Beyond Traditional Inefficiency Measures: Quantifying Health System Waste Through a Hierarchical Model of Inherited and Self-Generated Persistent Inefficiency

Tuaine Junior Unuia

A thesis submitted to
Auckland University of Technology
in partial fulfilment of the requirements for the
degree of Master of Business (MBus)

2025
Business School

Abstract

This study develops a hierarchical inefficiency model to quantify how persistent technical inefficiency is generated and transmitted through multi-tiered healthcare systems. Departing from conventional hospital-centric assessments, the model decomposes inefficiency into inherited (propagated from higher governance levels) and self-generated (locally produced) components across three administrative tiers: province, region, and hospital. By leveraging the nested structure of Canadian healthcare governance, the framework captures system-level inefficiencies embedded in institutional design rather than isolated provider-level performance.

Applied to panel data from hospitals in Alberta, Nova Scotia, and Ontario, the analysis shows that persistent inefficiency consistently originates at the top of the hierarchy, within provincial governance, where it is highest: 7.25% in Alberta, 7.02% in Nova Scotia, and 6.97% in Ontario. It then compounds as it flows downward, yielding total system inefficiencies of 16.45%, 14.63%, and 14.88%, respectively. These results demonstrate that hospital-level inefficiencies often reflect upstream structural constraints rather than solely local mismanagement.

While the decomposition provides a tractable diagnostic of how inefficiency propagates, it also raises a policy dilemma: Should reforms equip hospitals and regions to absorb inherited burdens, or should they target the persistent sources of inefficiency embedded in provincial governance? These findings challenge standard health economics approaches that localise inefficiency at the point of care. By reframing inefficiency as a cascading, structural phenomenon, this study offers a system-aligned perspective for identifying where meaningful reform must begin.

Table of Contents

List of Tables

Attestation of Authorship

"I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor used artificial intelligence tools or generative intelligence tools (unless it is clearly stated, and referenced, along with the purpose of use), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning."

11 July 2025

Acknowledgments

First and foremost, I thank God for His presence throughout this journey. I am thankful for His many blessings, and His strength carried me through the challenges I faced along the way. I am forever grateful for His grace.

This thesis is dedicated to my grandparents, **Tou Matapo Tuaine Unuia** and **Tangata-Tupu-Kiikii Unuia**, whose quiet sacrifices and unwavering support made this achievement possible. You instilled in me the belief that education matters and gave so much to help me pursue dreams that once lived only in your hearts. I hope you are both proud. This belongs to you.

I want to express my sincere and deepest gratitude to **Sean Kimpton** for his consistent support throughout this journey. From shaping my research and strengthening my academic writing to encouraging me in teaching and presentation opportunities, his guidance has been invaluable. I am especially grateful for the trust he placed in me and for being a supervisor who genuinely advocates for his students.

I also wish to thank **Dr. Antony Andrews** for his guidance in developing the modelling components of this thesis. His support in learning RStudio, coding, and applying Bayesian methods helped me build a strong technical foundation for this research.

I am deeply grateful to the many people, institutions, and scholarship providers who supported me throughout this journey, whether through financial assistance, encouragement, or guidance. Your contributions, seen and unseen, helped make this thesis possible. This thesis is a testament to the collective support and encouragement that pushed me forward.

Thank you all for your belief in me.

Chapter 1: Introduction

In an era of rising healthcare demands, ageing populations, and increasing financial constraints, the need for an efficient healthcare system has never been more critical. As medical advancements extend life expectancy and chronic diseases become more prevalent, healthcare systems worldwide face mounting pressure to deliver high-quality care while managing limited resources effectively. Governments and policymakers are increasingly focused on optimising healthcare expenditures, reducing waste, and improving service delivery to meet the growing needs of their populations (Organisation for Economic Co-operation and Development, 2017; World Health Organisation et al., 2018). However, achieving efficiency is not simply a matter of cutting costs; it requires a deeper understanding of how inefficiencies emerge and propagate within complex, multi-level healthcare systems.

Healthcare systems worldwide face persistent inefficiencies that undermine service delivery, inflate costs, and reduce patient outcomes. While traditional (in)efficiency measures focus on hospital-level performance, encompassing both technical and cost inefficiency, 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 problem but a cascading phenomenon that originates from systemic governance structures, resource allocation mechanisms, and bureaucratic bottlenecks (Organisation for Economic Co-operation and Development, 2017; World Health Organisation et al., 2018). These inefficiencies accumulate as healthcare resources flow through different administrative tiers, creating substantial waste before funds and services reach patients. Despite growing recognition of inefficiencies within healthcare, no studies to date have examined inefficiency as a multi-level structural issue, where each administrative level not only generates its own inefficiency but also inherits inefficiencies from the levels above.

In this study, we introduce a fundamental distinction in health system inefficiency by differentiating inherited inefficiency from self-generated inefficiency. We postulate that inherited inefficiency refers to structural and policy-driven constraints passed down from higher administrative levels, such as delays in funding, regulatory fragmentation, or misaligned resource allocation, which exist before lower levels of administration even begin to function. Self-generated inefficiency, by contrast, arises due to internal mismanagement, excess administrative overhead, or inefficient workforce distribution at each level. The combined effect of inherited and self-generated inefficiencies results in compounded inefficiency; whereby healthcare systems experience progressive waste at every stage of administrative decision-making and service provision. Understanding how inefficiency flows through hierarchical structures is crucial for designing interventions that address both structural and operational inefficiencies rather than focusing solely on frontline providers.

The Canadian healthcare system presents an ideal setting for studying hierarchical inefficiency due to its multi-tiered administrative structure. As a publicly funded but decentralised system, Canada delegates significant responsibility to provincial, regional, and hospital levels, each of which plays a role in determining healthcare access, funding allocation, and operational priorities. While the federal government provides broad regulatory oversight, provinces control budgets and policy direction, while regional health authorities oversee implementation, and hospitals operate independently within these constraints (Health Canada, 2023a; Marchildon et al., 2020; Martin et al., 2018). This structure makes it possible to analyse how inefficiencies emerge, propagate, and accumulate across different layers of governance, influencing overall system performance and patient care quality.

At the provincial level, governments control healthcare budgets, establish regulatory frameworks, and dictate the overarching policy direction (Health Canada, 2023a; Martin et al., 2018). However, bureaucratic inefficiencies, such as delayed budget approvals, fragmented decision-making, and administrative redundancies, can create inefficiencies before resources even reach regional authorities. For example, slow funding disbursements, politically motivated spending decisions, and inefficient procurement policies can reduce the effectiveness of available resources long before they are allocated to hospitals (Canadian Institute for Health Information, 2014). At the regional level, inefficiencies often arise from unequal resource distribution, staffing imbalances, or slow implementation of provincial policies, further exacerbating disparities in healthcare access. Some regions may experience hospital overcrowding, while others suffer from underutilised medical facilities due to misaligned resource coordination (Canadian Institute for Health Information, 2025; Marchildon et al., 2020; Martin et al., 2018). Additionally, delays in approving medical equipment purchases or infrastructure projects can further strain hospital operations.

Hospitals, at the final stage of the system, inherit inefficiencies from above while also generating their own. Even when funding reaches hospitals, inefficiencies persist due to operational bottlenecks, inefficient staff scheduling, and redundant administrative procedures. Hospital-level inefficiencies include ineffective bed management, over-reliance on unnecessary diagnostics, coordination issues within different units of hospitals and delays in patient discharge processing, all of which contribute to increasing healthcare costs and reducing system capacity (Fixler et al., 2014; Sutherland & Crump, 2013). By the time healthcare services are delivered to patients, a significant portion of the originally allocated resources has been lost to inefficiency.

The logic of this framework is that inefficiency is not a siloed or static phenomenon but a flowing, layered burden. Each level retains its base inefficiency while inheriting additional inefficiency from the level above. When upstream inefficiencies are not absorbed but merely passed down, they accumulate and intensify, undermining performance at the very point where services are delivered. This highlights the need for system-wide reforms that begin not just at the frontline but at the higher administrative layers, where inefficiencies originate and where their downstream impact can be either contained or amplified.

A practical way to conceptualise the cumulative nature of inefficiency in healthcare is through a hypothetical budget-based analogy, which demonstrates how inefficiencies accumulate at each level. Suppose a provincial government begins with a $100 budget for healthcare. Due to policy inefficiencies and administrative waste, only $96 remains effective when it reaches regional authorities. As regional inefficiencies compound, another $2 is lost, reducing the available funding to $94. The regions themselves introduce inefficiencies, further reducing the effective amount to $93. When hospitals receive the funding, an additional $1 is lost due to delays in budget transfers and miscommunication between administrative bodies, leaving $92 available for direct hospital operations. At the bottom of the hierarchy, hospitals introduce their own inefficiencies, such as overuse of resources, misallocation of staff, and bureaucratic delays, resulting in only $90 worth of actual healthcare services being delivered to patients. While inefficiency is fundamentally a technical phenomenon, measuring its impact in financial terms makes the concept more tangible and quantifiable. Table 1 illustrates how inefficiencies accumulate at different levels, highlighting the role of both inherited and self-generated inefficiencies in the overall loss of healthcare resources.

Table 1

Illustrative Example of How Inefficiency Accumulates Across the Healthcare System

Level of Hierarchy	Initial Budget	Remaining Budget After Inefficiency	Cumulative Impact from Initial Budget (%)	Why is There a Loss? (Possible reason for Technical Inefficiency)
Province	$100 → $96	4% lost at the provincial level	4%	Administrative overhead, political decisions, excessive bureaucracy. Example: The province spends too much on management and policy meetings.
Region (Inherited)	$96 → $94	2.08% lost just because the region happens to be in this province	6.08%	Delayed funding allocation and bureaucratic inefficiency. Example: Provinces take too long to send funding, requiring excessive paperwork, additional spending proofs, and multiple rounds of information requests from regions before approving funds
Region (Own Issues)	$94 → $93	1.06% lost due to regional inefficiency	7.14%	Poor resource allocation, staffing imbalances, bureaucratic delays, budget mismanagement, inefficient hospital coordination, and infrastructure limitations leading to unequal healthcare access, service delays, and increased operational costs

Hospital (Inherited)	$93 → $92	1.08% lost just because they are in the region	8.22%	Delayed budget approvals, miscommunication, uneven resource distribution, slow procurement processes, and fragmented hospital coordination causing service delays, equipment shortages, and operational bottlenecks
Hospital (Own Issues)	$92 → $90	2.17% lost due to hospital inefficiency	10.40%	Service coordination issues, unnecessary tests, overallocation of resources and patient discharge delays. Example: Hospitals waste money on unnecessary tests or inefficient patient management.
Total inefficiency in the system	**10% (approx.)**			
Note: This table presents a hypothetical example for illustrative purposes. Figures are not empirical estimates but are stylised to reflect the structure and flow of inefficiency described in the text.				

This structured breakdown in the table highlights how inefficiencies flow through different levels of the healthcare system, compounding their effects before resources reach patients. Each administrative level inherently loses efficiency due to inherited inefficiencies while also introducing its own inefficiencies, further reducing the system's ability to maximise resources.

Consistent with complexity science, the healthcare system functions as a complex adaptive system, where inefficiencies at one level, such as rigid policy mandates or regulatory delays, can cascade downward, generating unintended consequences at the operational level (Lipsitz, 2012). A striking example is the 3-day Medicare rule, which requires patients to remain hospitalised for at least three consecutive days to qualify for subsidised skilled nursing facility care. Although originally intended as a cost-containment measure, the policy has produced downstream inefficiencies, unnecessary hospital stays, excessive diagnostic testing, and delayed discharges that are not the result of hospital-level mismanagement, but rather structural constraints imposed from higher administrative tiers. Bruce et al. (2013) reinforce this interpretation by illustrating how cascade effects in critical care often originate from systemic defaults and policy triggers, such as automatic progression to intensive interventions, that reflect higher-order institutional and regulatory decisions. These upstream determinants set in motion treatment trajectories that are difficult to halt, resulting in overtreatment, resource strain, and ethical dilemmas at the bedside.

This highlights the core premise of my study: that inefficiency in healthcare is not solely a localised phenomenon but includes inherited inefficiencies that originate upstream in the governance and regulatory architecture. Recognising the difference between inherited and self-generated inefficiency is critical for diagnosing where waste emerges and how it accumulates across multi-tiered systems. Renger et al. (2017) further reinforce this perspective by showing how disruptions in one subsystem, such as delays or communication failures, can propagate through interconnected processes, creating cascading failures that undermine efficiency throughout the system. Their work underscores the need for system-level thinking and supports my call for a hierarchical inefficiency model that accounts for both the vertical flow of structural inefficiency and the horizontal inefficiencies generated within each level. This systems-based understanding is essential for designing targeted reforms that address the root causes of inefficiency rather than simply treating its symptoms.

Despite the recognition of inefficiency as a major challenge in healthcare, existing inefficiency models fail to incorporate the hierarchical structure of inefficiency flow. Traditional models treat inefficiency as an isolated phenomenon at the level of individual hospitals or healthcare providers without acknowledging that inefficiencies are structurally embedded within multi-tiered healthcare systems. This study introduces a novel hierarchical inefficiency model that accounts for the multi-level nature of inefficiency, incorporating both the flow of inherited inefficiency from higher levels and self-generated inefficiencies at each administrative tier. By leveraging a hierarchical Bayesian framework, this approach borrows strength across different levels of the system, allowing for a more accurate estimation of inefficiency at each level. Unlike traditional efficiency models, which assess inefficiency as a static measure, this model captures inefficiency as a dynamic, interdependent process where the efficiency of one level directly influences the inefficiency of another.

This approach is particularly valuable for policymakers and health system planners, as it enables targeted interventions that address not only frontline inefficiencies but also the structural inefficiencies embedded within healthcare governance. By understanding inefficiency as a multi-level phenomenon, policymakers can design more effective resource allocation strategies, improve administrative decision-making, and enhance overall healthcare system performance. As healthcare systems worldwide continue to struggle with resource constraints and growing demand, a hierarchical inefficiency framework offers a novel and more precise way to diagnose and address systemic inefficiencies.

Chapter 2: The Canadian Healthcare System

Canada's healthcare system was founded on the principle that all residents should have access to essential medical care based on need and not their ability to pay (Martin et al., 2018). The Canadian healthcare system is intended to provide universal access to hospital and physician services at no direct cost to patients. However, it is not a single national program. Instead, it comprises thirteen independently operated provincial and territorial insurance plans. Each plan is independently operated by its respective province or territory. However, all must follow the same five national principles set out in the Canada Health Act (1984): public administration, comprehensiveness, universality, portability, and accessibility. In exchange for compliance, provinces and territories receive federal health transfers to support their healthcare delivery systems (Marchildon et al., 2020).

The federal government plays a central role by setting national standards and providing financial support, primarily through the Canada Health Transfer. It also delivers healthcare services directly to specific groups, including Indigenous Peoples, veterans, and federal inmates (Boychuk, 2008; Health Canada, 2023a). The system's origins trace back to the 1950s and 1960s when Saskatchewan introduced public hospitals and medical insurance. These efforts laid the foundation for a national system, later consolidated under the Canada Health Act. While the system has succeeded in ensuring equitable access to core services, it remains fragmented in areas outside the Act's defined coverage, such as outpatient prescriptions and dental care (Martin et al., 2018).

While the federal government sets national standards and provides financial support, provinces are primarily responsible for designing and operating the health services that citizens interact with daily. At the provincial level, governments in Canada oversee the organisation, funding, and delivery of healthcare services within their jurisdictions. They are responsible for establishing healthcare policies, managing public health programs, regulating hospital and physician services, and ensuring alignment with federal standards such as those outlined in the Canada Health Act 1984. Provinces finance these services primarily through general taxation and play a central role in negotiating agreements with healthcare professionals, overseeing infrastructure development, and addressing regional disparities in service access. Their broad authority enables them to direct resources and shape long-term strategies that respond to demographic trends and evolving health needs, such as population ageing and rising wait times (Martin et al., 2018).

The actual delivery of healthcare services, however, is often administered through Regional Health Authorities (RHAs), which function as operational arms of provincial health systems. These RHAs are tasked with coordinating the provision of hospital care, community health services, and public health initiatives within their respective geographic areas. They work to implement provincial policies on a local scale, manage facilities and workforce deployment, and respond to specific health concerns within communities. This tier is essential for translating provincial priorities into practical outcomes and for tailoring care to local population needs. The decentralised nature of RHAs is intended to enhance responsiveness and efficiency, yet it has also contributed to uneven service quality, fragmented governance, and variations in access across regions (Martin et al., 2018).

For instance, Ontario, the most populous province with approximately 15.6 million residents, operates the country's largest and most complex system through Ontario Health, which integrates hospital, primary, community, and mental health services across diverse urban and rural regions (Ontario Health, 2023; Statistics Canada, 2021c). Alberta, with about 4.7 million residents, employs a centralised model under Alberta Health Services (AHS), which manages province-wide service delivery across both rural and urban areas (Alberta Health Services, 2024; Health Canada, 2023b; Statistics Canada, 2021a). Nova Scotia, home to just over 1 million people, presents a dual-authority model comprising the Nova Scotia Health Authority and the Izaak Walton Killam (IWK) Health Centre. The province continues to face challenges related to rural access, staffing shortages, and diagnostic delays (Nova Scotia Health, 2023; Statistics Canada, 2021b). These three provinces represent a meaningful cross-section of Canada's healthcare system, ranging from large to medium and small in scale and offer contrasting organisational approaches that provide a strong foundation for assessing structural inefficiency across tiers.

Public funding is primarily sourced from general tax revenues collected by federal, provincial, and territorial governments. Approximately 70% of total health spending is publicly financed, combining local expenditures and federal transfers such as the Canada Health Transfer (Canadian Institute for Health Information, 2023; Martin et al., 2018). Provinces may also levy non-compulsory health premiums to supplement their healthcare budgets. These optional health premiums are voluntary payments that individuals or families may choose to make to access services not included in the standard publicly funded healthcare coverage. While these payments are not necessary for core services such as hospital and physician care, they can help reduce out-of-pocket costs for extras such as prescription medications, dental care, or physiotherapy (Martin et al., 2018). This funding structure ensures that hospital and physician services are widely accessible at no direct cost to patients. However, it leaves other sectors, such as long-term care, mental health, and dental services, unevenly covered, resulting in variation across jurisdictions.

Healthcare costs in Canada have consistently grown faster than the economy and government revenues, raising concerns about long-term fiscal sustainability. While earlier increases were driven mainly by pharmaceutical spending, more recent cost growth is primarily linked to hospital expenditures and physician remuneration (Marchildon et al., 2020). These rising costs put pressure on provincial budgets and challenge the system's ability to expand or standardise coverage beyond core services.

Canada's layered governance structure, federal, provincial, regional, and institutional, not only shapes service delivery but also enables the transmission and compounding of inefficiencies, making it a natural setting for studying inherited and persistent inefficiency.

Marchildon et al. (2020) suggest that such variation is an expected feature of decentralised governance, but when compounded by uneven infrastructure and funding, it exacerbates systemic inefficiencies. These issues are particularly severe in rural and northern regions, where workforce shortages and underfunded facilities persist, thereby limiting access to timely, quality care and contributing to performance gaps across multiple tiers of the system (Martin et al., 2018). As of 2023, one in six Canadians lacked a regular family physician, and fewer than half could see a provider on the same or next day, significantly limiting early diagnoses, continuity of care, and timely specialist referrals (Flood et al., 2023)

These access issues reflect just one dimension of broader inefficiencies entrenched in the system. Despite high levels of public spending, Canada underperforms on many efficiency benchmarks. In 2021, the Commonwealth Fund ranked Canada 10th out of 11 high-income countries in overall healthcare system performance (Schneider et al., 2021). According to the Canadian Institute for Health Information (CIHI), between 12,600 and 24,500 preventable deaths occur each year due to delays in access, uncoordinated care, and system-level fragmentation (Canadian Institute for Health Information, 2014). Although Canada ranked 4th in healthcare spending as a share of GDP and 9th per capita, it consistently underperformed on availability and access indicators. Among 31 countries with universal healthcare, it ranked 28th for physician availability and 25th for hospital beds. It also ranked lowest for timely access, with 65.2% of patients waiting over a month to see a specialist and 58.3% waiting over two months for non-emergency surgery (Moir & Barua, 2024b). These rankings suggest that high expenditure alone does not guarantee access or availability, highlighting a misalignment between investment and outcomes in Canada's model.

The Fraser Institute reported a record-high median wait time of 30 weeks from referral to treatment in 2024, the longest since tracking began in 1993. Delays were particularly severe in orthopaedic (57.5 weeks) and neurosurgical (46.2 weeks) procedures, with diagnostic waits for magnetic resonance imaging (MRI) (16.2 weeks) and computed tomography (CT) scans (8.1 weeks) also elevated (Moir & Barua, 2024a). These delays have been linked to suffering, lost productivity, and reduced quality of life.

Allin Sara et al. (2015) highlight that inefficiencies are not just institutional but structural, stemming from misaligned incentives, inadequate use of performance data, and coordination failures. Health Canada (2025) acknowledges that staffing shortages, limited digital infrastructure, and regional inequities remain significant barriers to care. CIHI studies confirm that differences in hospital performance are shaped not only by internal management but also by upstream policy choices, including resource distribution and funding models (Canadian Institute for Health Information, 2014). The Canadian Medical Association (CMA) (2023) reported that physicians spend an average of 10.6 hours per week on administrative tasks, with 38% considered unnecessary. These include duplicative form-filling, inconsistent reporting requirements, and outdated billing systems. Nova Scotia's 10% reduction target is a rare example of reform, but efforts remain fragmented. Broader studies identify administrative fragmentation, lack of data comparability, and jurisdictional overlap as persistent barriers to integration (Flood et al., 2023; Marchildon et al., 2020).

Delays in disbursing funds and approving services exacerbate wait times and reduce responsiveness. These delays, already reflected in nationally recorded wait times, are tied to slow capital investments and rigid cost-containment protocols (Moir & Barua, 2024b). Naylor (2018) argues that incremental reform under fiscal pressure has created inertia that delays facility upgrades and human resource expansion. Similarly, the Office of the Auditor General of Canada (2015) reported chronic delays in infrastructure renewal due to fractured accountability. Although publicly funded, provincial health budgets are often shaped by political priorities. Since the 1980s, health spending has grown faster than social spending despite evidence linking higher social investment to better outcomes (Dutton et al., 2018). Misaligned budgeting models reduce allocative efficiency, meaning resources are not aligned with the services patients value most (Marchildon et al., 2020)

Interprovincial disparities reflect structural inefficiencies in Canada's decentralised system. Canadian Institute for Health Information (2014) found that differences in provincial performance are influenced as much by upstream decisions as by internal management. Allin Sara et al. (2015) linked regional inefficiencies to socioeconomic burdens such as obesity, low income, and population ageing. These challenges are especially acute in smaller provinces like Nova Scotia (Marchildon et al., 2020). International comparisons show that Canada's regional access and coordination performance trails behind other high-income systems (Schneider et al., 2021)

At the hospital level, inefficiencies include slow procurement, capacity bottlenecks, and poor coordination. Wang et al. (2018) found that Canadian acute hospitals operate at only 75% efficiency. Organisation for Economic Co-operation and Development (2023) reported that bed occupancy reached 92% nationally, leaving no surge capacity. Interoperability gaps delay diagnostics and duplicate care. Alternate level of care (ALC) stays, patients awaiting long-term care (LTC) placement consume 10–20% of hospital bed use in some provinces and are a visible symptom of poor coordination between hospital and long-term care sectors (Sutherland & Crump, 2013). These inefficiencies persist despite high spending and are linked to avoidable hospitalisations.

Frontline inefficiencies, including long wait times and diagnostic overuse, further strain the system. A 2015 audit by the Office of the Auditor General of Canada revealed that in rural and Indigenous communities, frontline workers often operated beyond their scope of practice due to staffing shortages and inadequate training (Office of the Auditor General of Canada, 2015). The audit also found that Health Canada failed to ensure proper training and compliance for nurses in several under-resourced areas. Infrastructure deficiencies, such as incomplete inspections and broken ventilation systems, were also left unaddressed. These shortcomings underscore broader systemic failures in ensuring safe, reliable, and equitable care in underserved regions. Additionally, poor service integration and misaligned responsibilities further hamper service delivery in rural areas (Rechel et al., 2009).

Collectively, these inefficiencies reflect a system in which waste and underperformance are distributed across multiple layers of the healthcare architecture. Despite recognising these challenges, most studies assess efficiency as a single-layer phenomenon, overlooking how decisions and constraints at one level shape outcomes at another. Hierarchical inefficiency refers to waste or underperformance that originates and accumulates across multiple levels of governance, federal, provincial, institutional, and frontline, rather than being isolated to a single tier. This presents a critical gap in the literature. This study seeks to bridge that gap and quantify inefficiency across the healthcare system by adopting a hierarchical modelling framework. Unlike conventional models that treat inefficiency as confined to hospitals or individual providers, this approach traces inefficiency both upward and downward, showing how governance failures cascade into institutional bottlenecks and how local constraints feed back into system-wide underperformance. The contribution lies in disentangling waste attributable to governance, regional disparities, institutional processes, and frontline constraints and providing a more granular understanding of where persistent inefficiencies originate and how they interact.

Chapter 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:

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

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

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 $p-$the 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:

$$u_l \sim \exp\left[Normal\left(\log(\mu_{u_l}), \sigma_{u_l}\right)\right] \quad (8)$$

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

$$u_j \sim \exp\left[Normal\left(\log(u_k), \sigma_{u_{j[k]}}\right)\right] \quad (10)$$

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:

$$u_l \sim \exp\left[Normal\left(\log\left(\mu_{u_l}\right), \sigma_{u_l}\right)\right] \qquad (11)$$

$$u_k \sim \exp\left[Normal\left(\phi_l . \log\left(u_l\right), \sigma_{u_{k[l]}}\right)\right] \qquad (12)$$

$$u_j \sim \exp\left[Normal\left(\phi_k . \log\left(u_k\right), \sigma_{u_{j[k]}}\right)\right] \qquad (13)$$

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 the 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:

$$
\begin{aligned}
&P(\boldsymbol{Z}, u_j, u_k, u_l, \alpha_j, \alpha_k, \alpha_l | \boldsymbol{\theta}) \\
&\qquad = P(Z|\alpha_j, u_j, \beta, \sigma_v, p'_{jt}) \times P\left(\alpha_j \middle| \alpha_k, \sigma_{\alpha_j}\right) \times P(\alpha_k|\alpha_l, \sigma_{\alpha_k}) \times P(\alpha_l|\sigma_{\alpha_l}) \\
&\qquad \times P\left(u_j \middle| u_k, \phi_k, \sigma_{u_{j[k]}}\right) \times P\left(u_k \middle| u_l, \phi_l, \sigma_{u_{k[l]}}\right) \times P(u_l|\mu_{u_l}, \sigma_{u_l})
\end{aligned}
$$

Expanding the likelihood in its parametric form, it becomes:

$$
\begin{aligned}
&P(\boldsymbol{Z}, \alpha_j, \alpha_k, \alpha_l, u_j, u_k, u_l, |\boldsymbol{\theta}) \\
&= \left[\prod_{j=1}^{J}\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma_v^2}}\, exp\left(-\frac{\left(z_{jt} - \boldsymbol{p}'_{jt}\beta - \alpha_j + u_j\right)^2}{2\sigma_v^2}\right)\right] \\
&\times \left[\prod_{j=1}^{J}\frac{1}{\sqrt{2\pi\sigma_{\alpha_j}^2}} \exp\left(-\frac{\left(\alpha_j - \alpha_k\right)^2}{2\sigma_{\alpha_j}^2}\right)\right] \times \left[\prod_{k=1}^{K}\frac{1}{\sqrt{2\pi\sigma_{\alpha_k}^2}} \exp\left(-\frac{(\alpha_k - \alpha_l)^2}{2\sigma_{\alpha_k}^2}\right)\right] \\
&\times \left[\prod_{l=1}^{L}\frac{1}{\sqrt{2\pi\sigma_{\alpha_l}^2}} \exp\left(-\frac{\alpha_l^2}{2\sigma_{\alpha_l}^2}\right)\right] \times \left[\prod_{j=1}^{J}\frac{1}{u_j\sigma_{u_j}\sqrt{2\pi}} \exp\left(-\frac{\left(\log(u_j) - \log(\phi_k \times \log(u_k))\right)^2}{2\sigma_{u_{j[k]}}^2}\right)\right] \\
&\times \left[\prod_{k=1}^{K}\frac{1}{u_k\sigma_{u_k}\sqrt{2\pi}} \exp\left(-\frac{(\log(u_k) - \log(\phi_l \times \log(u_l)))^2}{2\sigma_{u_{k[l]}}^2}\right)\right] \\
&\times \left[\prod_{l=1}^{L}\frac{1}{u_l\sigma_{u_l}\sqrt{2\pi}} \exp\left(-\frac{\left(\log(u_l) - \mu_{u_l}\right)^2}{2\sigma_{u_l}^2}\right)\right]
\end{aligned}
$$

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. However, due to data-sharing restrictions imposed by the Canadian Institute for Health Information (CIHI), the underlying dataset cannot be made publicly available.

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].

Chapter 4: Data and 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 (MED_EXP) 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 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 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 emphasised by Tsionas (2006) and Jacobs et al. (2006, pp. 21-27), ensuring that output measures reflect both the volume and complexity of care.

Table 2 presents the summary statistics of all variables used in the study by province. For transparency and clarity, expenditure-related variables are reported in their raw form, meaning they are neither combined, weighted, nor deflated. This approach allows the actual observed values to be displayed, providing an unadjusted view of the data's distribution and magnitude. However, during modelling, both the deflated and unadjusted data are used depending on the specific requirements of the analysis, ensuring accurate efficiency assessments while maintaining data transparency.

Table 2
Summary Statistics for Alberta

Variable	Observation	Min	1st Qu	Median	Mean	Std. Dev	3rd Qu	Max
MED compensation	295	0.06	0.33	0.71	12.91	34.36	3.72	212.54
UPP compensation	295	2.74	6.58	11.05	71.46	142.32	32.15	721.5
MOS compensation	295	0.54	1.13	2.17	14.5	28.91	6.68	147.55
Supplies expenses	295	0.31	0.95	1.67	13.8	30.83	5.33	168.93
Drug expenses	295	0.01	0.16	0.26	2.54	5.81	0.99	36.71
Contract out expenses	295	0.01	0.06	0.13	3.32	8.67	1.76	61.73
Sundry expenses	295	0.13	0.29	0.51	3.37	7.32	1.75	43.77
Capital	295	16	27	55	144.46	225.47	136	1095
Admitted RIW	295	97.58	676.38	1034.31	8619.87	17946.86	3768.07	87277.43
Outpatients	295	76.86	379.36	810.53	4003.56	7529.33	2484.94	43620.94

Table 3
Summary Statistics for Ontario

Variable	Observation	Min	1st Qu	Median	Mean	Std. Dev	3rd Qu	Max
MED compensation	587	0.08	2.1	6.96	19.04	31.77	23.51	249.3
UPP compensation	587	2.74	11.12	32.05	103.43	151.16	139.21	762.49
MOS compensation	587	0.48	1.84	5.32	19.72	31.92	24.3	197.72
Supplies expenses	587	0.56	1.97	6.28	24.7	39.76	28.81	217.39
Drug expenses	587	0.03	0.26	1.75	11.74	27.58	9.7	259.04
Contract out expenses	587	0.02	0.54	1.52	6.2	12.48	6.24	83.96
Sundry expenses	587	0.33	0.9	2.46	18.03	67.15	11.05	704.36
Capital	587	15	45	105	264.97	336.99	332.5	1510
Admitted_RIW	587	112.22	1000.23	3540.45	13156.74	19758.16	15687.27	90815.7
Outpatients	587	58.47	835.63	2562.05	4734.52	5485.4	6705.61	28056.25

Table 4
Summary Statistics for Nova Scotia

Variable	Observation	Min	1st Qu	Median	Mean	Std. Dev	3rd Qu	Max
MED compensation	60	0.11	1.56	4.36	8.27	13.2	6.77	55.78
UPP compensation	60	12.32	40.45	49.03	94.04	110.39	90.75	463.11
MOS compensation	60	1.75	7.6	10.43	20.91	25.61	22.42	100.66
Supplies expenses	60	1.91	6.88	10.87	23.29	35.73	21.04	142.88
Drug expenses	60	0.36	2.01	3.04	8.83	17.38	4.74	74.05
Contract out expenses	60	0.16	0.5	1	3.17	4.63	2.67	17.77
Sundry expenses	60	0.32	1.51	2.49	9.41	19.58	7.4	84.3
Capital	60	64	89.75	121	231.82	264.9	177.5	975
Admitted RIW	60	1492.39	3744.1	5451.31	11348.11	14529.94	10733.98	57386.72
Outpatients	60	247.53	757.22	1158.32	1582.4	1508.19	1719.67	6391.33

Discussion of Summary Statistics

The summary statistics provide a detailed view of hospital resource use and outputs in Alberta, Ontario, and Nova Scotia, highlighting differences across these provinces. Each province exhibits unique patterns in how hospitals allocate their resources, reflecting variations in scale, complexity, and operations.

In Alberta, hospitals show significant differences in spending. The average medical compensation is $12.91 million, but it can vary greatly due to a high standard deviation of $34.36 million. Supplies expenses average $13.80 million, with some hospitals spending as little as $0.31 million. Drug and contract expenses are generally low on average, but they also vary widely. This shows how differently hospitals manage outside services and purchase supplies. Alberta hospitals handle an average of 8,619 admitted cases, indicating their activity levels, while they see an average of 4,003 outpatient visits, showing varied patient care.

Ontario has the largest number of hospitals (587) and a wider range of spending and activity. Medical compensation is much higher, at an average of $19.04 million, with some reaching up to $249.30 million. This is likely due to larger hospitals needing more services. Supplies expenses average $24.70 million, which is almost double that of Alberta, highlighting the larger scale of operations. Ontario also has more outpatient visits and admitted cases, reflecting its role in serving a bigger and more diverse population. This increased workload leads to higher averages for drug and contract expenses, showing the complexity of Ontario's healthcare system.

In Nova Scotia, there are only 60 hospitals, so operations are smaller, but there are still important trends. The average medical compensation is lower at $8.27 million, but unit-producing personnel compensation averages $94.04 million, indicating a different staffing approach. Supplies and drug expenses are lower, averaging $23.29 million and $8.83 million, respectively. Notably, Nova Scotia has a higher average number of admitted cases than Alberta, suggesting a focus on inpatient care. However, outpatient visits average only 1,582, which is much lower than in Alberta or Ontario, indicating different service delivery methods.

Chapter 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 (LOO-CV) 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 unpenalised 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 unpenalised 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 generalisability 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 the 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 5, 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 5
Posterior means, standard deviations and 95% credible intervals of the translog HIM-IF & HIM Model

Variable	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]
$\boldsymbol{\beta}_{\log(MED\_EXP)}$	0.023 (0.013) [0.002, 0.045]	0.023 (0.013) [0.001, 0.045]
$\boldsymbol{\beta}_{\log(W\_MOSUPP\_EXP)}$	0.360 (0.031) [0.310, 0.411]	0.360 (0.031) [0.309, 0.412]
$\boldsymbol{\beta}_{\log(W\_SUPDRG\_EXP)}$	0.075 (0.015) [0.050, 0.099]	0.075 (0.015) [0.050, 0.099]
$\boldsymbol{\beta}_{\log(BEDS)}$	0.529 (0.023) [0.4941, 0567]	0.529 (0.024) [0.491, 0.567]
$\boldsymbol{\beta}_{\log(ORU)}$	-0.253 (0.024) [-0.291, -0.214]	-0.253 (0.024) [-0.292, -0.214]
$\boldsymbol{\beta}_{\log(ARU)}$	-0.452 (0.026) [-0.492, -0.408]	-0.452 (0.025) [-0.491, -0.410]
$\boldsymbol{\beta}_{\mathbf{trend}}$	0.001 (0.001) [-0.001, 0.003]	0.001 (0.001) [-0.001, 0.003]
$\boldsymbol{\beta}_{\log(MED\_EXP)^2}$	0.007 (0.004) [0.001, 0.014]	0.007 (0.004) [0.001, 0.013]
$\boldsymbol{\beta}_{\log(W\_MOSUPP\_EXP)^2}$	0.060 (0.044) [-0.011, 0.135]	0.060 (0.045) [-0.013, 0.134]
$\boldsymbol{\beta}_{\log(W\_SUPDRG\_EXP)^2}$	-0.033 (0.011) [-0.051, -0.015]	-0.034 (0.011) [-0.052, -0.015]
$\boldsymbol{\beta}_{\log(BEDS)^2}$	-0.079, (0.025) [-0.121, -0.037]	-0.079 (0.026) [-0.122, -0.037]
$\boldsymbol{\beta}_{\log(ORU)^2}$	-0.080 (0.014) [-0.104, -0.057]	-0.080 (0.014) [-0.104, -0.057]
$\boldsymbol{\beta}_{\log(ARU)^2}$	-0.156 (0.014) [-0.178, -0.133]	-0.155 (0.014) [-0.178, -0.133]
$\boldsymbol{\beta}_{(\mathbf{trend})^2}$	-0.001 (0.001) [-0.003, 0.000]	-0.001 (0.001) [-0.003, 0.000]
$\boldsymbol{\beta}_{\log(MED\_EXP \times W\_MOSUPP\_EXP)}$	-0.068 (0.026) [-0.110, -0.025]	-0.068 (0.025) [-0.109, -0.026]
$\boldsymbol{\beta}_{\log(MED\_EXP \times W\_SUPDRG\_EXP)}$	0.040 (0.011) [0.022, 0.058]	0.041 (0.011) [0.023, 0.058]
$\boldsymbol{\beta}_{\log(MED\_EXP \times BEDS)}$	0.016 (0.017) [-0.012, 0.043]	0.016 (0.017) [-0.012, 0.044]
$\boldsymbol{\beta}_{\log(MED\_EXP \times ORU)}$	-0.004 (0.013) [-0.026, 0.017]	-0.004 (0.013) [-0.026, 0.017]
$\boldsymbol{\beta}_{\log(MED\_EXP \times ARU)}$	-0.028 (0.011) [-0.046, -0.010]	-0.028 (0.011) [0.023, 0.058]
$\boldsymbol{\beta}_{\log(MED\_EXP \times trend)}$	0.005 (0.002) [0.002, 0.008]	0.005 (0.002) [0.002, 0.008]
$\boldsymbol{\beta}_{\log(W\_MOSUPP\_EXP \times W\_SUPDRG\_EXP)}$	-0.081 (0.035) [-0.139, -0.023]	-0.080 (0.036) [-0.139, -0.021]
$\boldsymbol{\beta}_{\log(W\_MOSUPP\_EXP \times BEDS)}$	0.031 (0.061) [-0.069, 0.131]	0.032 (0.062) [-0.070, 0.135]
$\boldsymbol{\beta}_{\log(W\_MOSUPP\_EXP \times ORU)}$	0.111 (0.039) [0.047, 0.176]	0.111 (0.040) [0.046, 0.177]
$\boldsymbol{\beta}_{\log(W\_MOSUPP\_EXP \times ARU)}$	-0.039 (0.035) [-0.097, 0.019]	-0.039 (0.037) [-0.099, 0.021]
$\boldsymbol{\beta}_{\log(W\_MOSUPP\_EXP \times trend)}$	-0.006 (0.004) [-0.013, 0.001]	-0.006 (0.005) [-0.014, 0.001]
$\boldsymbol{\beta}_{\log(W\_SUPDRG\_EXP \times BEDS)}$	0.105 (0.022) [0.068, 0.142]	0.105 (0.022) [0.068, 0.141]
$\boldsymbol{\beta}_{\log(W\_SUPDRG\_EXP \times ORU)}$	0.018 (0.020) [-0.015, 0.050]	0.018 (0.020) [-0.015, 0.050]
$\boldsymbol{\beta}_{\log(W\_SUPDRG\_EXP \times ARU)}$	0.017 (0.018) [-0.013, 0.047]	0.017 (0.018) [-0.013, 0.047]
$\boldsymbol{\beta}_{\log(W\_SUPDRG\_EXP \times trend)}$	0.002 (0.003) [-0.002, 0.007]	0.002 (0.003) [-0.002, 0.007]
$\boldsymbol{\beta}_{\log(BEDS \times ORU)}$	-0.121 (0.030) [-0.172, -0.072]	-0.121 (0.031) [-0.172, -0.071]
$\boldsymbol{\beta}_{\log(BEDS \times ARU)}$	0.051 (0.029) [0.004, 0.099]	0.051 (0.030) [0.003, 0.100]
$\boldsymbol{\beta}_{\log(BEDS \times trend)}$	-0.004 (0.003) [-0.009, 0.001]	-0.004 (0.003) [-0.009, 0.002]
$\boldsymbol{\beta}_{\log(ORU \times ARU)}$	0.162 (0.025) [0.121, 0.202]	0.162 (0.024) [0.122, 0.202]
$\boldsymbol{\beta}_{\log(ORU \times trend)}$	-0.005 (0.003) [-0.010, -0.001]	-0.005 (0.003) [-0.010, -0.001]
$\boldsymbol{\beta}_{\log(ARU \times trend)}$	0.004 (0.002) [0.000, 0.008]	0.004 (0.002) [0.000, 0.008]

$\boldsymbol{\mu_{u_l}}$	-2.931 (0.483) [-3.726, -2.135]	-3.064 (0.469) [-3.837, -2.284]
$\boldsymbol{\sigma_v}$	0.045 (0.002) [0.043, 0.048]	0.045 (0.002) [0.043, 0.048]
$\mathbb{E}\left[\boldsymbol{\sigma_{u_{j[k]}}}\right]$	0.575 (0.420) [0.060, 1.382]	0.568 (0.414) [0.060, 1.363]
$\mathbb{E}\left[\boldsymbol{\sigma_{u_{k[l]}}}\right]$	0.432 (0.324) [0.036, 1.050]	0.419 (0.317) [0.033, 1.033]
$\boldsymbol{\sigma_{u_l}}$	0.315 (0.238) [0.025, 0.779]	0.304 (0.236) [0.022, 0.758]
$\boldsymbol{\sigma_{\alpha_l}}$	0.080 (0.060) [0.007, 0.196]	0.080 (0.060) [0.006, 0.194]
$\boldsymbol{\sigma_{\alpha_k}}$	0.060 (0.046) [0.006, 0.150]	0.059 (0.047) [0.005, 0.150]
$\boldsymbol{\sigma_{\alpha_j}}$	0.335 (0.032) [0.291, 0.394]	0.329 (0.031) [0.286, 0.383]
Province Influence Factors		
$\boldsymbol{\phi_{AB}}$	0.608 (0.250) [0.167, 0.966]	-
$\boldsymbol{\phi_{NS}}$	0.498 (0.292) [0.045, 0.952]	-
$\boldsymbol{\phi_{ON}}$	0.499 (0.289) [0.047, 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 ($\boldsymbol{\alpha}$= 0.300) and HIM ($\boldsymbol{\alpha}$= 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 ($\beta_{\log(ARU)}$=-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) indicate that a 1% increase in administrative and operational staff use is associated with a 0.36% rise in resource requirements. While administration supports hospital functioning, over-expansion risks excessive bureaucracy that diverts resources away from patient care, further distancing hospitals from the efficient frontier. Similarly, supplies and drug expenditures ($\beta_{\log(W\_SUPDRG\_EXP)}$= 0.075) also leads 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 elasticity, indicating that a 1% increase in staffed and operational beds is associated with roughly a 0.53% rise in required inputs. This reflects the strong influence of physical capacity on hospital performance, but also shows that simply adding more beds does not solve inefficiency. If expansion exceeds patient demand, resources risk being underutilised, so bed growth should be approached as a precautionary measure and carefully aligned with actual service needs.

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, 0.002 [0.002, 0.008]) indicates that the input requirements for clinical staff increased systematically over time, suggesting a form of technological regress in labour productivity. This may reflect rising administrative burdens, increasing case complexity, or diminishing marginal returns to workforce expansion, whereby each additional unit of medical labour 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, 0.003 [−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 specialisation in high-volume ambulatory procedures. By contrast, the marginally positive interaction between inpatient activity and time ($\beta_{\log(ARU \times trend)}$) = 0.004, 0.002 [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 6 to 8 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 6
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

Note: Values are derived from the Hierarchical Inefficiency Model with Influence Factors (HIM-IF). Reported percentages represent the decomposition of persistent inefficiency into provincial, regional, and hospital components within each province.

As presented in Table 6, 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 to function, they are already burdened with inefficiency from the layers above. 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%. 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), is 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 7

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							

Note: Values are derived from the Hierarchical Inefficiency Model with Influence Factors (HIM-IF). Reported percentages represent the decomposition of persistent inefficiency into provincial, regional, and hospital components within each province.

Table 7 above presents the hierarchical decomposition of inefficiency in Nova Scotia's health system, which contrasts with 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. However, this is only part of the explanation. Nova Scotia's population density, 17.4 people per km^2 (Statistics Canada, 2021b), is more than triple that of Alberta, which stands at just 5.7 people per km^2 (Statistics Canada, 2021a). 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.

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 8
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[ǂ] (Base + Inherited)**	**Total Inefficiency[ǂ] (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				

Note: Values are derived from the Hierarchical Inefficiency Model with Influence Factors (HIM-IF). Reported percentages represent the decomposition of persistent inefficiency into provincial, regional, and hospital components within each province.

Table 8 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 km²), 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. However, 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%. This puts the province between Nova Scotia (85.37%) and Alberta (83.55%), reinforcing the view that Ontario's system, while more expansive and institutionally complex, does not necessarily translate that complexity into greater inefficiency. However, its larger scale may help distribute inefficiency across more institutions, softening the impact at any single node, even if variation still exists.

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 Data Envelopment Analysis (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 exhibited greater variation in efficiency, 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.

Across the three provinces, Alberta records the highest total inefficiency (16.45%), followed by Ontario (14.88%) and Nova Scotia (14.63%). 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 provides what is, to my knowledge, the first detailed empirical decomposition of inefficiency across multiple tiers of a national healthcare system, not just in Canada but in the broader healthcare efficiency literature. While numerous studies have assessed hospital-level or system-wide inefficiency using frontier methods such as DEA or SFA, none have mapped how inefficiency is transferred, inherited, and compounded across administrative layers, from province to region to hospital, with this level of granularity. This framework offers a new lens for understanding where inefficiency originates, how it travels, and where it is absorbed or amplified. The results also identify specific regions with higher cumulative inefficiencies, regions that, if named, could be targeted for deeper evaluation or corrective action. Unfortunately, despite multiple formal requests, the Canadian Institute for Health Information (CIHI), which provided the data, declined to release the names of the regions or hospitals involved. While this may reflect institutional or political sensitivities, it also limits the practical applicability of the findings. As this paper moves toward publication, I hope that CIHI and other stakeholders might engage with this analysis, either by revisiting access policies or by using this framework internally to identify and address the structural inefficiencies the model has revealed.

Chapter 6: Policy Implications

The empirical results from the HIM-IF model offer important and actionable insights for policymakers aiming to improve the efficiency of Canada's healthcare system. One of the clearest signals from the analysis comes from the estimated elasticities of input use. For example, administrative expenditures show one of the strongest effects on total input requirements (elasticity=0.360), meaning that even small increases in non-clinical staffing and operations can significantly raise hospital resource demands. This finding reinforces the need to assess whether administrative functions are proportionate and delivering value, especially as system complexity grows. Similarly, the elasticity for hospital beds (0.529) suggests that expanding physical capacity is resource-intensive and must be carefully justified. While additional beds may signal responsiveness to public concerns or political priorities, they can also lead to disproportionate growth in input use unless tightly aligned with service demand.

More positively, the model shows that Canadian hospitals operate under increasing returns to scale (RTS $\approx$ 1.42), meaning they become more efficient as they grow. This opens up a clear policy opportunity: encouraging regional integration, shared services, or hospital collaboration could allow systems to capitalise on size-related efficiency. However, this must be done with attention to the local context to avoid overstretching resources or reducing access. Importantly, the interaction between administrative and medical expenditures is negative, suggesting that when these two areas work in tandem rather than in silos, total resource use declines. This highlights the value of better team coordination and integrated workflow design as tools to curb inefficiency from within.

The findings of this study lay the foundation for a broader rethinking of how inefficiency arises in Canada's healthcare system. For the first time, this study decomposes inefficiency across provincial, regional, and hospital levels, showing that waste is not just a hospital-level problem but a cascading one. Inefficiency begins at the top and flows downward through institutional layers, eventually surfacing at the hospital level as operational strain. This layered view reframes inefficiency as a systems issue, and it calls for a shift in policy thinking: solutions must begin upstream if we expect results downstream.

One of the clearest findings is that provincial-level governance is the largest single contributor to overall inefficiency. In all three provinces studied, Alberta, Nova Scotia, and Ontario, the provincial layer introduces more waste than any other tier. If reforms aim to deliver real gains, they must start here. Improvements in budget timing, regulatory streamlining, and strategic oversight will ripple down the chain and help hospitals perform more efficiently. Too often, hospitals are judged for inefficiencies that are, in reality, inherited from higher levels of government. Without addressing these upstream burdens, hospital-level improvements will hit a ceiling.

At the regional level, most inefficiencies are moderate but not insignificant. Some regions, such as Ontario's regions 2 and 13, emerge as particularly high-burden zones, where inherited inefficiency meets local operational issues to create compounding effects. Yet, these regions cannot be named due to data privacy restrictions. If the Canadian Institute for Health Information (CIHI) or provincial governments were to revisit data-sharing protocols, even under secure conditions, it would greatly improve the ability of health leaders to intervene where the need is most acute. In the meantime, health authorities can still apply this model internally to identify their own pressure points and plan accordingly.

Importantly, the study also challenges the assumption that bigger systems are necessarily less efficient. Ontario, despite its size and complexity, performs relatively well because it spreads risk and redundancy across its network. That said, geography still matters; Alberta and parts of northern Ontario face real logistical barriers to care, including remote areas, long travel times, and reliance on air transport. In these cases, targeted policy responses are required, including flexible funding models, investments in rural infrastructure, and better regional coordination to absorb shocks without passing them directly to hospitals.

This model provides a practical roadmap for reform. Distinguishing between inherited and self-generated inefficiency enables targeted, tier-specific solutions. Provincial reforms should tackle systemic waste at the top; regional authorities should be equipped to manage what they inherit, and hospitals should be supported in addressing their internal processes. This tiered approach mirrors how the healthcare system actually functions, as a network of interdependent actors, rather than treating each unit in isolation.

This study also speaks to the importance of research–policy partnerships. The Canadian Institute for Health Information (CIHI)'s decision not to release regional or hospital identifiers is understandable from a privacy standpoint, but it limits how these insights can be turned into action. We urge CIHI and other data custodians to consider safe and controlled pathways for data sharing that would allow for more targeted reforms. With the right safeguards, better data access can help all stakeholders, from hospital administrators to health ministers, work toward a common goal: a more efficient, fair, and responsive healthcare system for all Canadians.

Chapter 7: Concluding Remarks

This study reconceptualises healthcare inefficiency as a hierarchical and cascading phenomenon rather than a siloed, hospital-level issue. By decomposing inefficiency across provincial, regional, and hospital levels, it provides the first empirical framework in the healthcare efficiency literature to trace how inefficiency flows downward through institutional layers. Using Canadian data, the results demonstrate that inefficiency often originates at the top of the system, in provincial governance, and accumulates as it moves through regions to the point of care. Ironically, the very administrative structures designed to ensure oversight and efficiency may be contributing to the problem by transmitting inefficiency downstream.

The empirical estimates from the HIM-IF model reinforce this systems-level interpretation. Input elasticities associated with administrative expenditures (0.360) and bed capacity (0.529) reflect how total input requirements scale with changes in these resource categories, conditional on output. In the context of the input distance function, these values indicate that administrative overhead and physical infrastructure are substantial contributors to input use. While not measures of inefficiency per se, they highlight areas where resource intensity is high and, therefore, where careful planning and demand alignment are essential. Similarly, the estimated returns to scale (1.42) indicate that hospitals operate under increasing returns to scale, meaning that as inputs expand, output tends to grow more than proportionally. This implies that larger or better-integrated hospitals may exhibit more productive use of inputs, particularly when coordination and scale are managed effectively. The interaction effects, such as the negative coefficient between medical and administrative expenditures, suggest that coordination across clinical and support functions may reduce overall input needs, offering operational insights for staffing and workflow strategies.

From a policy perspective, these findings provide both a diagnostic and a strategic direction. If inefficiency is primarily inherited from higher levels, then structural reform must begin upstream. Provincial-level delays in funding, fragmented policy frameworks, and rigid regulations are not background noise; they are central drivers of inefficiency throughout the system. Regional governance, meanwhile, plays a critical role in either absorbing or exacerbating inherited inefficiencies, with certain regions in Ontario, for example, emerging as focal points of cumulative waste. These patterns argue for more tailored and transparent accountability mechanisms that reflect the layered architecture of healthcare delivery.

Looking forward, the hierarchical nature of the data offers a valuable foundation for future research. One natural extension would be to estimate province-specific or region-specific input elasticities, capturing local variations in production technology and resource use. Hierarchical Bayesian modelling provides the tools to share information across levels while retaining flexibility, enabling more precise and context-specific policy recommendations. Such an approach would allow jurisdictions to benchmark themselves not just nationally but within the structural realities of their own administrative context.

Beyond its empirical contributions, this study proposes a shift in how health policymakers conceptualise inefficiency. It argues that inefficiency should not be treated as a static performance shortfall at the hospital level but rather as a dynamic systems issue that can stem from upstream governance decisions. Efficiency interventions, therefore, must be as multilayered and interconnected as the health system itself. Without acknowledging the cascading structure of inefficiency, downstream reforms will remain constrained by upstream inertia. By viewing inefficiency as something that flows through the system rather than simply appearing at the point of care, this framework offers policymakers a more grounded perspective on healthcare performance. It gently invites broader reflection on whether traditional efficiency assessments in healthcare might benefit from a more layered, system-aware approach.

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

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

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. *Am J Respir Crit Care Med, 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. (2023). *National Health Expenditure Trends: 2023 Data Tables*. https://www.cihi.ca/en/national-health-expenditure-trends

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

Canadian Medical Association (CMA). (2023). *Admin Burden Working Group: Improving Physician Wellness by Reducing Physician Admin Burden*. O. T. Association. https://policybase.cma.ca/link/policy14530

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.

Dutton, D. J., Forest, P.-G., Kneebone, R. D., & Zwicker, J. D. (2018). Effect of provincial spending on social services and health care on health outcomes in Canada: an observational longitudinal study. *Cmaj, 190*(3), E66-E71.

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

Fixler, T., Paradi, J. C., & Yang, X. (2014). A data envelopment analysis approach for measuring the efficiency of Canadian acute care hospitals. *Health Services Management Research, 27*(3-4), 57-69.

Flood, C. M., Thomas, B., & McGibbon, E. (2023). Canada's primary care crisis: Federal government response. *Healthc Manage Forum, 36*(5), 327-332. https://doi.org/10.1177/08404704231183863

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., 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

Health Canada. (2023b). *Bilateral Agreement with Alberta to Improve Care*. https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/shared-health-priorities/working-together-bilateral-agreements/alberta-improve-care.html

Health Canada. (2025). *Health Workforce*. https://www.canada.ca/en/health-canada/services/health-care-system/health-human-resources.html

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

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). 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.

Martin, D., Miller, A. P., Quesnel-Vallée, A., Caron, N. R., Vissandjée, B., & Marchildon, G. P. (2018). Canada's universal health-care system: achieving its potential. *The Lancet, 391*(10131), 1718-1735.

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

Moir, M., & Barua, B. (2024a). *Comparing performance of universal health care countries*. Fraser Institute. https://www.fraserinstitute.org/studies/comparing-performance-of-universal-health-care-countries-2024

Moir, M., & Barua, B. (2024b). *Waiting your turn: Wait Times for Health Care in Canada,* . Fraser Institute. https://www.fraserinstitute.org/studies/waiting-your-turn-wait-times-for-health-care-in-canada-2024

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.

Naylor, C. D. (2018). Canada as single-payer exemplar for universal health care in the United States: a borderline option. *Jama, 319*(1), 17-18.

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

Office of the Auditor General of Canada. (2015). *Report 4 - Access to health Services for Remote First Nations Communities*. https://www.oag-bvg.gc.ca/internet/English/parl_oag_201504_04_e_40350.html

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

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

Organisation for Economic Co-operation and Development. (2023). *Health at a Glance 2023: OECD Indicators*. https://doi.org/10.1787/7a7afb35-en.

Rechel, B., Doyle, Y., Grundy, E., & McKee, M. (2009). How can health systems respond to population ageing?

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.

Schneider, E. C., Shah, A., Doty, M. M., Tikkanen, R., Fields, K., Williams, R., & II, M. M. (2021). Reflecting poorly: health care in the US compared to other high-income countries. *New York: The Commonwealth Fund, 4*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://search.issuelab.org/resources/38740/38740.pdf

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*. https://www.cfhi-fcass.ca/docs/default-source/health-system-performance/ALC_EN_Final.pdf

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.

Wang, L., Grignon, M., Perry, S., Chen, X. K., Ytsma, A., Allin, S., & Gapanenko, K. (2018, Dec). The Determinants of the Technical Efficiency of Acute Inpatient Care in Canada. *Health Serv Res, 53*(6), 4829-4847. https://doi.org/10.1111/1475-6773.12861

World Health Organisation, Organisation 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.