) [0.000, 0.008]
$\boldsymbol{\mu}_{\boldsymbol{u}_{\boldsymbol{l}}}$	-2.931 (0.483) [-3.726, -2.135]	-3.064 (0.469) [-3.837, -2.284]
$\boldsymbol{\sigma}_{\boldsymbol{v}}$	0.045 (0.002) [0.043, 0.048]	0.045 (0.002) [0.043, 0.048]
$\mathbb{E}\left[\boldsymbol{\sigma}_{\boldsymbol{u}_{\boldsymbol{j}[\boldsymbol{k}]}}\right]$	0.575 (0.420) [0.060, 1.382]	0.568 (0.414) [0.060, 1.363]
$\mathbb{E}\left[\boldsymbol{\sigma}_{\boldsymbol{u}_{\boldsymbol{k}[\boldsymbol{l}]}}\right]$	0.432 (0.324) [0.036, 1.050]	0.419 (0.317) [0.033, 1.033]
$\boldsymbol{\sigma}_{\boldsymbol{u}_{\boldsymbol{l}}}$	0.315 (0.238) [0.025, 0.779]	0.304 (0.236) [0.022, 0.758]
$\boldsymbol{\sigma}_{\boldsymbol{\alpha}_{\boldsymbol{l}}}$	0.080 (0.060) [0.007, 0.196]	0.080 (0.060) [0.006, 0.194]
$\boldsymbol{\sigma}_{\boldsymbol{\alpha}_{\boldsymbol{k}}}$	0.060 (0.046) [0.006, 0.150]	0.059 (0.047) [0.005, 0.150]
$\boldsymbol{\sigma}_{\boldsymbol{\alpha}_{\boldsymbol{j}}}$	0.335 (0.032) [0.291, 0.394]	0.329 (0.031) [0.286, 0.383]
Province Influence Factors		
$\boldsymbol{\phi}_{\boldsymbol{AB}}$	0.608 (0.250) [0.167, 0.966]	-
$\boldsymbol{\phi}_{\boldsymbol{NS}}$	0.498 (0.292) [0.045, 0.952]	-

A remarkable observation from the results is the similarity between the HIM-IF and HIM models in terms of posterior means and standard deviations across all estimated parameters. The global intercepts for HIM-IF (α= 0.300) and HIM (α= 0.342) are nearly identical within their respective credible intervals. Likewise, the first-order input and output elasticities exhibit no substantial differences, with their 95% credible intervals overlapping entirely between the two models.

This strong consistency suggests that the two specifications capture hospital production technology in a nearly identical manner, indicating that the estimated relationships hold consistently across specifications. The output elasticities $\beta_{\log(ORU)}$and $\beta_{\log(ARU)}$ are negative due to the imposed normalisation of the input distance function, which ensures homogeneity. This transformation aligns with the theoretical expectation that the original input distance function is decreasing in outputs.

Similarly, the positive input elasticities indicate that the input distance function is increasing in inputs, meaning that as hospitals allocate more resources—such as labour, supplies, and capital—their input requirements expand. This is a fundamental property of a well-specified input distance function, ensuring that an increase in any input results in a greater measured distance from the efficient frontier. The fact that input elasticities remain positive at the geometric mean of the data further confirms that the monotonicity condition is satisfied, reinforcing the validity of the model's specification.

Keeping up the focus of this study, we interpret the estimated parameters from the HIM-IF model, ensuring that the input distance function framework is correctly accounted for in the interpretation. Since the model follows an input-oriented approach, positive coefficients for inputs indicate that increasing these resources moves hospitals further from the efficient frontier, meaning that they contribute to inefficiency. Conversely, negative coefficients for outputs suggest that increasing output levels allows hospitals to become more efficient by reducing the required level of inputs to provide the same level of care.

The negative coefficient for outpatient resource use ($\beta_{\log(ORU)}$=-0.253) implies that as hospitals increase outpatient activity, the input distance function decreases—meaning hospitals can proportionally reduce their inputs *less* while maintaining the same output level. This indicates that higher outpatient volumes intensify input use, pushing hospitals closer to the efficiency frontier (i.e., less "slack" in inputs exists as outputs grow).

Similarly, the larger negative coefficient for acute resource use ($\beta_{\log(ORU)}$=−0.452) suggests that inpatient care imposes even stronger input constraints: As hospitals expand inpatient services, the potential to proportionally reduce inputs diminishes further. This aligns with the notion that inpatient care is more resource-rigid than outpatient care, requiring hospitals to operate with proportionally fewer input savings at higher volumes.

Since RTS measures how outputs respond to proportional changes in inputs and the function is structured as an inverse relationship, that is $RTS = \frac{-1}{-0.253-0.452} \approx 1.42$. This implies that a 1% increase in all inputs leads to an approximately 1.42% increase in total inpatient and outpatient resource use. The reciprocal nature of RTS in the input distance function confirms that hospitals operate under increasing returns to scale, meaning that as inputs expand, output grows more than proportionally, reflecting greater input productivity at larger operational scales.

All the first-order input coefficients with respect to the input distance function at the geometric mean are positive, which satisfies the monotonicity condition of the input distance function. This ensures that the function is increasing in inputs, meaning that as hospitals expand their resource use—whether through medical personnel, administrative costs, supplies, or bed capacity—their total input requirements rise, pushing hospitals further away in their input use from the efficient frontier.

The coefficient for medical personnel expenditures ($\beta_{\log(MED\_EXP)}$= 0.023) indicates that a 1% increase in the use of medical staff results in a 0.023% increase in total input requirements, pushing hospitals further from the efficient input frontier. However, the small magnitude suggests that increases in medical staffing contribute only marginally to overall hospital input expansion. This relatively low marginal effect of medical personnel suggests that hiring additional clinicians contributes less to inefficiency compared to expansions in administrative or physical infrastructure.

Administrative expenditures ($\beta_{\log(W\_MOSUPP\_EXP)}$ = 0.360) show a stronger effect, meaning that a 1% increase in administrative and operational staff use leads to a 0.36% rise in total input requirements. This highlights that greater reliance on administrative resources significantly contributes to hospital input expansion, further distancing hospitals from the efficient frontier. Similarly, supplies and drug expenditures ($\beta_{\log(W\_SUPDRG\_EXP)}$= 0.075) also lead to higher input use, though their impact is smaller compared to administrative resource use. This highlights the need for cost-effective procurement and supply management, as increased use of medical supplies and pharmaceuticals adds to total input requirements but at a lower rate than labour and operational expenditures.

The coefficient for hospital beds ($\beta_{\log(BEDS)}$= 0.529) is the largest among inputs, indicating that expanding bed capacity substantially raises input use. This underscores that hospitals must align bed expansion with patient demand to prevent unnecessary resource growth, reinforcing the importance of efficient capacity planning. The strong positive coefficient further implies that bed capacity is a major driver of input inefficiency, highlighting the potential value of policies that encourage bed-sharing strategies or outpatient care substitution to improve resource use.

Turning to second-order effects, the squared term for medical personnel expenditures ($\beta_{\log(MED\_EXP)^2}$ =0.007) is positive, indicating that as medical personnel use increases, its impact on total input requirements grows at an increasing rate. This means that beyond a certain point, additional medical staff contribute more to total input expansion than initially expected, potentially due to diminishing marginal productivity. Thus, hospitals must be cautious about over-reliance on increased medical staffing, as it may not yield proportional service improvements.

In contrast, the squared term for weighted supplies and drugs expenditures ($\beta_{\log(W\_SUPDRG\_EXP)^2} = -0.033$) is negative, implying that while greater use of medical supplies and drugs initially raises total input requirements, the rate of increase slows down as expenditure levels grow. This points towards the presence of economies of scale in procurement and usage, where bulk purchasing and optimised supply chains reduce marginal input expansion over time.

Similarly, the negative squared term for hospital beds ($\beta_{\log(BEDS)^2}$ -0.079) indicates that the effect of additional beds on total input use diminishes as more beds are added. It follows that hospitals may approach an optimal capacity where expanding bed availability does not proportionally increase input requirements, reinforcing the need for careful bed utilisation strategies to prevent excessive resource allocation.

Several interaction effects provide valuable insights into how different hospital resources complement or substitute each other in determining total input use. These interactions highlight the interdependencies between labour, infrastructure, and service provision, shaping how hospitals allocate resources to maintain operational efficiency. The negative interaction between medical and administrative expenditures $\beta_{\log(MED\_EXP \times W\_MOSUPP\_EXP)} = -0.068$) suggests that higher staffing levels combined with increased administrative support improve resource utilisation. This implies that coordinating clinical and administrative teams more effectively reduces unnecessary input expansion, potentially streamlining hospital operations and making better use of existing resources.

Conversely, the positive interaction between medical personnel and supplies/drug expenditures ($\beta_{\log(MED\_EXP \times W\_SUPDRG\_EXP)}$=0.040) indicates that higher medical staffing levels are associated with greater use of medical supplies and pharmaceuticals. That is, hospitals with larger medical personnel teams also consume more treatment-related resources, potentially due to increased patient care intensity or inefficiencies in resource consumption. This interaction highlights the need for cost-conscious management of medical supply chains to ensure that growing staff levels do not lead to disproportionate input demands.

Another key interaction emerges between medical personnel and acute care resource use ($\beta_{\log(MED\_EXP \times ARU)}$=−0.028), which is negative, indicating that hospitals with higher inpatient activity utilise medical personnel more efficiently. As inpatient volumes rise, the additional demand for medical personnel increases at a diminishing rate, suggesting better workforce allocation in hospitals handling larger inpatient caseloads. This reinforces the idea that high inpatient volume hospitals benefit from economies of scale in labour utilisation.

The positive interaction between administrative expenditures and outpatient care ($\beta_{\log(W\_MOSUPP\_EXP \times ORU)}$ =0.111) suggests that higher administrative support leads to increased input use in outpatient services. This could stem from greater bureaucratic overhead in managing outpatient care, implying that hospitals must carefully assess the efficiency of administrative processes to avoid unnecessary input growth. In contrast, the negative interaction between hospital beds and outpatient services ($\beta_{\log(BEDS \times ORU)} = -0.121$) suggests that hospitals with greater bed capacity and higher outpatient volumes allocate inputs more effectively. This indicates potential operational synergies, where hospitals with more beds can manage outpatient flow more efficiently, reducing resource duplication and optimising patient transitions between inpatient and outpatient care.

The positive interaction between hospital beds and acute care ($\beta_{\log(BEDS \times ARU)}$ =0.051) suggests that when both inpatient activity and bed capacity increase, total input use rises. This may reflect capacity constraints or resource-intensive inpatient treatments, which require higher staffing levels and medical supply consumption as hospitals accommodate more complex inpatient cases. Also, the positive interaction between outpatient and acute care services ($\beta_{\log(ORU \times ARU)}$ =0.162) suggests that hospitals experiencing simultaneous increases in outpatient and inpatient services require greater total inputs. This may be due to higher patient turnover, increased administrative burden, or shared resource constraints between the two service areas. The finding underscores the need for carefully balancing inpatient and outpatient care to optimise resource allocation and prevent excessive input expansion.

The posterior distribution for the linear time trend coefficient (β_{trend}= 0.001) spans zero, indicating no statistically credible evidence of system-wide technological progress or regress in hospital production during the study period. While the positive mean estimate might superficially suggest mild technological progress—where fewer inputs are required over time to produce the same output—the wide credible interval [−0.001, 0.003] renders this conclusion unreliable. Similarly, the squared trend term ($\beta_{(\text{trend})^2}$ = −0.001) offers no meaningful evidence of accelerating or decelerating technological change, as its 95% credible interval [−0.003, 0.000] includes zero. This overall pattern of temporal stability implies that, at an aggregate level, hospitals neither consistently improved nor declined in their input efficiency over time.

Beneath this aggregate neutrality lie important input- and output-specific technological dynamics. Among inputs, the credibly positive interaction between medical personnel expenditures and time ($\beta_{\log(MED\_EXP \times trend)}$) = 0.005, CI [0.002, 0.008]) indicates that the input requirements for clinical staff increased systematically over time—suggesting a form of technological regress in labor productivity. This may reflect rising administrative burdens, increasing case complexity, or diminishing marginal returns to workforce expansion, whereby each additional unit of medical labor contributes progressively less to output. In contrast, administrative expenditures ($\beta_{\log(W\_MOSUPP\_EXP \times trend)}$ = −0.006), pharmaceutical and supply costs ($\beta_{\log(W\_SUPDRG\_EXP \times trend)}$= 0.002), and bed capacity ($\beta_{\log(BEDS \times trend)}$ = −0.004) all exhibited trend interactions with 95% credible intervals that include zero, suggesting no meaningful change in the technological contribution of these inputs over the study period.

Turning to outputs, the significantly negative interaction between outpatient activity and time ($\beta_{\log(ORU \times trend)}$) = −0.005, CI [−0.010, −0.001]) provides credible evidence of technological progress in outpatient care. Over time, hospitals required fewer inputs per unit of outpatient service delivered, likely reflecting innovations such as telehealth, improved care coordination, and specialization in high-volume ambulatory procedures. By contrast, the marginally positive interaction between inpatient activity and time ($\beta_{\log(ORU \times trend)}$) = 0.004, CI [0.000, 0.008]) suggests a potential decline in inpatient efficiency, though the evidence is not conclusive. This ambiguity may reflect variation across hospitals—some achieving efficiency gains through lean practices, while others experienced rising input demands due to increasing patient acuity or complexity.

The study now turns to the core empirical contribution of this study: measuring how inefficiency flows across Canada's multi-tiered healthcare system. Tables 3 to 5 present a breakdown of inefficiency at three hierarchical levels—province, region, and hospital—for Alberta, Nova Scotia, and Ontario. These findings illustrate both inherited inefficiency, which flows down from higher administrative levels, and self-generated inefficiency, which arises within each layer. Together, they provide a comprehensive picture of how inefficiency accumulates and varies across the country's healthcare governance architecture.

Table 3: Hierarchical Decomposition of Inefficiency in Alberta's Health System

Province	Base Provincial Inefficiency	Province → To Region	Region	Regional Inefficiency (Base + Inherited)	Region → To Hospital	Hospital Inefficiency† (Base + Inherited)	Total Inefficiency† (Province + Region + Hospital)
Alberta (83 hospitals)	7.25 %	4.17 %	1	4.51 %	2.07 %	2.44 %	14.83 %
			2	4.18 %	1.57 %	1.80 %	13.75 %
			3	5.03 %	2.84 %	7.48 %	21.08 %
			4	5.13 %	2.89 %	3.80 %	17.04 %
			5	4.80 %	2.45 %	2.78 %	15.53 %
Grand average				4.73 %	2.36 %	3.66 %	16.45 %
Grand Average Efficiency Score							83.55 %

† Averaged at Regional Level

As presented in Table 3, Alberta emerges as the most inefficient of the three provinces, with a total system inefficiency of 16.45%, corresponding to an efficiency score of 83.55%. The province starts with a relatively high base provincial inefficiency of 7.25%, which reflects structural inefficiencies at the top of the system—such as bureaucratic rigidity, funding delays, or poorly coordinated provincial policies. However, this inefficiency is not confined to the provincial level. It is passed down through the administrative hierarchy as inherited inefficiency. The inefficiency transmitted from the province to its regions—measured here at 4.17%—represents the inefficiency that regional health authorities must contend with simply because they happen to operate within Alberta's provincial governance system.

It is important to emphasise that this "passing down" of inefficiency does not reduce the inefficiency at the higher level. The province retains its full base inefficiency of 7.25%; the 4.17% inherited by the region is an additional burden that the regional level must absorb. In other words, the regional level begins its operations, which are already laden with structural inefficiencies over which it has little or no control. On top of this inherited inefficiency, the region generates its own inefficiency—from internal management limitations, fragmented coordination, or delayed policy execution—bringing the total regional inefficiency to an average of 4.73%. The difference between the inherited portion (4.17%) and the total regional inefficiency (4.73%) represents the region's self-generated inefficiency.

A similar pattern unfolds as inefficiency cascades from the regional to the hospital level. Alberta's average inherited inefficiency from region to hospital is 2.36%, meaning that before hospitals even begin functioning, they are already burdened with inefficiency from the above layers. Hospitals then add their own inefficiencies, such as inefficient discharge processes, staff shortages, or excessive diagnostics, resulting in an average hospital-level inefficiency of 3.66%. Together, the cumulative inherited inefficiency from province and region totals 11.98% (7.25% + 4.73%), which represents 72.8% of Alberta's total observed inefficiency of 16.45%. This figure underscores the central argument of this study: the majority of observed hospital inefficiency stems from systemic governance failures at higher administrative levels.

The total cumulative inefficiency reaches a striking 21.08% in Region 3, where hospital-level inefficiency peaks at 7.48%, clearly indicating a compounding process. These results show that inherited inefficiency can snowball when not absorbed or managed at intermediate levels, creating severe downstream inefficiencies at the point of care.

Geographical and demographic factors further complicate Alberta's capacity to absorb inefficiency. With a land area of over 661,000 km² and a sparse population density of just 5.7 people per km², the province faces formidable logistical challenges in delivering healthcare across vast and remote areas (Statistics Canada, 2021a). Specialised infrastructure, such as air ambulances (e.g., STARS), are often essential for reaching northern communities, but they also drive-up operational complexity and cost(Alberta Health, 2021). These structural realities magnify the effects of inherited inefficiency, especially when regional governance lacks the agility or resources to absorb the upstream burdens effectively.

Table 4: Hierarchical Decomposition of Inefficiency in Nova Scotia's Health System

Province	Base Provincial Inefficiency	Province → To Region	Region	Regional Inefficiency (Base + Inherited)	Region → To Hospital	Hospital Inefficiency† (Base + Inherited)	Total Inefficiency† (Province + Region + Hospital)
Nova Scotia (13 hospitals)	7.02 %	3.41 %	1	4.17 %	2.06 %	2.97 %	14.79 %
			2	4.10 %	2.00 %	2.76 %	14.48 %
			3	4.07 %	2.02 %	2.75 %	14.43 %
			4	4.03 %	2.01 %	3.15 %	14.84 %
			5	4.03 %	2.00 %	2.92 %	14.58 %
Grand average				4.08 %	2.02 %	2.91 %	14.63 %
Grand Average Efficiency Score							85.37 %

† Averaged at Regional Level

Table 4 presents the hierarchical decomposition of inefficiency in Nova Scotia's health system, which contrasts Alberta on how inefficiency flows and is managed across administrative layers. Although the province begins with a similar level of base provincial inefficiency—7.02% compared to Alberta's 7.25%—the way inefficiency propagates through the rest of the system diverges meaningfully. With a total system inefficiency of 14.63%, Nova Scotia's efficiency score of 85.37% reflects a more controlled and less compounding accumulation of inefficiency across tiers.

Where the difference becomes more subtle—but still meaningful—is at the regional level. Regions in Nova Scotia inherit an average of 3.41% inefficiency from the province, compared to 4.17% in Alberta. While both provinces transmit structural inefficiency to their middle tier, Nova Scotia's regional burden begins slightly lighter. However, once self-generated inefficiency is considered, the distinction narrows. Nova Scotia's total regional inefficiency averages 4.08%, while Alberta's stands at 4.73%. That places the regional self-generated inefficiency in Nova Scotia at about 0.67%, whereas Alberta's is closer to 0.56% when comparing aggregate averages. So, rather than Nova Scotia's regions clearly outperforming Alberta's, we observe a more evenly matched middle tier, with both provinces exhibiting modest levels of regional inefficiency generation. The earlier impression of Nova Scotia's superior regional governance doesn't hold up fully when viewed through this lens.

At the hospital level, Nova Scotia continues to demonstrate a more stabilised pattern of inefficiency. Hospitals inherit an average of 2.02% inefficiency from their respective regions—comparable to Alberta—but generate only 2.91% in additional inefficiency on their own. By contrast, Alberta's hospitals add 3.66% on average, and in some cases, far more. Alberta's Region 3, for instance, records a staggering 7.48% hospital-level inefficiency, underscoring how quickly inefficiency can escalate when upstream burdens meet internal mismanagement or structural stress.

In Nova Scotia, no such extremes are observed. Hospital inefficiency remains contained across all five regions, fluctuating only slightly between 2.75% and 3.15%. This tight clustering suggests that while hospitals are not immune to inefficiency, they are generally operating within more consistent and manageable bounds. Unlike Alberta, where inefficiency at the hospital level can spike dramatically in response to inherited strain, Nova Scotia's hospitals appear better equipped—or better structured—to prevent downstream inefficiency from compounding.

The scale of the system may contribute to this. Nova Scotia has only 13 hospitals compared to Alberta's 83, and such a smaller network inherently presents fewer coordination points, less institutional variability, and arguably clearer lines of accountability. But this is only part of the explanation. Nova Scotia's population density—17.4 people per km², Statistics Canada (2021b) is more than triple that of Alberta, which stands at just 5.7 people per km². These underlying geographic and demographic differences fundamentally shape the logistical context in which healthcare operates. In Alberta, the vast physical distances and dispersed population can amplify the difficulty of resource distribution, staff deployment, and infrastructure coordination. In contrast, Nova Scotia's compact geography allows for tighter integration of services, more direct oversight, and reduced travel burdens—all of which likely contribute to smoother hospital operations and steadier performance metrics.

Importantly, the absence of outliers in Nova Scotia's hospital-level inefficiency cannot be dismissed as a byproduct of system size alone. Smaller systems can still suffer from mismanagement and disorganisation (Giancotti et al., 2017). What distinguishes Nova Scotia is not merely its scale but how that scale interacts with geography and administration. The relatively high population density, combined with a modest number of hospitals, seems to support a system where inherited inefficiencies do not snowball uncontrollably. Hospitals, despite receiving structural burdens from above, manage to absorb them without significant internal escalation.

Cumulatively, the inherited inefficiency from the provincial and regional levels in Nova Scotia totals 11.10% (7.02% + 4.08%), which represents 76.1% of the province's total observed inefficiency of 14.63%. This further reinforces the core argument of this study: inefficiency is primarily a systemic governance issue rather than a hospital-level problem alone.

This points to a key policy implication: while reducing inefficiency at the top remains vital, provinces can also focus on fortifying hospital governance structures to contain the inefficiencies they inevitably inherit. Nova Scotia's experience suggests that well-managed hospitals, operating within a tightly coordinated and geographically coherent system, can serve as effective barriers against the full transmission of systemic inefficiency, regardless of what occurs upstream. However, the question still remains: how much can be achieved by improving hospital operations alone, when a significant portion of inefficiency is already embedded at the provincial level? In Nova Scotia, as in Alberta, the largest single source of inefficiency originates from the provincial tier. This reinforces that hospital-level discipline, while necessary, cannot substitute for systemic reform. Addressing the structural inefficiencies at the provincial level—where funding delays, policy fragmentation, and administrative overhead are first introduced—is essential if efficiency gains at lower levels are to be sustained or scaled. Without tackling inefficiency at its source, downstream containment will always face a ceiling.

Table 5: Hierarchical Decomposition of Inefficiency in Ontario's health system

Province	Base Provincial Inefficiency	Province → To Region	Region	Regional Inefficiency (Base + Inherited)	Region → To Hospital	Hospital Inefficiency† (Base + Inherited)	Total Inefficiency† (Province + Region + Hospital)
Ontario (123 hospitals)	6.97 %	3.46 %	1	4.14 %	2.08 %	2.76 %	14.47 %
			2	4.19 %	2.09 %	5.11 %	17.15 %
			3	4.12 %	2.08 %	2.87 %	14.58 %
			4	4.18 %	2.08 %	2.55 %	14.28 %
			5	4.14 %	2.04 %	2.73 %	14.44 %
			6	4.13 %	2.02 %	2.91 %	14.63 %
			7	4.17 %	2.07 %	2.86 %	14.62 %
			8	4.11 %	2.02 %	2.90 %	14.60 %
			9	4.11 %	2.02 %	2.62 %	14.28 %
			10	4.22 %	2.10 %	2.85 %	14.66 %
			11	4.16 %	2.07 %	2.64 %	14.36 %
			12	4.12 %	2.04 %	2.72 %	14.40 %
			13	4.19 %	2.09 %	5.02 %	17.04 %
			14	4.15 %	2.07 %	3.02 %	14.78 %
Grand average				4.15 %	2.06 %	3.11 %	14.88 %
Grand Average Efficiency Score							85.12 %
† Averaged at Regional Level							

Table 5 presents the hierarchical decomposition of inefficiency in Ontario's health system, the third and final province under review. As Canada's most populous province, with nearly 16 million residents and the largest geographic area among the three (over 1,076,000 km^2), Ontario offers a complex healthcare landscape. It operates with 123 hospitals, which is still notably more than Nova Scotia's 13 or Alberta's 83, spanning from highly urbanised southern regions to sparsely populated and logistically challenging northern territories. This mix of scale and geographic diversity adds a further dimension to how inefficiency is distributed and managed within its multi-tiered system (Statistics Canada, 2021c).

Ontario begins with the lowest base provincial inefficiency of the three provinces, at 6.97%, compared to 7.02% in Nova Scotia and 7.25% in Alberta. While this difference is not large in absolute terms, it marks Ontario as starting with a slightly leaner provincial structure. However, as inefficiency flows down to the regional and hospital levels, the pattern becomes more complex.

At the regional level, Ontario's regions inherit an average of 3.46% inefficiency from the provincial tier—more than Nova Scotia (3.41%) but less than Alberta (4.17%). Once self-generated inefficiency is accounted for, total regional inefficiency averages 4.15%, placing Ontario between Alberta (4.73%) and Nova Scotia (4.08%). This implies that Ontario's regions operate with a moderate level of internal inefficiency and only modestly add to what they inherit. Yet, when disaggregated by region, some variation emerges. For example, Regions 2 and 13 show total inefficiencies of 17.15% and 17.04% respectively, indicating points in the system where upstream burdens and internal dynamics compound more visibly.

At the hospital level, Ontario's numbers again occupy a middle position. Hospitals inherit an average of 2.06% inefficiency from the regional tier, nearly identical to Nova Scotia (2.02%) and slightly below Alberta (2.36%). Hospital inefficiency, however, averages 3.11%, exceeding Nova Scotia's 2.91% but staying below Alberta's 3.66%. As with regional inefficiency, the pattern here is mostly stable, but outliers appear—most notably in Region 2, where hospital inefficiency reaches 5.11%, and in Region 13 at 5.02%, suggesting certain clusters of operational or contextual strain.

Taken together, Ontario's total system inefficiency averages 14.88%, translating to an efficiency score of 85.12%. Cumulatively, the inherited inefficiency from the provincial and regional levels totals 11.12% (6.97% + 4.15%), which represents 74.7% of Ontario's total observed inefficiency of 14.88%. This highlights that, even in Ontario's large and institutionally sophisticated system, inefficiency primarily originates upstream rather than locally.

Ontario's internal diversity partly explains this mixed picture. The densely populated south benefits from established infrastructure and administrative coherence(Statistics Canada, 2021c). In contrast, northern Ontario's vast distances and rugged terrain—much like Alberta—create structural challenges for delivery, often requiring specialised services like air ambulances(Alberta Health, 2021; Statistics Canada, 2021a). As such, while Ontario's average inefficiency remains moderate, its geographic heterogeneity continues to exert pressure on certain regions and hospitals more than others.

It is also worth briefly noting Ontario's performance through a different methodological lens. A DEA study of 113 acute-care hospitals in Ontario by Chowdhury and Zelenyuk (2016) found that efficiency varies significantly across hospital size, teaching status, and geography. Interestingly, that study noted that small and rural hospitals were more variable in efficiency but not necessarily less efficient, challenging common assumptions that smaller units are automatically less productive. Furthermore, teaching hospitals were found to be relatively efficient, contrary to many studies in other jurisdictions. These findings, while not directly comparable to the hierarchical inefficiency measures used in this study, still reinforce the broader point: organisational and contextual factors—not scale alone—determine efficiency at the hospital level.

Ontario's hospital sector, with over 120 institutions, reflects the complexity and opportunity that come with scale. On the one hand, the sheer number of hospitals allows for redundancy, flexibility, and greater institutional specialisation, which can help diffuse inherited inefficiencies and protect the system from systemic shocks. On the other hand, it introduces coordination challenges that may explain why certain regions (e.g., Regions 2 and 13) still register elevated inefficiency levels despite the province's otherwise favourable starting point.

This brings us back to the core insight of this study. Even in Ontario, where hospital-level inefficiency is relatively contained and regional variation is modest, the base provincial inefficiency remains the single largest contributor to overall system inefficiency. The province may have more institutional capacity to absorb and manage that burden, but the structural inefficiencies introduced at the top continue to shape outcomes downstream. As in

Alberta and Nova Scotia, this underscores the limits of reforming only the frontline or regional levels. Provincial governance remains the critical leverage point for achieving sustainable reductions in system-wide inefficiency.

This study offers, to the best of current knowledge, the first detailed empirical decomposition of inefficiency across multiple administrative tiers within a national healthcare system—not only in Canada, but within the broader healthcare efficiency literature. While prior research has assessed hospital-level or system-wide inefficiency using frontier methods such as Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA), no previous study has mapped how inefficiency is transmitted, inherited, and compounded across governance layers—from province to region to hospital—with this degree of granularity. The hierarchical inefficiency framework developed here provides a new lens for understanding where inefficiency originates, how it propagates, and where it is absorbed or amplified within the system. The analysis also identifies specific regions with disproportionately high cumulative inefficiency—regions that could serve as targets for more focused evaluation and policy intervention. However, the practical applicability of these findings is constrained by data confidentiality policies: the Canadian Institute for Health Information (CIHI), despite multiple formal requests, declined to release identifiers for the regions and hospitals included in the study. While understandable from a privacy and institutional standpoint, restricted access to granular identifiers limits the ability to directly translate research findings into operational reforms. Greater engagement with stakeholders and more flexible access policies would enhance the potential for empirical models such as this to contribute directly to system-level improvement efforts.

6. Policy Implications

The empirical results from the HIM-IF model provide important insights for policymakers seeking to enhance efficiency in Canada's healthcare system. A central contribution of this study is the clear distinction between inherited and self-generated inefficiency across provincial, regional, and hospital tiers, offering a more precise diagnostic framework for targeted interventions.

First, the findings indicate that system-level inefficiencies—particularly those originating at the provincial governance layer—are the largest contributors to overall waste. Provincial decisions on budget timing, regulatory design, and administrative structures significantly influence downstream hospital performance. Elasticities estimated from the production frontier suggest that administrative expenditures disproportionately affect resource use (elasticity = 0.360), while hospital bed expansion is similarly resource-intensive (elasticity = 0.529). These results imply that hospital-level management improvements alone will have limited impact without reforms targeting systemic governance bottlenecks. Policies should prioritise streamlining funding flows, aligning incentives across administrative tiers, and ensuring that expansions in administrative and physical capacity are tightly linked to actual service demand.

Second, regional health authorities (RHAs) introduce moderate but non-negligible levels of inefficiency. Their role as intermediaries is critical in either amplifying or mitigating governance failures inherited from higher levels. Some regions, particularly those serving remote or socially vulnerable populations, experience compounding inefficiencies when upstream and local challenges intersect. Although data privacy restrictions currently limit direct identification of specific health zones, the model highlights the importance of region-specific performance monitoring and adaptation. Regional authorities should be empowered to adjust inherited frameworks to local conditions, supported by tailored funding models and performance-based accountability systems. Hospitals, in turn, should strengthen internal coordination between administrative and clinical operations, as the negative interaction between administrative and medical expenditures suggests that integrated workflows can mitigate input growth pressures.

Third, the results underscore the importance of data availability and analytic infrastructure in improving system responsiveness. Current limitations on access to regional and hospital-level identifiers hinder precise intervention targeting. While protecting patient and institutional confidentiality remains essential, establishing controlled-access data environments would enhance policymakers' ability to deploy resources more effectively. Greater investment in integrated health information systems, real-time dashboards, and standardised efficiency metrics across administrative levels would further strengthen the transparency and agility of healthcare governance.

Finally, the findings challenge the assumption that larger, more complex systems are inherently less efficient. Ontario's relatively strong performance, despite its size, illustrates how risk pooling and coordinated resource management can offset inefficiencies typically associated with decentralised governance. However, geographical and infrastructural barriers, particularly in Alberta and northern Ontario, continue to impose structural inefficiencies that require targeted responses. Investments in rural health infrastructure, flexible funding mechanisms for remote areas, and improved regional coordination will be essential to addressing these challenges.

Overall, the HIM-IF framework provides a practical roadmap for reform by distinguishing inefficiencies that must be addressed at the system level from those that can be resolved locally. Efficiency improvement strategies must recognise the interdependence of administrative layers and prioritise governance reforms alongside operational enhancements to achieve sustainable performance gains.

7. Concluding remarks

Improving healthcare system efficiency remains a critical challenge for policymakers worldwide, particularly in decentralised and multi-tiered governance structures. This study addresses this challenge by introducing a hierarchical model that distinguishes between inherited and self-generated inefficiency across multiple administrative levels.

Applying the model to panel data from Ontario, Alberta, and Nova Scotia over the period 2015–2019, the findings demonstrate that a substantial portion of hospital inefficiency originates upstream, at the provincial level, and is subsequently inherited by regional health authorities and hospitals. In Alberta, for instance, inherited inefficiency from provincial governance accounts for approximately 44% of observed hospital inefficiency. Similar patterns of inefficiency propagation are observed even in provinces with differing governance models, such as Ontario's recentralised system and Nova Scotia's smaller-scale health authority. These results highlight that systemic factors, rather than purely local management failures, are the primary drivers of inefficiency within Canada's healthcare system.

The results carry important implications for the design of healthcare reforms. Efficiency improvement efforts that focus solely on hospital management risk addressing only symptoms rather than structural causes. Meaningful gains require tackling inefficiencies embedded within provincial funding mechanisms, regulatory frameworks, and administrative processes. Regional and hospital-level interventions should be understood as complementary to, rather than substitutes for, upstream reforms.

Although the empirical application focuses on Canada, the hierarchical inefficiency framework developed here broadly applies to other decentralised health systems, including those of Australia, Germany, and the United

States. In these contexts, cascading inefficiencies across administrative layers are likely pervasive yet under-measured. This framework offers a generalisable approach for diagnosing where inefficiencies originate and how they can best be addressed.

Several avenues for future research emerge from this work. Extending the framework to dynamic settings with evolving governance reforms would provide insights into the persistence or correction of inefficiencies over time. Incorporating quality-adjusted output measures would allow for a richer understanding of whether inherited inefficiencies translate into higher costs and poorer patient outcomes. Application of the model to other multi-layered public service sectors, such as education or social care, could further validate its utility and extend its policy relevance.

Recognising inefficiency as a cascading, systemic phenomenon rather than a hospital-centric issue represents an important shift for health economics. By providing a new methodological lens and empirical evidence, this study underscores that tackling healthcare inefficiency effectively requires reforms at both the operational and structural levels—an insight increasingly vital as health systems globally confront rising fiscal pressures, demographic change, and growing demands for accountability.

References

Alberta Health. (2021). *Helicopter Emergency Medical Services Report 2021. Government of Alberta*. https://open.alberta.ca/publications/helicopter-emergency-medical-services-report-2021

Alberta Health Services. (2024). About AHS. https://www.albertahealthservices.ca/about/about.aspx

Allin Sara, Grignon Michel, Veillard Jeremy, & Wang Li. (2015). How can health system efficiency be improved in Canada? *Healthcare Policy, 11*(1), 33-45. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4748364/

Allin Sara, Grignon Michel, & Wang Li. (2016). The determinants of efficiency in the Canadian health care system. *Health Economics, Policy and Law, 11(1)*(39-65). https://doi.org/10.1017/S1744133115000274

Andrews, A., & Emvalomatis, G. (2024). Efficiency measurement in healthcare: the foundations, variables, and models–a narrative literature review. *Economics, 18*(1), 20220062.

Aritz, A., Elias Teixeira, K., Amanda, L., Zhedong, L., Joaquín, M.-M., & Håvard, R. (2024). Automatic cross-validation in structured models: Is it time to leave out leave-one-out? *Spatial Statistics, 62*, 100843. https://doi.org/https://doi.org/10.1016/j.spasta.2024.100843

Baker, G. R., Norton, P. G., Flintoft, V., Blais, R., Brown, A., Cox, J., Etchells, E., Ghali, W. A., Hébert, P., Majumdar, S. R., O'Beirne, M., Palacios-Derflingher, L., Reid, R. J., Sheps, S., & Tamblyn, R. (2004). The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ: Canadian Medical Association Journal, 170*(11), 1678-1686. https://doi.org/10.1503/cmaj.1040498

Boychuk, G. W. (2008). *National health insurance in the United States and Canada: Race, territory, and the roots of difference*. Georgetown University Press.

Bruce, C. R., Fetter, J. E., & Blumenthal-Barby, J. S. (2013). Cascade Effects in Critical Care Medicine: A Call for Practice Changes. *American Journal of Respiratory and Critical Care Medicine, 188*(12), 1384-1385. https://doi.org/10.1164/rccm.201309-1606ED

Canadian Institute for Health Information. (2014). *Measuring the Level and Determinants of Health System Efficiency in Canada*. https://www.cihi.ca/en/measuring-the-level-and-determinants-of-health-system-efficiency-in-canada

Canadian Institute for Health Information. (2025). *Health Workforce in Canada, 2019 to 2023: Overview - Methodology Notes*. https://www.cihi.ca/en/health-workforce-in-canada

Cantor Victor John M, & Poh Kim Leng. (2018). Integrated analysis of healthcare efficiency: a systematic review. *Journal of Medical Systems, 42*, 1-23.

Chowdhury, H., & Zelenyuk, V. (2016). Performance of hospital services in Ontario: DEA with truncated regression approach. *Omega, 63*, 111-122. https://doi.org/10.1016/j.omega.2015.10.004

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1973). Transcendental logarithmic production frontiers. *The review of economics and statistics*, 28-45.

Colombi, R., Martini, G., & Vittadini, G. (2017). Determinants of transient and persistent hospital efficiency: The case of Italy. *Health economics, 26*, 5-22.

Eklom, B., & Callander, E. (2020). A Systematic Review of Hospital Efficiency and Productivity Studies: Lessons from Australia, UK and Canada.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal, 1*(3), 515-534. https://doi.org/10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and computing, 24*, 997-1016.

Giancotti, M., Guglielmo, A., & Mauro, M. (2017). Efficiency and optimal size of hospitals: Results of a systematic search. *PloS one, 12*(3), e0174533.

Health Canada. (2023a). *Canada's Health Care System*. https://www.canada.ca/en/health-canada/services/canada-health-care-system.html

Herr, A. (2008). Cost and technical efficiency of German hospitals: does ownership matter? *Health economics, 17*(9), 1057-1071.

Hollingsworth, B., & Peacock, S. J. (2008). *Efficiency measurement in health and health care*. Routledge.

Jacobs, R., Smith, P. C., & Street, A. (2006). *Measuring efficiency in health care: analytic techniques and health policy*. Cambridge University Press.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai,

Lipsitz, L. A. (2012, Jul 18). Understanding health care as a complex system: the foundation for unintended consequences. *Jama, 308*(3), 243-244. https://doi.org/10.1001/jama.2012.7551

Liu, Z., & Rue, H. (2022). Leave-group-out cross-validation for latent gaussian models. *arXiv preprint arXiv:2210.04482*.

Marchildon G. (2013). Canada: Health system review. *Health systems in transit, 15*(1), 1-179.

Marchildon, G. P., Allin, S., & Merkur, S. (2020). Canada: health system review. *Health systems in transit, 22*(3), 1-194.

McConnell, C. R., Brue, S. L., & Flynn, S. M. (2018). *Microeconomics: principles, problems, and policies*. McGraw-Hill.

Mutter, R. L., Rosko, M. D., & Wong, H. S. (2008). Measuring hospital inefficiency: the effects of controlling for quality and patient burden of illness. *Health services research, 43*(6), 1992-2013.

Nova Scotia Health. (2023). About Us. https://www.nshealth.ca/about-us

Ontario Health. (2023). About Us. https://www.ontariohealth.ca/about-us

Organization of Economic Co-operation and Development. (2017). *Tackling wasteful spending on health*. OECD Publishing Paris.

Renger, R., Foltysova, J., Ienuso, S., Renger, J., & Booze, W. (2017). Evaluating system cascading failures. *Evaluation Journal of Australasia, 17*(2), 29-36.

Rosko, M. D., & Mutter, R. L. (2008). Stochastic frontier analysis of hospital inefficiency: a review of empirical issues and an assessment of robustness. *Medical care research and review, 65*(2), 131-166.

Statistics Canada. (2021a). *Focus on Geography Series, 2021 Census; Alberta. Government of Canada*. https://www12.statcan.gc.ca/census-recensement/2021/as-sa/fogs-spg/Page.cfm?Lang=E&Dguid=2021A000248&topic=1

Statistics Canada. (2021b). Focus on Geography series 2021 Census; Nova Scotia. Government of Canada. https://www12.statcan.gc.ca/census-recensement/2021/as-sa/fogs-spg/Page.cfm?dguid=2021A000212&lang=E&topic=1

Statistics Canada. (2021c). Focus on Geography series 2021 Census; Ontario. Government of Canada. https://www12.statcan.gc.ca/census-recensement/2021/as-sa/fogs-spg/Page.cfm?Lang=E&Dguid=2021A000235&topic=1

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological), 36*(2), 111-133.

Sutherland, J. M., & Crump, R. T. (2013). Alternative Level of Care: Canada's Hospital Beds, the Evidence, and Options. *Healthcare Policy*(9), 24-36.

Tsionas, E. G. (2006). Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics, 21*(5), 669-676.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing, 27*, 1413-1432.

World Health Organization. (2010). *The world health report: health systems financing: the path to universal coverage*. World Health Organization. https://iris.who.int/handle/10665/44371

World Health Organization, Organization of Economic Co-operation and Development, & The World Bank. (2018). *Delivering quality health services: a global imperative for universal health coverage*. WHO Publication.

Worthington, A. C. (2004). Frontier efficiency measurement in health care: a review of empirical techniques and selected applications. *Medical care research and review, 61*(2), 135-170.