OF MEN AND MACHINES: THE CHALLENGE OF INCREASING DIAGNOSTIC PRECISION IN THE TURKISH HEALTH SYSTEM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEHTAP TUĞBA KAYNAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF ECONOMICS

SEPTEMBER 2025

Approval of the thesis:

OF MEN AND MACHINES: THE CHALLENGE OF INCREASING DIAGNOSTIC PRECISION IN THE TURKISH HEALTH SYSTEM

submitted by **MEHTAP TUĞBA KAYNAK** in partial fulfillment of the requirements for the degree of **Master of Science in Economics, the Graduate School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Sadettin KİRAZCI
Dean
Graduate School of Social Sciences ____________________

Prof. Dr. Elif AKBOSTANCI ÖZKAZANÇ
Head of Department
Department of Economics ____________________

Assoc. Prof. Dr. İlhan Can ÖZEN
Supervisor
Department of Economics ____________________

Examining Committee Members:

Prof. Dr. Selcen ÖZTÜRK (Head of the Examining Committee)
Hacettepe University
Department of Economics ____________________

Assoc. Prof. Dr. İlhan Can ÖZEN (Supervisor)
Middle East Technical University
Department of Economics ____________________

Assoc. Prof. Dr. Münis Seven AĞIR YAĞCI
Middle East Technical University
Department of Economics ____________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Mehtap Tuğba KAYNAK

Signature:

ABSTRACT

OF MEN AND MACHINES: THE CHALLENGE OF INCREASING DIAGNOSTIC PRECISION IN THE TURKISH HEALTH SYSTEM

KAYNAK, Mehtap Tuğba
M.S., The Department of Economics
Supervisor: Assoc. Prof. Dr. İlhan Can ÖZEN

September 2025, 151 pages

This thesis analyses underdiagnosis as a systemic cost and burden to the Turkish healthcare system, analysing it through three dimensions of diagnostic capacity strain: human, institutional, and technological. Building on Arrow's (1963) economics of uncertainty, underdiagnosis is examined as a systemic outcome of asymmetric information, incentive misalignments, and constrained capacity, reflecting the volumetric capacity increases and health system reforms experienced by the Health Transformation Program increased access, but also intensified pressures on physicians and institutions. Thus, while imaging volumes increased, the marginal informational yield of additional scans was limited by overworked doctors, uneven institutional structures, and variable machine quality. This paradox helps explain why overutilization and underdiagnosis coexist as consequences of systemic imbalance. Then, the empirical analysis combines nationwide patient- and hospital-level datasets from 2015–2016 and 2019–2023. Logistic regression and instrumental variable (IV) models estimate how physician workload, hospital tier, repeat MRI behaviour, and machine quality affect the probability of underdiagnosis, measured via ICD-10 R-coded diagnoses. Results indicate that higher daily patient loads, peripheral hospital environments, and low MRI quality significantly increase diagnostic uncertainty. IV estimates further confirm that both physician workload and repeat MRI rates causally elevate underdiagnosis probabilities. The findings demonstrate that diagnostic precision is constrained simultaneously by human effort, institutional structure, and technological reliability. Türkiye’s experience illustrates a broader lesson: expanding access without embedding quality safeguards risks entrenching inefficiency and underdiagnosis. Policy implications include recalibrating performance incentives, ensuring adequate staffing, monitoring repeat imaging as a quality metric, and ensuring that access gains translate into real improvements in patient outcomes.

Keywords: Underdiagnosis, Diagnostic Capacity, Health Economics, Magnetic Resonance Imaging (MRI)

ÖZ

İNSAN VE MAKİNELER: TÜRKİYE SAĞLIK SİSTEMİNDE TANISAL DOĞRULUĞU ARTIRMANIN ZORLUKLARI

KAYNAK, Mehtap Tuğba
Yüksek Lisans, İktisat Bölümü
Tez Yöneticisi: Doç. Dr. İlhan Can ÖZEN

Eylül 2025, 151 sayfa

Bu tez, yetersiz tanıyı Türk sağlık sistemine sistemik bir maliyet ve yük olarak inceleyerek, tanı kapasitesi zorlanmasının üç boyutu olan insan, kurumsal ve teknolojik boyutları analiz etmektedir. Arrow'un (1963) belirsizlik ekonomisine dayanarak, yetersiz tanı, asimetrik bilgi, teşvik uyumsuzlukları ve kısıtlı kapasitenin sistemik bir sonucu olarak incelenmekte ve Sağlık Dönüşüm Programı'nın deneyimlediği hacimsel kapasite artışlarını ve sağlık sistemi reformlarını (erişimi artırmakla birlikte hekimler ve kurumlar üzerindeki baskıları da yoğunlaştırmıştır) yansıtmaktadır. Dolayısıyla, görüntüleme hacimleri artarken, ek taramaların marjinal bilgi verimi, aşırı çalışan doktorlar, eşitsiz kurumsal yapılar ve değişken makine kalitesi nedeniyle sınırlı kalmıştır. Bu paradoks, aşırı kullanım ve yetersiz tanının neden sistemik dengesizliğin sonuçları olarak bir arada var olduğunu açıklamaya yardımcı olmaktadır. Ardından, ampirik analiz 2015-2016 ve 2019-2023 yıllarına ait ülke çapındaki hasta ve hastane düzeyindeki veri setlerini birleştirmektedir. Lojistik regresyon ve enstrümantal değişken (IV) modelleri, hekim iş yükü, hastane kademesi, tekrarlayan MRI davranışı ve makine kalitesinin, ICD-10 R kodlu tanılar aracılığıyla ölçülen yetersiz tanı olasılığını nasıl etkilediğini tahmin etmektedir.

Sonuçlar, daha yüksek günlük hasta yüklerinin, çevresel hastane ortamlarının ve düşük MRI kalitesinin tanı belirsizliğini önemli ölçüde artırdığını göstermektedir. IV tahminleri ayrıca hem hekim iş yükünün hem de tekrarlayan MRI oranlarının yetersiz tanı olasılıklarını nedensel olarak artırdığını doğrulamaktadır. Bulgular, tanısal kesinliğin aynı anda insan çabası, kurumsal yapı ve teknolojik güvenilirlik tarafından kısıtlandığını göstermektedir. Türkiye'nin deneyimi daha geniş bir dersi ortaya koymaktadır: Kalite güvencelerini yerleştirmeden erişimi genişletmek, verimsizliği ve yetersiz tanıyı daha da derinleştirme riskini taşımaktadır. Politika uygulamaları arasında performans teşviklerinin yeniden kalibre edilmesi, yeterli personel sağlanması, tekrarlayan görüntülemenin bir kalite ölçütü olarak izlenmesi ve erişim kazanımlarının hasta sonuçlarında gerçek iyileştirmelere dönüşmesinin sağlanması yer almaktadır.

Anahtar Kelimeler: Tanı Eksikliği, Tanı Kapasitesi, Sağlık Ekonomisi, Manyetik Rezonans Görüntüleme (MRG)

To my future ideals, ever worthwhile

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Assoc. Prof. Dr. İlhan Can Özen, whose profound knowledge in the field, invaluable guidance, and critical insights have been essential in shaping this thesis. His constructive comments, rigorous evaluations, and intellectual support were crucial in refining the research, ensuring its academic depth, and guiding it to completion.

I am also sincerely grateful to Prof. Dr. Selcen Öztürk, for her careful evaluation and valuable feedback. As a scholar whose related studies I have benefited from in my research, receiving her perspectives was especially meaningful. Her comments not only strengthened the quality of this study but also provided a meaningful academic connection to the sources that informed my work.

My heartfelt thanks also go to Assoc. Prof. Dr. Münis Seven Ağır for her thoughtful comments and valuable contributions to the improvement of this thesis. Having served as my minor advisor during my undergraduate years, she was among the first professors to whom I turned with questions about pursuing an academic path in Economics. Her presence on the jury, therefore, carries a special personal meaning for me, symbolizing both continuity and progression in my academic journey. Her evaluation and support not only strengthened this thesis but also made the process of completing it feel more meaningful.

In addition, I would like to acknowledge Ender Yıldırım and the METU MEMS Research Center for their generous technical support within the scope of the TÜBİTAK 1004-supported MAESTRO Project (Project No: 22AG008, *Micro Medical Technologies Platform*). As a research assistant in this project, I benefited greatly from the remote computing resources made available, which were invaluable for carrying out much of the data processing and analysis required for this thesis.

I am especially grateful to my mother, who has always been by my side throughout my academic life with her endless love, and support. Her unwavering belief in me gave me strength during the most challenging moments of this journey. I would also like to thank my dear friends Alp Eren Sarı, Kristina Jakimovska, Gamze Sarı, and Tarık Karaman, whose constant motivation and belief in me carried me through the difficulties of this thesis process and made the journey more meaningful. Lastly, I would like to acknowledge my own perseverance and dedication throughout this challenging process, for not giving up even when it seemed overwhelming.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CMS	Centers for Medicare & Medicaid Services
CT	Computed Tomography
HTP	Health Transformation Program
MRI	Magnetic Resonance Imaging
OECD	Organisation for Economic Co-operation and Development
SUT	Sağlık Uygulama Tebliği (Health Implementation Guideline)
2SLS	Two-Stage Least Squares
GHI	General Health Insurance
IMF	International Monetary Fund
IV	Instrumental Variable
MAM	Mammography
MoH	Ministry of Health
MRI	Magnetic Resonance Imaging
PBSP	Performance-Based Supplementary Payment
PPPs	Public-Private Partnerships
OLS	Ordinary Least Squares
PET	Positron Emission Tomography
SSI	Social Security Institution
TRA	Turkish Radiology Association
US	United States
UK	United Kingdom
USG	Ultrasound
WB	World Bank
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

There is no doubt that health is one of the most important areas that has advanced the most in human life with technology. On the one hand, together with widespread use of advanced diagnostic imaging technologies as magnetic resonance imaging (MRI) and computed tomography (CT), healthcare systems worldwide have witnessed a significant development in diagnosis and treatment as making it easier and faster, and treatment options are becoming more widely available. On the other hand, these developments are also raising new concerns: Do additional diagnostic tools lead to improved diagnoses, or do they occasionally create potential risks or new bottlenecks as a result of their overutilization? All these concerns create the challenge of improving diagnostic accuracy. This can be described as the paradoxical dichotomy of diagnostic accuracy between human expertise and machine capacity, or quantity and quality, or expanding access and ensuring accuracy. From this point on, the thesis will examine how the situations that lead to this paradoxical dichotomy develop through the factors of the rapid expansion of the Turkish health system.

Over the past two decades, many countries have expanded their health system capacity through spending, infrastructure, technology, and workforce. On the human capacity, according to OECD, the number of doctors increased from 2.9 million in 2001 to 3.5 million in 2011 and 4.3 million in 2021. This increase grew faster than the population size, from 3.2 in 2011 to 3.7 in 2021. However, it highlights that the increase in the number of doctors was 2.5 or fewer doctors per 1,000 people in Türkiye among OECD countries (OECD, 2023, p. 176). In addition to human capacity, physical infrastructure has also risen. In 2022, the average number of hospital beds per 1,000 people in OECD countries was 4.3. The latest developments in technological capacity are reported in OECD as how technology is demonstrated in health systems as a key diagnostic tool and a significant cost driver. It is reported that the majority of OECD countries have seen a sharp growth in the availability and utilization of CT, PET, and MRI scanners; however, overutilization of these diagnostic procedures provides no benefit to patients (OECD, 2023, p. 116). Thus, rapid growth in MRI use has raised concerns about the potential for overutilization. For example, in the United States, approximately 20% of advanced imaging studies have been investigated as repetitive or of low clinical value (Hendee et al., 2010; Colla et al., 2017).

In Türkiye, as in many countries, diagnostic technologies have become the backbone of modern medicine. In this regard, Türkiye's healthcare system has experienced significant expansion in its health system capacity through its infrastructure, technology, and workforce in the recent two decades, consistent with the global expansion trend. In particular, the use of MRI in Türkiye has increased sharply over the last two decades. In 2015, Türkiye, Germany, the United States, Japan, and France had the highest rates of MRI examinations per capita. All of them had more than 100 MRI exams per 1,000 people among OECD countries, while Türkiye ranked first with 144 scans, while the OECD average was 57 scans. While this increase in utilization of MRI exams in the US was more than doubled between 2000 and 2015, it was tripled in Türkiye between 2008 and 2015. (OECD, 2017, p.170). These shed light on the quantitative surge in healthcare capacity and access in Türkiye and suggest potential overutilization.

It is reported that there is increasing evidence that MRI scans are systematically prescribed to patients with various health problems in Türkiye, resulting in overutilization of MRI scans (OECD, 2017, p.170). This increase in MRI utilization was also considered a multifaceted problem by the Turkish Radiology Association. They highlight the problem of a low number of radiologists per population, with 5 radiologists per 100,000 people, which is approximately one-half to one-third of that in most European countries. Thus, the combination of the low number of radiologists and the high number of examinations demonstrates the high workload per radiologist, which is required to report 200 or even 300 examinations per day. (Turkish Radiology Association, 2018, p. 2).

This heavy schedule forces radiologists to evaluate each trial in less than five minutes, and this is significantly insufficient for a standard radiologic examination and creates *‘examination intensity’*, which makes the service unsustainable. As a result, while radiologists work at risk, patients might face serious risks to diagnostic reliability (Turkish Radiology Association, 2018, p. 2). This illustrates the quantity-quality paradox in Türkiye: more is not always better in healthcare.

1.1. Research Problem

A critical consequence of overuse without adequate quality is the risk of underdiagnosis (Newman-Toker, 2014). Underdiagnosis occurs as a result of negligence or flawed actions during the diagnostic process and refers to situations where a diagnosis is missed, resulting in delayed or missed diagnoses that are unclear or nonspecific (p.44). Therefore, underdiagnosis has become a significant problem because missed diagnoses and delays negatively impact both the quality and efficiency of healthcare and patient outcomes. According to a US study, approximately 5% of adults (12 million) experience a diagnostic error each year, and nearly half of these errors have the potential to cause serious harm (Singh et al., 2014, p. 729). Thus, the thesis is motivated by this quality challenge: the concern in the Turkish health care system. Despite the widespread use of diagnostic technologies and the increasing number of tests, many patients still face a lack of definitive diagnosis for their conditions. This situation arises from the use of non-specific diagnosis codes in disease definition. In this context, ICD-10 R [1]codes, which specify "Symptoms, signsand abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)" are used for these undiagnosed conditions rather than definitive disease diagnoses.

A recent study suggests that approximately 10-12% of all diagnoses fell into R codes between 2016 and 2022 in Türkiye (Ülgü and Birinci, 2023). This indicates that one

[1] The International Classification of Diseases, 10th Revision (ICD-10) is the WHO’s alphanumeric system for coding diagnoses. Chapter XVIII (R00–R99) is titled *“Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified.”* (e.g., R07 chest/throat pain, R51 headache, R55 syncope) rather than a definitive disease diagnosis at that visit (CMS, 2024). In this thesis, “R-codes” denote encounters recorded with symptom- or sign-based ICD-10 codes and will be used as a proxy for underdiagnosis.

in ten patients is not diagnosed with any specific disease. However, in a healthcare system where the number of diagnostic technologies used for disease diagnosis is constantly increasing and is widely used, the presence of R codes at a certain rate reveals the problem of underdiagnosis, which raises the research question of this thesis. Thus, the prevalence of over 10% of R codes in Türkiye demonstrates the concerns regarding underdiagnosis. Each unresolved case not only burdens the patient but also burdens the healthcare system through repeated visits and potentially excessive resource use. In this context, another study shows that MRI scans are of insufficient quality to make a diagnosis, requiring patients to undergo repeat imaging or further investigations in Türkiye (Başar et al., 2019). As Arrow (1963) noted, medicine is explained as adjustments to uncertainty regarding the prevalence of illness and the effectiveness of treatment. Thus, there can be many tests and procedures that cannot resolve the diagnosis uncertainty. Considering all these, diagnostic precision is emerging as a new challenge for the health system in Türkiye. Therefore, the first motivating question for this research is: Does the intensive use of MRI in Türkiye actually lead to better diagnoses, or has overuse led to a prevalence of underdiagnosis?

The economic impact of diagnostic quality serves as another driving force behind this thesis. As an advanced imaging technology, the proliferation of MRIs is expensive and has a significant budgetary impact on the healthcare system. However, the low diagnostic yield from these scans represents a waste of resources and opportunity costs. Inadequate diagnosis can lead to costs such as repeated scans and increased specialist visits. Therefore, from a health economics perspective, diagnostic quality has a direct impact on system efficiency. Here, this thesis is guided by the following question: What are the economic costs of inadequate diagnosis in the Turkish healthcare system, and conversely, could improving diagnostic accuracy through higher-quality MRI lead to cost savings?

1.2. Literature Gaps

There are many studies in the literature that highlight the increasing use of medical imaging technologies (Smith-Bindman et al., 2008; Larson et al., 2011; Wang et al., 2021; Nesemeier et al., 2023). This increase is attributed to factors such as increased technology availability, increased patient and physician demand, and improvements in reimbursement systems. However, this situation also brings with it the problem of overutilization, which increases the costs of MRI and CT imaging utilization (Emery et al., 2013). Reasons for overutilization include patient expectations and physician concerns. Furthermore, many studies investigate the potential risks of overutilization of imaging technologies. Some studies investigate how overuse leads to quality issues in healthcare and leads to over-testing (Chassin and Galvin, 1998). Some studies also indicate that the overuse of medical imaging is a significant factor in increasing global healthcare expenditures (Armao et al., 2012; Bernardy et al., 2009; Brownlee et al., 2017; Chassin and Galvin, 1998; Emanuel and Fuchs, 2008; Hendee et al., 2010; Smith-Bindman et al., 2008). Other studies emphasize the importance of ensuring and maintaining the quality of care in medical imaging services (Gadeka and Esena, 2020).

Many studies have been published in the literature on Türkiye's health care reforms. Many of these focus on quantitative improvements and health outcomes and describe the Health Transformation Program's successes in expanding coverage, access, and service utilization. For example, Tatar et al. (2011) and Atun et al. (2013) analysed the HTP and emphasized achievements such as larger insurance coverage, increased hospitalizations, improved equity in access, and improved patient satisfaction.

On the other hand, there are various studies on the adequacy of medical imaging devices and diagnostic capacity in Türkiye. These studies are conducted to improve the efficiency and quality of the country's healthcare system. In this context, some studies focus on the distribution and use of medical imaging devices in hospitals, thereby providing information on the adequacy and accessibility of these devices (Semin et al., 2006; Sonğur and Top, 2016; Temiz and Kara, 2024). Some studies also evaluate the performance of medical imaging devices in Türkiye and their impact on early disease diagnosis in terms of quality and performance. Such studies emphasize the importance of effective device use and maintenance to improve the quality of healthcare services (Özen et al., 2022). On the other hand, there are few qualitative analyses concerning the overutilization of examinations (Yıldırım & Yıldırım, 2011).

The thesis contributes to the literature by examining diagnostic uncertainty and analysing the performance of diagnostic technologies, which have become increasingly widespread both in Türkiye and the world in the past two decades. In this context, unlike most studies in the literature, this study does not merely provide a quantitative analysis of the number of services provided by diagnostic technologies in healthcare. Instead, it analyses the effectiveness of these services, based on underdiagnosis, under the influence of diagnostic capacity factors within the Turkish healthcare system. Thus, it aims to examine the prevalence of underdiagnosis and the resulting systemic and economic costs despite the increased use of MRI technologies in the Turkish healthcare system.

1.2. Research Questions

Based on the above motivations, the thesis addresses several research questions related to the high rate of MRI utilization and underdiagnosis in Türkiye. To this end, the aim is to empirically evaluate the determinants of underdiagnosis in Türkiye during 2015-2016:

RQ1: How can 'diagnostic capacity' be defined and measured in the Turkish healthcare system, and how does this capacity vary by human, institutional, technological, and regional factors?

RQ2: How do supply-side factors in the healthcare system affect diagnostic accuracy and underdiagnosis rates under diagnostic capacity factors in the Turkish healthcare system?

RQ3: How are MRI utilization and MRI quality related to underdiagnosis? What are the impacts of MRI quality and MRI utilization levels on patient and system outcomes in the Turkish healthcare system?

RQ4: What are the consequences of underdiagnosis for the health system, particularly the economic costs and inefficiencies, and how are these outcomes linked to capacity shortages or strain?

The study intends to close a gap in the literature on high imaging utilization and diagnostic efficiency in a middle-income country setting by tackling these concerns. Given the swift changes in the healthcare system, the growing involvement of the private sector, and the particular difficulties in striking a balance between quality and accessibility, the Turkish instance is especially instructive. The ultimate objective is to obtain knowledge that can direct policy to maximize the use of MRIs by optimizing their benefits in accurately detecting diseases while minimizing resource loss and patient uncertainty brought on by underdiagnosis.

1.3. Research Contribution

This thesis will offer a thorough understanding of the connection between diagnostic quality and capacity, determining its causes as well as its consequences, focusing on the use of MRI in Türkiye. First, it will introduce a methodological contribution: the use of "R-code" as a quantitative indicator of underdiagnosis in hospital data. This is the first study in Türkiye to measure underdiagnosis by looking at the proportion of cases with ICD-10 R-coded diagnoses. Defining the underdiagnosis criterion allows us to compare diagnostic performance across hospitals and regions in a standardized manner. Second, the study contributes a causal analysis linking MRI utilization quality to patient diagnostic outcomes in Türkiye: lower MRI quality significantly increases underdiagnosis rates, suggesting affected by diagnostic capacity factors.

Following the introduction, the remainder of this thesis is organized as follows. Chapter 2 analyses the capacity expansion of healthcare services driven by the reforms under the Health Transformation Program. It develops a diagnostic capacity framework in three dimensions: technological, human, and institutional, and highlights regional disparities in healthcare delivery. Chapter 3 presents the theoretical framework and reviews the literature on underdiagnosis in diagnostic imaging. Based on Arrow's uncertainty theory, this chapter contextualizes underdiagnosis through concepts such as supplier-induced demand[2], defensive

[2] Supplier-induced demand (SID) refers to physicians influencing patients' demand beyond clinical necessity, largely due to financial or institutional incentives (Evans, 1974; McGuire, 2000). Classic studies show how provider behavior and payment systems affect utilization (Fuchs, 1978; Gruber & Owings, 1996; Clemens & Gottlieb, 2014), while recent work highlights SID in diagnostic imaging (Zabrodina et al., 2020; Dreger et al., 2022). A detailed discussion will follow in Chapter 3.

medicine[3], patient expectations, and capacity constraints. Chapter 4 describes the data sources and empirical methodology. It details patient- and hospital-level datasets covering the years 2015–2016 and 2019–2023, presents descriptive statistics and visualizations, and describes the econometric approach, including logistic regressions and instrumental variable (IV) models. Chapter 5 presents the econometric models and results. It discusses logistic regression models of underdiagnosis across human, institutional, and machine capacity dimensions, followed by IV estimates that address potential endogeneity in physician workload and MRI quality. Finally, Section 6 concludes by summarizing the key findings, situating them within broader theoretical and policy debates, and outlining implications for health system reforms and future research.

[3] Defensive medicine refers to medical practices primarily motivated by physicians' fear of malpractice liability rather than by clinical necessity. The concept first appeared in academic literature in the early 1970s (Hershey, 1972), gained prominence during the U.S. malpractice crises of the 1980s (Office of Technology Assessment, 1994), and was later formalized into two main types: *positive defensive medicine* (ordering unnecessary tests, referrals, or hospitalizations) and *negative defensive medicine* (avoiding high-risk patients or procedures) (Kessler & McClellan, 1996). Empirical studies such as Studdert et al. (2005) and Katz et al. (2005) demonstrated the widespread prevalence of defensive practices and their tangible effects on diagnostic decision-making. A more detailed discussion of defensive medicine will be provided in Chapter 3.

CHAPTER 2

DIAGNOSTIC CAPACITY IN TÜRKİYE: EVALUATION OF THE TURKISH HEALTHCARE TRANSFORMATION

This chapter reviews the institutional and operational context of diagnostic services in Türkiye's healthcare system, examining how systemic constraints contribute to underdiagnosis in imaging. Accordingly, the thesis will first outline the institutional structure shaped by the Health Transformation Program (HTP), financing reforms, and political factors. Then it will define a framework of diagnostic capacity under human workforce, technology, institutional arrangements, and time resources, and their application to Türkiye. Finally, it will examine the reasons behind the explosive rise of MRI-based radiography, emphasizing policy reactions, overuse, and repeat imaging. Throughout, we relate each factor to the thesis hypothesis that human, technological, and institutional strain in the system impairs diagnostic accuracy and leads to underdiagnosis.

2.1. Institutional Structure of Healthcare in Türkiye

This chapter analyses the reforms under the Health Transformation Program (HTP) and the changes and impacts they introduced to the Turkish healthcare system. In this context, it examines the HTP's objectives and achievements, along with criticisms of its implementation. The role of performance-based incentives is also discussed alongside the HTP program.

Thus, this section will cover the transformation of Türkiye's healthcare system, focusing on the structural factors behind the proliferation of diagnostic imaging, particularly MRI technologies.

2.1.1. Institutional Reforms and the Health Transformation Program

The Health Transformation Program (HTP) launched in 2003 was a comprehensive reform initiative to address the problems in the health sector, including: ii) lagging health outcomes compared to other OECD and middle-income countries; iii) inequitable access to health services; iv) fragmented structure in health service financing and delivery, leading to inefficiency and undermining financial sustainability; and iv) low service quality and limited responsiveness to patients. In this regard, the program aimed to make the health system more effective by improving governance, efficiency, user and provider satisfaction, and long-term financial sustainability (OECD, 2008, p.36).

In this context, HTP reforms can be analysed as institutional and organizational changes, financial changes, and service utilization reforms: Institutional and organizational changes include the restructuring of the Ministry of Health, the establishment of a General Health Insurance System (GHI) that combine the SSK, Bağ-Kur, "Emekli Sandığı" (Retirement Fund), and Green Card programs under the Social Security Institution (SSI) system, granting autonomy to public hospitals and establishing a strong preventive and primary health care system based on the family medicine model (OECD, 2008, p.36).

Financial reforms include increasing health spending, expanding health insurance coverage, introducing the performance-based payment system, and merging health insurance programs into a unified universal health insurance system. In this context, the Green Card system was eliminated, and the poor population was integrated into the universal health insurance system with the implementation of the Social Security and Universal Health Insurance Law in October 2008. Green Card holders officially joined the General Health Insurance Scheme (GHI) (Atun et al.2013). As a result, insurance coverage increased from 64% to 98% of the population. (Gürsoy, 2014; Okçan and Çakar, 2015). World Bank studies noted that these changes resulted in better financial protection for the poor and equity gains (Aran and Rokx, 2014). Service delivery reforms primarily focused on expanding primary healthcare services, hospital capacity, and services. In this context, first, HTP launched a primary health care model centred on family medicine. With this model, each family physician provided a range of services to a maximum of 4,000 registered citizens. In addition, Public-Private Partnerships (PPPs) were implemented in hospital construction to expand hospital capacity. This led to the development of "City Hospital" models, where private companies construct, equip, and maintain the facilities, while the Ministry of Health provides clinical services. Furthermore, a *performance-based supplementary payment* (PBSP) system was introduced to increase efficiency and reduce waiting times (Vujicic et al., 2009). This system rewarded physicians and hospitals based on service volume and efficiency scores for imaging, surgical, and outpatient consultations. Additionally, innovations were made in health information systems. For example, a nationwide electronic health record system, "e-Pulse", was launched (Atun et al.2013).

With all these reforms, the HTP addressed some inequalities by improving access and patient satisfaction metrics. Furthermore, the HTP fundamentally restructured the healthcare system through market mechanisms to increase efficiency and private sector participation. The program embraced public-private partnerships and competitive procurement processes, encouraged private hospital growth, and implemented performance-based financing.

2.1.2. Performance-Based Payments and the Rise of Pay-for-Service

Firstly, the performance-based payment system was implemented to encourage job motivation and productivity among public sector health personnel.in MoH hospitals in 2004 (Atun et al. 2013; Vujicic et al., 2009). This system introduced financial bonuses for physicians and hospital departments based on increased patient visits, procedures, and diagnostic test volume. Within this framework, MRI and CT scans became the focus of this incentive structure because they are high-cost and billable procedures. Because the reward system is directly linked to service volume, it created a favourable environment for supplier-induced demand (Okçan and Çakar, 2015; Sparkes, 2015). This has raised many concerns. The WB study notes that PBSP encourages unnecessary procedures and reduces the quality and safety of healthcare (Aran and Rokx, 2014, p.10). Some also argued that PBSP sometimes pressures physicians to increase efficiency at the expense of quality (Yıldırım & Yıldırım, 2011). In addition, Turkish Medical Association surveys have revealed that most physicians believe the PBSP mechanism leads to more unnecessary diagnoses and even an increased risk of malpractice (Turkish Medical Association Ethical Committee, 2009). Aktas et al. (2023), based on qualitative interviews with physicians, report that physicians respond to performance-based payment incentives by “selecting” high-volume cases and even “upcoding” patient diagnoses to meet targets.

Imaging services are reimbursed according to a national fee schedule known as the Health Practice Guideline (SUT). The SUT lists standard prices for all medical procedures. According to this practice, there is a fixed reimbursement rate for an MRI of a specific body part, as stated in the SUT. Public hospitals receive these reimbursements in their revolving funds. Private hospitals receive the same reimbursement rate from the Social Security Institution (SGK). In addition, private hospitals are allowed to charge patients an additional fee, which, by law, allows them to bill up to 200% of the SGK rate (SGK, 2011). This system generated more hospital revenue through higher imaging volumes, providing a financial incentive for both public and private healthcare providers to perform more imaging. This revenue was then used to purchase more MRI machines and radiology staff, further increasing capacity (Özen et al., 2022).

In addition, some imaging services in public hospitals are outsourced. Many public hospitals have contracted with private companies to provide on-site imaging services. Under these contracts, a private provider brings personnel to the public hospital and is paid per scan or through a service agreement by the hospital. This outsourcing model became widespread in the 2000s, particularly for MRI and CT scans, allowing the Ministry of Health to rapidly expand capacity without requiring the Ministry of Health to purchase each device directly. However, it has also created distortions in the payment model: private contractors are typically paid a flat fee per scan, incentivizing maximum efficiency. Many outsourced radiology units are reported to operate 24/7 and perform 120–140 MRI exams per day on a single device (Rylands-Monk, 2014). By operating scanners at this high volume, these subcontractors earn more, and the hospital can meet patient demand and collect social security payments accordingly. However, this high-volume, fee-for-service environment can reduce quality.

In summary, while the reimbursement system has made advanced diagnostic methods like MRI widely affordable and accessible for patients, concerns have also emerged regarding overuse and quality of care. However, low reimbursement rates and volume-based incentives risk promoting quantity over quality. Accordingly, private hospitals can earn more by performing more scans, while radiologists in public hospitals must interpret a large number of images quickly (Rylands-Monk, 2014; Turkish Association of Radiology, 2018). Meanwhile, the increasing use of outsourcing in public hospitals feeds this cycle. All these financial dynamics can contribute to MRI overuse and potentially underdiagnosis. One radiologist explained that the payment model "leads to radiologists receiving proportionately less payment for their work compared to 2003, despite reporting significantly more images" (Rylands-Monk, 2014). In addition to these, Başar et al. (2019) found a 34% repeat MRI rate in a large public hospital and attributed the high number of repeats to inadequate image quality and expedited protocols. Okçan and Çakar (2015) argue that reforms in Türkiye foster a mindset of productivity at the expense of clinical relevance, a common criticism of healthcare systems undergoing neoliberal commodification. Thus, this implies that the financial incentives in the Turkish system are misaligned with diagnostic quality.

In conclusion, the HTP radically restructured the healthcare provision and financing. It achieved universal coverage, increased healthcare resources, including staff and facilities, but also introduced intense productivity targets and expanded private sector roles. Studies have shown that while performance-based pay systems and market liberalization aimed to increase efficiency, they often forced clinicians to see more patients in a limited time. This suggests that the reforms prioritized breadth of care over depth. In the next section, the study will examine how these institutional pressures are reflected in the various dimensions of diagnostic capacity (technological, human, institutional, and regional), which collectively affect diagnostic accuracy.

2.2. Diagnostic Capacity Framework in Türkiye

The study defines "diagnostic capacity" as the ability of a healthcare system to accurately diagnose diseases using diagnostic tools in a timely and appropriate manner. It separates capacity into technological, human, and dimensions and assesses each dimension in Türkiye. This identifies gaps and inequities, such as imaging device distribution inequities and workforce shortages, that are likely contributing to underdiagnosis under systemic pressure.

2.2.1. Defining Diagnostic Capacity

Diagnostic capacity is a multifaceted concept that encompasses how technology, human resources, institutional settings, and time interact to determine how well a healthcare system detects disease. Thus, diagnostic capacity refers to a system's ability to accurately and timely diagnose patient conditions, and this depends on the availability of diagnostic imaging tools, their effective use, and the competence of the professionals (Wright et al., 2021, p. 2) According to this definition, the study defines the basic dimensions of diagnostic capacity as technological, human, and institutional, in the Turkish health system.

2.2.2. Human Capacity

Human capacity includes the number and workload of healthcare professionals performing diagnostics, including doctors, nurses, technicians, and other support staff.

Figure 1 shows the trend in the number of specialist physicians, general practitioners, and medical assistants in Türkiye from 2002 to 2023. According to the graph, all three categories of healthcare professionals have significantly increased. This increase reflects both policy-driven workforce expansion and the broader transformation of the Turkish healthcare system under the Health Transformation Program launched in 2003.

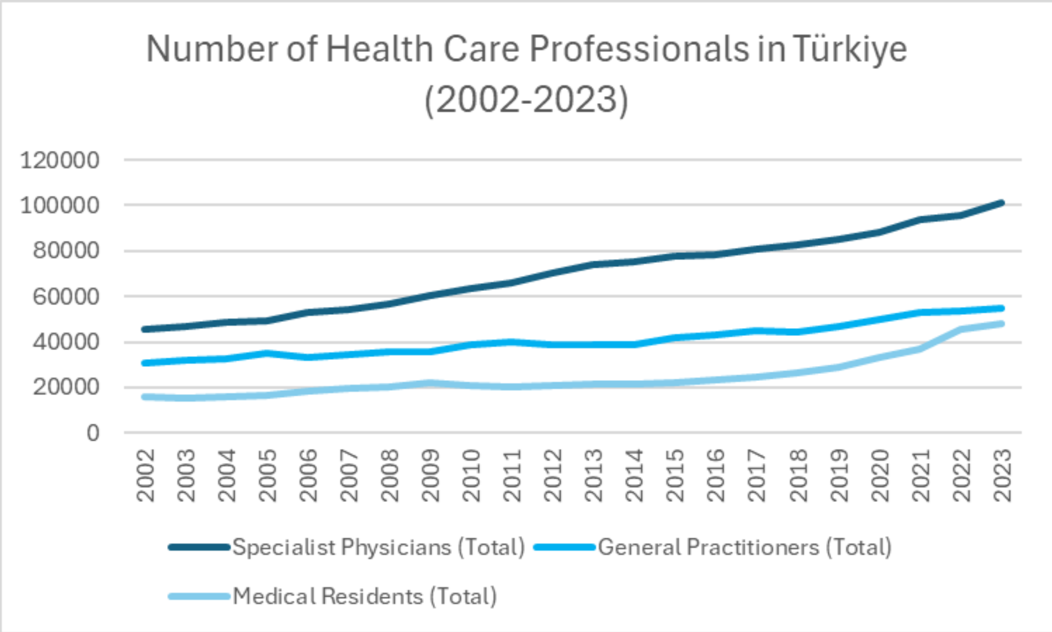

Figure 1. Number of Healthcare Professionals in Türkiye

Source: Author's own calculations based on data from *MEDAS Health Statistics Database* (Ministry of Health, Republic of Türkiye, 2023)

According to Figure 2, Türkiye experienced a gradual increase in the number of physicians relative to its population between 2002 and 2023. In 2002, this ratio was approximately 1.45 physicians per 1,000 people, reaching about 2.4 physicians per 1,000 people by 2023, indicating an improvement in workforce density. This growth, particularly over the last decade, coincides with the expansion of medical school quotas, the integration of more medical assistants into the healthcare system, and targeted recruitment policies aimed at addressing regional inequalities in access to healthcare. Despite these gains, Türkiye's physician-to-population ratio still lags behind the OECD average of around 3.6 per 1,000, underscoring a continued gap in human resource capacity (OECD, 2023).

Moreover, the uneven distribution of physicians across regions means that improvements in national averages may mask persistent shortages in peripheral provinces. These dynamics highlight both progress in expanding the workforce and the structural challenge of aligning physician supply with population needs. This imbalance in workforce allocation continues to shape diagnostic capacity, as physician shortages in certain regions exacerbate delays in access and contribute to unequal health outcomes.

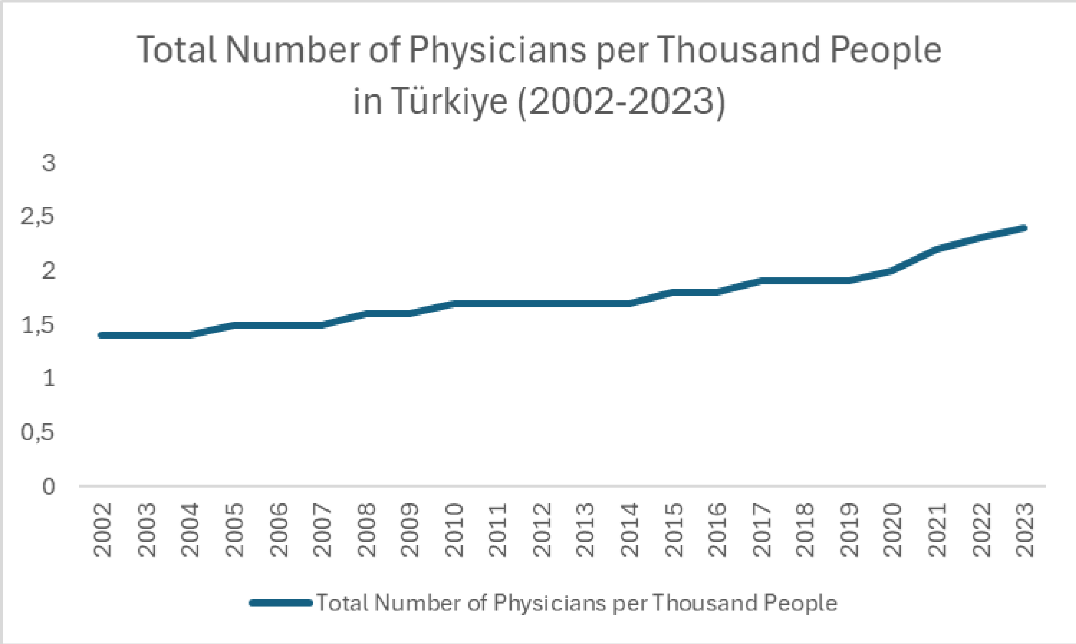

Figure 2. Total Number of Physicians per Thousand People

Source: Author's own calculations based on data from *MEDAS Health Statistics Database* (Ministry of Health, Republic of Türkiye, 2023)

While this increase signals progress in terms of workforce density, it still lags behind many high-income countries, suggesting that physician supply is a structural constraint on healthcare delivery. According to a recent OECD study, Türkiye has approximately 2.2 physicians per 1,000 people, below the OECD average of approximately 3.7. Furthermore, following the reforms implemented in Türkiye, there has been a significant increase in the use of public healthcare services. One of these is the annual number of physician consultations per capita. This number increased from 3-4 in the early 2000s to approximately 8 by the end of the 2010s (OECD, 2017). This situation has brought to the fore the paradox of quantity over quality. Despite the increasing number of consultations, the number of physicians per capita has not grown at the same rate, raising concerns about the quality of the healthcare system. In this context, according to the Turkish Medical Association (TTB), some hospital appointment systems allocate up to 100 patients per doctor per day, forcing doctors to limit each visit to just 5–10 minutes or less, despite the typical consultation duration being 20 minutes (TTB, 2021). All of this suggests that Turkish physicians perform a remarkably high volume of consultations. Workforce constraints are particularly severe in radiology. The Turkish Radiology Society reports approximately 5 radiologists per 100,000 people; with such a small staff, individual radiologists face extremely high workloads. TRD surveys have revealed that many radiologists are required to report 200-300 imaging studies per day, leaving only about five minutes per case. However, such workloads are "significantly insufficient" for a standard examination. As a result, radiologists work under intense stress and are at risk of burnout. The TRD warns that this "examination intensity" poses a risk to both professionals and patients: radiologists are working at risk, and patients face serious risks to diagnostic safety.

On the other hand, while auxiliary staff such as nurses and technicians also influence diagnostic capacity, their numbers are not directly observable in the study. Instead, their role is indirectly captured through temporal variation in services. In most hospitals in Türkiye, capacity varies significantly during the day and after hours. Most diagnostic services are provided during weekday working hours except emergencies, which are also available at night, on weekends, and on holidays. Therefore, workload can increase during the morning and early afternoon hours. Also, outpatient clinics are generally closed on weekends and public holidays, with only on-call teams handling urgent cases. Thus, staffing levels typically fall outside weekday working hours, especially at night, weekends, and public holidays. Thus, time- and day-of-week effects function as proxies for auxiliary staff availability, complementing human capacity in shaping diagnostic performance (Turkish Radiology Association, 2018). In summary, Türkiye’s shortage of medical staff and their heavy workloads mean the system has limited human capacity. This constraint contributes to underdiagnosis, as fatigued doctors may miss subtle findings or be unable to review charts in depth.

2.2.3. Institutional Capacity

In Türkiye, health care institutions and organizations serve three purposes: prevention, treatment, and rehabilitation. According to Law No. 5510, which classifies health service providers according to their functions and size:

> "ARTICLE 70 - For the purposes of implementing this Law, healthcare providers are classified by the Ministry of Health as primary, secondary, and

> tertiary healthcare providers. The referral chain between these healthcare providers is determined by the Institution, taking into account the diagnosis, preliminary diagnosis, and the specialties of physicians and dentists, nationwide or on a provincial or district basis, in consultation with the Ministry of Health. Family physicians are considered primary healthcare providers. For the Institution to provide healthcare services, universal health insurance holders and their dependents must comply with the referral chain rules." (Official Gazette of the Republic of Türkiye, 2006).

Türkiye's hospital system operates within a three-tiered healthcare model that organizes medical services into three levels: primary, secondary, and tertiary care. In this context, the Ministry of Health Regulation No. 31746 is the current document that categorizes the stage of healthcare service providers (Official Gazette of the Republic of Türkiye, 2022). According to this, Primary health care providers are healthcare institutions where patients are diagnosed and treated on an outpatient or inpatient basis, providing easy access to citizens' residential areas and providing affordable and effective healthcare services. It includes community health centres, family health centres, occupational physicians, 112 emergency health services, and polyclinics that are not affiliated with any hospital (Official Gazette of the Republic of Türkiye, 2022).

Secondary healthcare providers are healthcare institutions that provide outpatient or inpatient diagnostic, treatment, and rehabilitation services. They include public hospitals that are not training or research hospitals, integrated district hospitals (E1), oral and dental health hospitals and centres, private hospitals, private medical centres, medical centres, and specialized centres (Official Gazette of the Republic of Türkiye, 2022).

Tertiary healthcare providers are high-level hospitals equipped with advanced technology or infrastructure to provide education and research services for diseases requiring advanced examinations and specialized treatment, as defined in relevant legislation. Tertiary healthcare providers accept patients referred from outside the institution or from outside the province for tertiary healthcare services and meet their healthcare needs. Training and research hospitals affiliated with the Ministry, as well as hospitals belonging to state and foundation universities, are directly classified as tertiary healthcare providers (Official Gazette of the Republic of Türkiye, 2022). In this context, institutional capacity refers to the organization, structure, and role of healthcare institutions in the diagnostic process. Since diagnostic capability is impacted by the resources and duties that vary between hospital levels in Türkiye, it can be examined by hospital type and referral system.

There are studies that suggest that the institutional level is associated with diagnostic quality within the specific hospital type. For example, Özen et al. (2022) found that higher-tier institutions, on average, produce significantly higher-quality MRI exams than lower-tier institutions. This suggests that institutional capacity is important in terms of the expertise, protocols, and equipment involved. Teaching hospitals may generally have more stringent protocols, experienced staff, and double-reading of images, resulting in better accuracy. In contrast, smaller community hospitals may have less robust quality control, potentially leading to more incomplete or under-yielding scans.

Institutional capacity also includes how hospitals are organized and staffed, and thus, there are fundamental elements that measure institutional capacity in hospital services. These include the number of patient beds, the number of beds per population, the number of inpatients, the number of days in bed, the average length of stay, and the bed turnover rate, which indicates how many patients use a bed annually. Finally, institutional capacity is linked to whether a facility is a central referral centre or a peripheral provider. Because Türkiye doesn't have mandatory interhospital referrals, peripheral hospitals can attempt diagnoses that could be performed at specialized centres. However, a pattern exists where complex cases generally flow to tertiary centres.

After the HTP was launched in 2003 in Türkiye, the capacity and utilization of health services were rapidly increased. Quantitative expansions in Turkish healthcare can be seen in key health indicators on institutional capacity in Figures 3 and 4 below. According to these, hundreds of new hospitals and health institutions were established (Atun et al., 2013); the number of hospitals increased from 1156 to 1556, from 2002 to 2023.

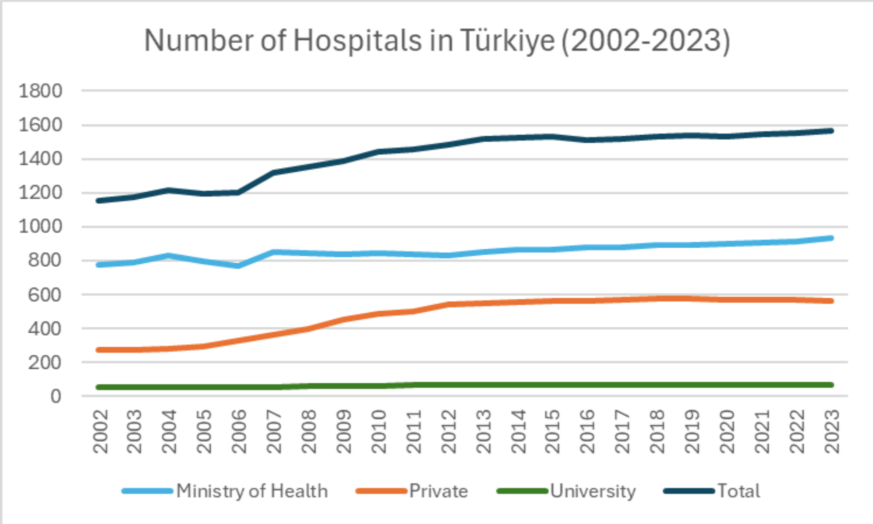

Figure 3. Number of Hospitals in Türkiye

Source: Author's own calculations based on data from *MEDAS Health Statistics Database* (Ministry of Health, Republic of Türkiye, 2023)

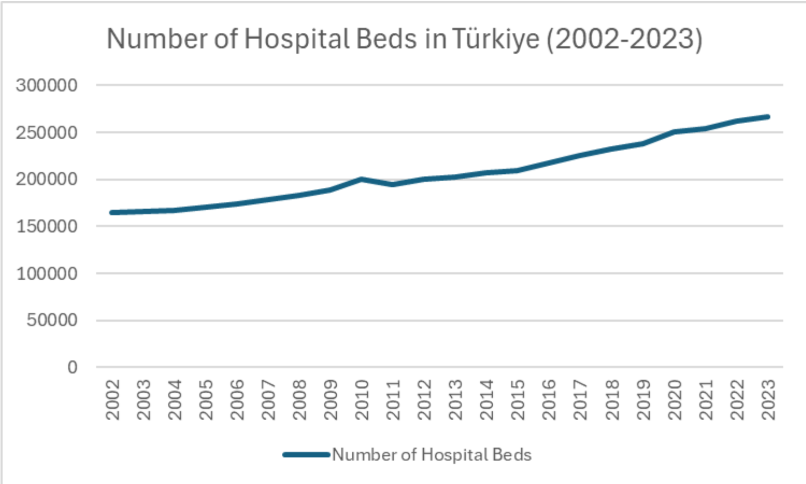

Figure 4. Number of Hospital Beds in Türkiye

Source: Author's own calculations based on data from *MEDAS Health Statistics Database* (Ministry of Health, Republic of Türkiye, 2023)

2.2.4. Technological Capacity

Technological capacity encompasses the availability and utilization of diagnostic devices such as MRI, CT, USG, and X-ray machines. In this context, the thesis will analyse parameters such as the availability of diagnostic devices at hospital and provincial levels, as well as the number of imaging scans per device, which reflect how intensively these devices are used in the Turkish healthcare system.

After the implementation of HTP reforms, Türkiye has expanded its technological capacity in the healthcare system, particularly in advanced imaging. The number of diagnostic technologies has risen, and their utilization has increased. In particular, utilization of MRI technologies has increased significantly in both public and private hospitals through large-scale investments in radiology infrastructure, an increase in the number of private service providers under SGK contracts, and easier access to patients.

Figure 5 shows trends in the number of MRI, CT, and USG units in Türkiye between 2008 and 2023. According to the figure, the number of MRI devices increased from approximately 500 to over 1,000, the number of CT devices from 800 to approximately 1,300, and the number of ultrasound devices from 2,000 to over 6,000. The expansion of MRI and CT technologies is consistent with private sector expansion, reimbursement schemes, and an increase in investment in advanced imaging technologies after the implementation of policy reforms in the HTP.

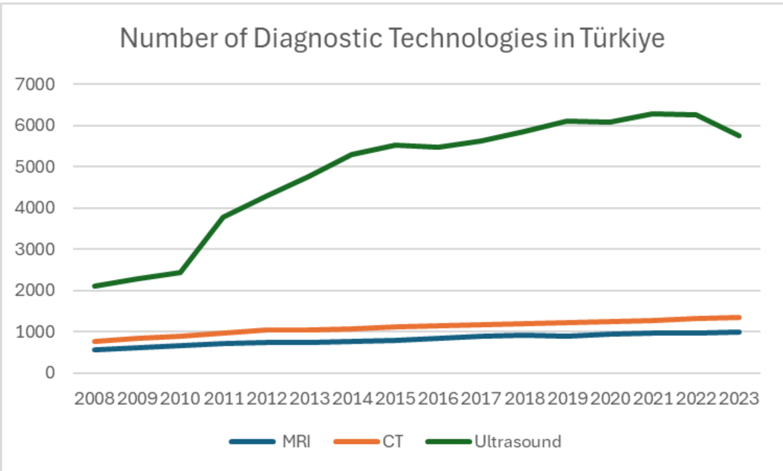

Figure 5: Number of Diagnostic Technologies in Türkiye (2008-2023)

Source: Source: Author’s own calculations based on *Health Statistics Yearbooks, 2008–2023* (Ministry of Health, Republic of Türkiye).

2.2.5. Regional Differences in Diagnostic Capacity

Geographic regions in Türkiye are divided into 26 subregions at the NUTS-2 level or 12 broader regions at the NUTS-1 level. Geographical regional distributions have shown historically distinct economic, social, and health development patterns. Accordingly, western and more urban regions, such as Istanbul, the Aegean, and Western Anatolia, including Ankara, have higher per capita physician density and more advanced medical equipment compared to less developed eastern and southeastern regions. To assess the diagnostic capacity of the Turkish healthcare system on a regional basis, three basic indicators for NUTS-1 regions for the year 2023 are presented in Table 1: Technological capacity as the number of MRI devices per 1,000,000 people, utilization density as the number of MRI examinations per 1,000 consultations, and institutional capacity as the total number of physicians per 100,000 people. According to Table 1, the number of MRI devices nationwide is 11.7 per million people. The number of devices per capita is above average in Istanbul (14.3), Western Anatolia (13.6), and the Aegean Region (12.0) because these regions are economically stronger and more urbanized, with better hospital infrastructure and private sector investments (OECD, 2017). The regions with the lowest MRI density are Southeastern Anatolia (9.4) and the Western Black Sea Region (9.7), demonstrating the unequal access to advanced diagnostic tools across Türkiye.

When looking at MRI examinations per 1,000 consultations, Northeastern Anatolia (43.7) and Istanbul (42.8) show the highest MRI utilization rates, while Southeastern Anatolia (27.3) and Central Eastern Anatolia (28.3) have the lowest. This suggests that overutilization may reflect supplier-induced demand (Zabrodina et al., 2020), while Southeastern Anatolia, with its lowest rate, is at risk for underutilization due to infrastructure deficiencies or limited specialist access, reflecting concerns about diagnostic underutilization (Hutubessy et al., 2002; Huang et al., 2023).

Table 1. MRI Technologies according to NUTS-1 Regions, 2023

	Number of MRI Devices in Hospitals per 1.000.000 Population by NUTS-1, 2023	Number of MRI Exams in Hospitals per 1.000 Consultation by NUTS-1, 2023	Number of Total Physicians per 100.000 Population by NUTS-1, 2023
Türkiye	**11,7**	**35,2**	**239**
İstanbul	14,3	42,8	298

Table 1. (continued)

Western Anatolia	13,6	34,8	368
Aegean	12	34,4	247
Eastern Blacksea	11,7	37,7	222
Mediterranean	11,6	33,9	216
Northeastern Anatolia	11,1	43,7	191
Western Marmara	11	38,1	205
Eastern Marmara	10,6	34,5	213
Mideastern Anatolia	10,5	28,3	182
Central Anatolia	10	31,4	204
Western Blacksea	9,7	32,7	204
Southeastern Anatolia	9,4	27,3	155

Source: Author's own illustration based on data from *Health Statistics Yearbook 2023* (Ministry of Health, Republic of Türkiye, 2024)

Considering the total number of physicians per 100,000 people, the national average is 239, but it varies significantly by region: Western Anatolia (368) and Istanbul (298) have the highest physician density. Southeastern Anatolia (155) and Northeastern Anatolia (191) have well below average physician density, which can lead to excessive workload per physician and an increased risk of underdiagnosis due to time pressures (Başar et al., 2019; Clemens and Gottlieb, 2014).

Consequently, there are multiple disparities in diagnostic capacity across Türkiye: For example, despite a low physician density (191), Northeastern Anatolia has a relatively high rate of MRI use (43.7%). This suggests an overreliance on diagnostic imaging methods to compensate for a lack of clinical knowledge or personnel. In addition, there are studies that have examined regional capacity differences in healthcare in Türkiye. Specifically, regarding radiological imaging devices, Çınaroğlu and Başer (2017) show that Türkiye's imaging technology is clustered: while urban centres may have abundant MRI and CT equipment, many smaller provinces have very few. This uneven technological capacity requires patients in rural or impoverished areas to wait long periods or travel to major cities for diagnosis. Conversely, hospitals with high-tech equipment are experiencing screening bottlenecks under pressure, while others are underutilized. In either case, inadequate spatial distribution may lead to diagnostic delays or underdiagnosis.

2.3. Radiology Growth in Türkiye

One of the quantitative advancements Türkiye has made since the HTP is in the field of radiology. Over the past two decades, technological capacity has grown rapidly, driven by both public investment and private sector growth. While there were only 58 MRI units nationwide in 2002, this number has risen to 1,001 by 2023, as can be seen in Figure 5 above.

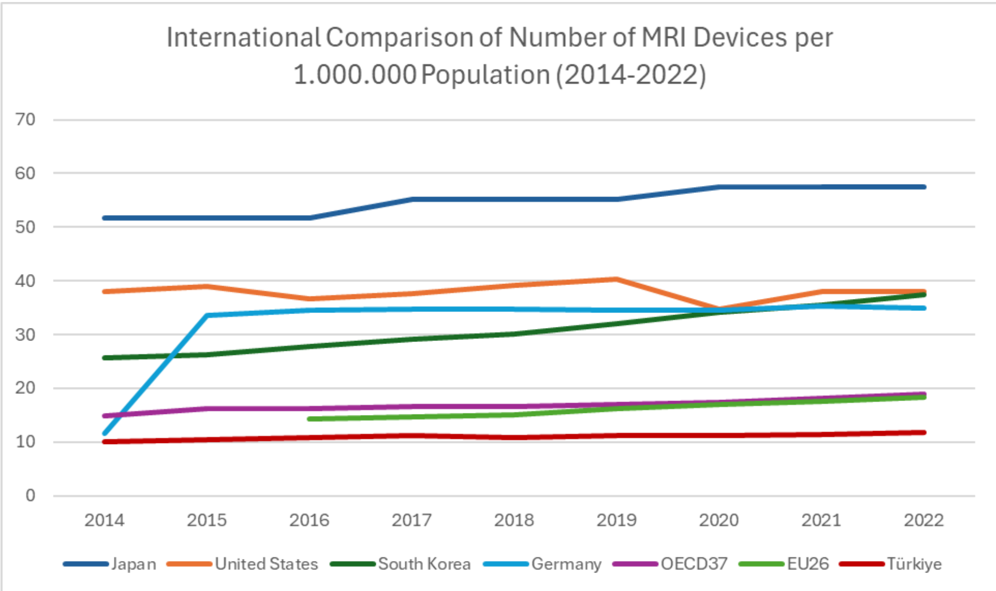

Figure 6: MRI Devices per 1,000,000 Population (2014–2022)

Source: Source: Author’s own calculations based on *Health Statistics Yearbooks, 2014–2022* (Ministry of Health, Republic of Türkiye).

Figure 6 shows the evolution of MRI device availability per 1,000,000 population in selected countries from 2014 to 2022. There is a noticeable difference in the relative availability of MRI devices per capita in Türkiye compared to other countries. According to the chart, in 2022, Türkiye had only 11.7 devices per million population, compared to 58 in Japan, 39 in the US, and over 25 in countries like South Korea and Germany. While the OECD average is above 18, Türkiye remains well below this threshold. This suggests that Türkiye's physical diagnostic capacity is limited in terms of equipment stock. Despite a slight increase from 2014 to 2022, Türkiye's investment in imaging infrastructure has not kept pace with the growing demand. The low number of machines per capita in Türkiye raises concerns that every machine is overutilized, especially in high-demand public hospitals. In this context, limited device availability per capita can lead to overuse of existing machines, unequal regional access, and reduced diagnostic yield; these factors will be examined in the empirical sections of this thesis.

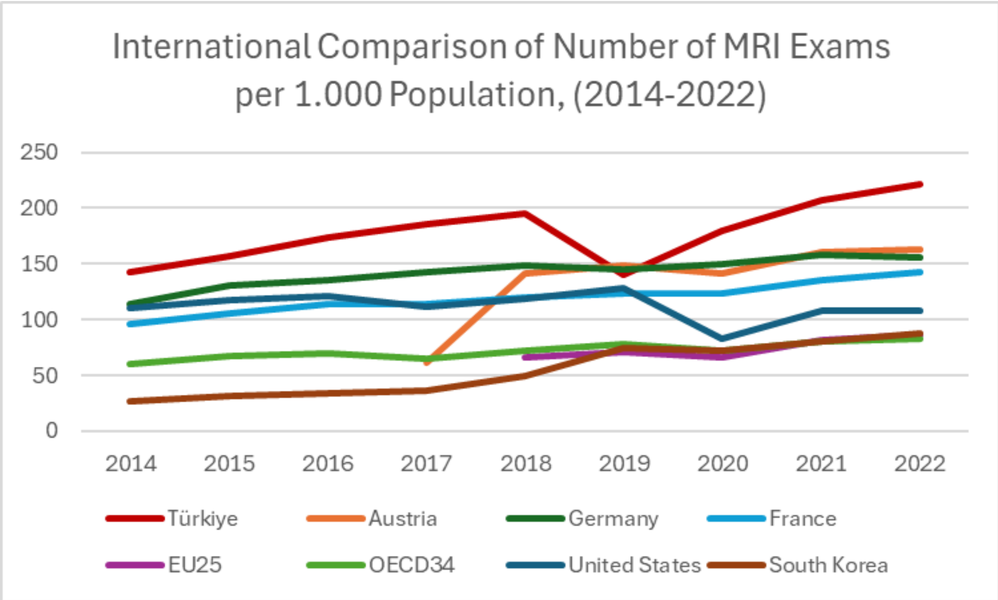

Figure 7: Number of MRI Exams per 1,000 Population (2014–2022)

Source: Author’s own calculations based on *Health Statistics Yearbooks, 2014–2022* (Ministry of Health, Republic of Türkiye).

Paradoxically, while the number of devices per capita is below the OECD level, it performs the highest number of MRI examinations per capita in Figure 7. Accordingly, in 2022, while OECD and EU averages hover around 100, Türkiye surpassed 220 examinations per 1,000 people, compared with approximately 160 in Austria and around 150 in Germany and France. Despite having fewer MRI devices, Türkiye's per capita use is higher than in countries with twice the equipment. This likely reflects structural pressure on the system to meet the increasing demand for diagnostics due to the referral-based reimbursement model, the lack of gatekeeping in primary care, and patient expectations (Jahanmehr et al., 2019; Zabrodina et al., 2020). The result may be supplier-induced demand, repeated scans due to scheduling pressures, or defensive medical practices. Thus, high per capita use with limited infrastructure often indicates quantitative overuse at the expense of diagnostic quality.

Looking at the scans per device further reveals the intensity of utilization in Türkiye. Figure 8 shows that Türkiye performed over 18,000 scans per MRI machine in 2022. This figure is well above the OECD average (~5,000) and significantly higher than other upper-middle-income or high-income countries such as Germany (~9,000) and South Korea (~3,000). This exceptionally high utilization suggests that MRI devices were operating at or near full capacity throughout the year in Türkiye. This raises concerns about scan quality, maintenance schedules, waiting times, and repeated scans or underdiagnosis. Diagnostic capacity is not only related to the number of machines; it also includes time availability, human resources, and technical quality (Martella et al., 2023; Başar et al., 2019). When the number of scans per scanner is extremely high, diagnostic accuracy can be compromised due to shortened scanning protocols, reporting delays, or inadequate follow-up. This can contribute to a high repeat screening rate in Türkiye and underdiagnosis of complex cases.

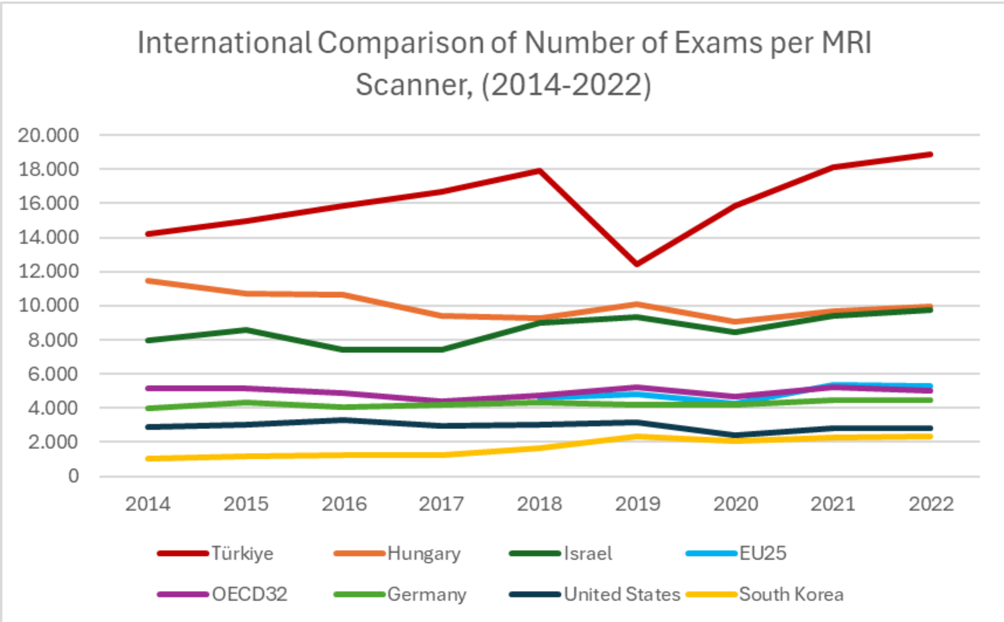

Figure 8: Number of MRI Exams per MRI Scanner (2014–2022)

Source: Author’s own calculations based on *Health Statistics Yearbooks, 2014–2022* (Ministry of Health, Republic of Türkiye).

Overall, the expansion of radiology services in Türkiye has been remarkable, marked by rapid increases in MRI, CT, and USG capacity following the HTP. While this growth has improved access and reduced geographic disparities, it has also generated new challenges, including risks of supplier-induced demand, defensive medicine practices, and diagnostic inefficiency. These developments underscore the importance of examining how rising volumes intersect with diagnostic quality and uncertainty. Building on this context, the next chapter reviews the theoretical and empirical literature on underdiagnosis, defensive medicine, and supplier-induced demand, which provides the conceptual foundation for the econometric analyses in later chapters.

CHAPTER 3

UNDERDIAGNOSIS IN DIAGNOSTIC IMAGING: THEORETICAL FRAMEWORK AND LITERATURE REVIEW

3.1. Defining Underdiagnosis

Before moving on to the theoretical framework, it is necessary to understand what underdiagnosis means and how it is defined. In this context, in this section, the concept of underdiagnosis is first defined, and then the rationale for using R-codes as a practical proxy for underdiagnosis will be explained. Finally, the section explores how diagnostic uncertainty shapes patient experience, influencing care pathways, repeat visits, and trust in the health system. Together, these subsections establish both the theoretical basis and the empirical approach used to study underdiagnosis in diagnostic imaging.

3.1.1. Concept of Underdiagnosis

Underdiagnosis is the situation that usually occurs when the diagnostic process fails, resulting from acts of negligence such as underuse of a test (Newman-Toker, 2014). This definition is practically applied as a diagnostic code in the ICD-10 symptom-based codes, falling within the range R00-R99 "symptoms, signs, and abnormal clinical/laboratory findings not elsewhere classified" It is described in CMS as “Unspecified” codes “titled “unspecified” are for use when the information in the medical record is insufficient to assign a more specific code” (CMS, 2024, p.9). It is allowed to record the appropriate "unspecified" code when there is not enough clinical information available to give a more specific code, such as when a diagnosis of pneumonia is confirmed but the precise kind is not known (CMS, 2024, p.17). As a result, a healthcare system's excessive R-code utilization implies that a large number of patients remain unaddressed, which may indicate systemic diagnostic inefficiency. Starting from his point of view, the study operationalizes underdiagnosis with a binary indicator, R_Code_Prob that flags whether a patient's visit contained any R codes, implying that no definitive diagnosis was made for that encounter.

3.1.2. Rationale and Limitations of R-codes as a Measure

This study assumes that using R codes as a proxy for underdiagnosis is an imperfect but practical approach. On the one hand, frequent use of R codes can signal diagnostic shortcomings: If a healthcare provider frequently uses “symptoms, unspecified” codes, this may indicate an inability to fully diagnose diseases, possibly due to time constraints, insufficient expertise, or lack of resources. Previous research highlights that diagnostic uncertainty is a prevalent issue and a recognized cause of diagnostic error rates (Alam et al. 2017).

Accordingly, underdiagnosis can be characterized as a latent medical error if the patient leaves without receiving the proper diagnosis, which can be harmful in the long term because of the delayed diagnosis (Balogh et al., 2015; Graber, 2013). Actually, it is not explicitly incorrect treatment; thus, it is crucial to keep in mind that not all R code usages result in "failures." Because medicine is inherently uncertain, there are situations in which a symptom code is acceptable, such as in the early stages of a disease when a wait-and-see strategy is clinically sound or when present techniques are actually unable to provide a conclusive diagnosis. As a result, mistakes and underdiagnosis are not synonymous. Therefore, underdiagnosis is not a synonym for error. However, notable variations in R code usage rates between settings might indicate performance problems at the system level. For instance, while Hospital A uses R codes in only 5% of visits, Hospital B uses them in 15% of raises questions about the reasons for these differences in terms of complexity, training, or resources between the two institutions. Thus, the consistently high rates of underdiagnosis likely reflect a significant diagnostic inefficiency (OECD, 2025).

Numerous studies have used R codes as indicators of underdiagnosis. A study in the UK investigated the frequency of ill-defined (R-coded) hospitalizations in older adults and the factors predicting these. An R-coded primary diagnosis was observed in 21.6% of elderly patient admissions (Walsh et al. 2011). Similarly, nonspecific diagnoses were observed in 15-50% of emergency department discharges in Denmark. More importantly, studies have shown that 11-20% of emergency department deaths occurring within 8-30 days were preceded by a nonspecific (R-coded) discharge. Accordingly, nonspecific diagnoses accounted for one-third of the patients and contributed to mortality in terms of the total number of deaths (Christensen, 2016). These statistics demonstrate that patients with R codes have serious illnesses that are overlooked or delayed and support the use of the R code as an indicator of underdiagnosis.

In Türkiye, the Turkish Ministry of Health officially uses the ICD-10 guideline, and ICD-10-R code definitions are directly used in diagnostic systems. This rule went into effect with the SUT in 2007, and health reports that do not contain ICD-10 codes are deemed unacceptable because they do not comply with the SUT. Thus, in healthcare services, ICD-10 codes must be included in diagnostic reports. Numerous studies have also been conducted using R code. For example, an examination of patients presenting to the emergency department in Antalya, Türkiye, revealed that 15.71% were identified with the R code (Kılıçaslan et al., 2002). Another study evaluated patients presenting to the emergency department in Ankara and found that the most common diagnoses in the emergency department were abdominal pain (R10.4) at 10.5% and nausea and vomiting (R11) at 7.2% (Yüksel, 2020). Additionally, in an analysis conducted in a paediatric emergency department, code R68.8 (general symptoms and other signs) ranked first with 26.77%, followed by R50.9 (fever of unknown origin and other origin - fever, unspecified) with 14.27%, and R11 (nausea and vomiting) ranked second with 10.18%. This means that 50% of paediatric emergency department patients are observed with an R code. The study also found that the initial readmission was 1.35% within 24 hours and 4.3% within the first 72 hours. When examining the diagnosis code of repeat patients, it was observed that R68.8 was first, R50.9 second, and R11 cases were third in order of recurrence (Akçay and Gül, 2022). According to these data, it is common practice to label a visit as an R code in situations where the precise disease cannot be determined. This allows for the measurement of underdiagnosis in Türkiye.

3.1.3. Diagnostic Uncertainty and Patient Experience

Underdiagnosis has tangible effects on patients, and it also contributes to a state of medical uncertainty since leaving the hospital without a clear answer requires further consultations, testing, or referrals, effectively prolonging the diagnostic process. Studies on diagnostic uncertainty show that patients with "unspecified disease" codes, R codes, typically experience increased follow-up healthcare use, including longer hospital stays and higher readmission rates. Marshall et al. (2022) found that children hospitalized with uncertain diagnoses had noticeably longer stays and higher 30-day readmission rates compared to similar patients with definitive diagnoses. These show that underdiagnosis tends to delay resolution, potentially leading to a series of additional healthcare visits.

An underdiagnosis can be considered as a "missed opportunity"—the patient is not mistreated, but neither is the proper therapy provided since the underlying reason is still unclear. This is distinct from an outright misdiagnosis, which occurs when a definitive but inaccurate diagnosis is established. However, over time, this can also be dangerous since the underlying condition of the illness may worsen, or patients may lose trust and become anxious when switching between medical professionals. Thus, underdiagnosis has been recognized as a systemic safety problem, a hidden error that has the potential to significantly lower healthcare quality (Newman-Toker, 2014; OECD, 2025).

As the OECD (2025) aptly notes, “underdiagnosis is inherently burdensome; chronically underdiagnosed patients will continue to seek care as their condition worsens, leading to higher costs and poorer outcomes” (OECD, 2025). Therefore, reducing underdiagnosis rates is important not only for individual patient well-being but also to prevent the accumulation of “hidden” costs and harms within the healthcare system. From that point of view, the thesis asserts that underdiagnosis is fundamentally burdensome. Therefore, lowering underdiagnosis rates is crucial for the health of individual patients as well as for avoiding the buildup of "hidden" expenses and damages within the healthcare system.

3.2. Theoretical Framework

The thesis is based on Arrow’s theory on uncertainty in medical care (Arrow, 1963), which highlights that healthcare markets are characterized by uncertainty and information asymmetry, where patients rely on physicians' expertise for decision-making, creating a principal-agent problem. In such environments, patients rely heavily on physician recommendations, and the demand for services like MRI is largely physician-driven. When combined with financial or institutional incentives, this uncertainty can easily lead to overutilization of services. Therefore, the question of the study arises whether the high use of MRI technologies in Türkiye reflects a truly greater need for more diagnostic procedures or the influence of physician practice styles and incentives under uncertainty.

3.2.1 Arrow’s Theory of Uncertainty in Medical Care

According to Kenneth Arrow (1963), uncertainty is intrinsic to medicine. He argues that the "special economic problems of medical care" (p. 941) arise from “adaptations to the existence of uncertainty in the incidence of disease and in the efficacy of treatment.” Therefore, fundamental risk and uncertainty are unavoidable due to the unpredictable nature of disease, and the outcomes of interventions cannot be guaranteed. Since patients lack knowledge about their own health and the treatment of their diseases, they trust their physicians. Arrow explains this ‘*asymmetric information’* as "the physician's knowledge of the outcomes and possibilities of treatment is necessarily much greater than the patient's...and their relationship is coloured by this knowledge."(p. 951).

Asymmetric information (Akerlof, 1970) creates a principal–agent relationship (Jensen & Meckling, 1976) between patient and physician. The patient (principal), who lacks the expertise to make decisions about the disease and its treatment, delegates decision-making authority to the physician (agent) and trusts them to act in the patient’s best interests. However, the principal-agent dynamic creates incentive problems. The physician may have ‘*incentives’* that are not aligned with the patient's health goals. The information gap resulting from asymmetric information leaves the patient lacking the necessary information to justify the recommended tests or treatments. In this situation, Arrow observed that physicians tend to increase utilization and expenditures. Information asymmetry requires trust, but it can also lead to agency conflicts if incentives are misaligned. Arrow (1963) also describes '*moral hazard*', which occurs from the unique nature of health insurance, where both patients and providers are insulated from the financial consequences of healthcare decisions. He observes that when patients are insured, they may demand more care than they would if they paid full price, and physicians may be more likely to recommend more expensive care. This creates a fertile environment for moral hazard, particularly in outpatient or general practice settings where oversight is weaker than for surgery. Here, patients protected from costs may be less discriminatory, and physicians whose income comes from reimbursed services face less resistance when requesting these services. Thus, Arrow's analysis highlights how both asymmetric information and lack of cost coverage led to inefficiency and potential overutilization in the healthcare sector.

In this theoretical framework, diagnostic aggression arises as a result of all these factors. Accordingly, under conditions of uncertainty and asymmetric information, physicians are often inclined to overtest rather than undertest; they order additional tests, imaging studies, or procedures to avoid missing a potential diagnosis. This tendency can be enhanced by medical and financial incentives: In many healthcare systems, a physician's performance is evaluated by the number of procedures performed, and they may be rewarded for over-testing. In contrast, undertesting can lead to potentially serious consequences if they miss a serious condition is missed. Consequently, the protection from uncertainty leads to aggressive diagnostic testing, which in turn gives rise to defensive medicine. In Arrow's framework, such overuse is not mere irrationality; it is a predictable outcome when uncertainty is high, information is unequal, and incentives favour extra intervention over restraint.

3.2.2. Arrow's Theoretical Framework on MRI Utilization in Türkiye

Arrow's theory of uncertainty in healthcare sheds light on why and how diagnostic imaging can be overutilized in healthcare systems. It is possible to evaluate his theoretical framework in the overutilization of MRIs in the Turkish healthcare system over the past two decades. In chapter 2, the study has demonstrated statistically that Türkiye expanded its hospital capacity, including advanced imaging, following health reforms after the HTP: Compared to other OECD countries, the number of MRI devices per capita is well below the national average, while the number of scans per MRI device and the number of scans per population are among the highest among these countries.

This high volume of MRI scans suggests that MRIs are used much more intensively in Türkiye than elsewhere, suggesting the overuse hypothesis. First, doctors in Türkiye are overworked and may resort to high-tech diagnostic tests like MRIs, shortening physical examinations and observations. This can be directly linked to the *theory of uncertainty*: a doctor under pressure and uncertain about a definitive diagnosis may use imaging as an extension of the examination. Secondly, *'the incentives'* identified in Arrow's framework help explain this pattern. In Türkiye, financial incentives operate under performance-based payment plans that reward productivity measured by the number of services provided, as implemented after the HTP reforms. This corresponds to the incentives framework mentioned by Arrow, where providers use their knowledge advantages to increase service volume. Thirdly, with universal health coverage implemented after HTP, patients pay little or nothing out of pocket for an MRI, which encourages demand, consistent with Arrow's *moral hazard* scenario. Patients with UHC often perceive MRIs as a necessary diagnostic step for reassurance and may insist on undergoing MRIs, further leading to overutilization.

This conclusion is also supported by recent international studies: One study found widespread overuse of medical imaging in low- and middle-income countries, with an average of 35% of MRI scans being unnecessary (Protheroe, 2024). Similar trends are seen in high-income countries; for example, in the United States, it is estimated that one-third of MRI or CT scans are inappropriate or of low value (Hendee, 2009, p. 32).

Globally, factors contributing to MRI overuse include insurance, patient expectations, defensive medicine, and fee-for-service (Hendee, 2009, p. 32). These characteristics are present in Türkiye, but what distinguishes it is the extent of overutilization: the combination of demand and supply market willingness to use MR devices has led Türkiye to reach very high usage rates, far surpassing many countries. Therefore, Arrow's theory offers a lens through which to understand the overuse of MRI in Türkiye as a result of three factors: a) *uncertainty* because of over workload of doctors and thus widespread use of MR as a diagnostic safety net; b) *asymmetric information with incentive structures*: doctors, who hold superior medical knowledge, may exploit this imbalance by recommending additional imaging when their pay for performance system rewards higher volumes of testing. and c) *moral hazard*: when patients are largely insulated from the costs of imaging through insurance coverage, they may be more willing to accept or even request additional scans.

Although doctors turn to advanced technological diagnostic methods under uncertainty, their intensive use does not mean that every patient is diagnosed with certainty; underdiagnosis remains a systemic problem. In Türkiye, approximately 10% of all hospital visit diagnoses carry an ICD-10 "R code". In this context, ICD-10 R-codes serve as an observable proxy for Arrow’s concept of diagnostic uncertainty: they signal encounters where the physician, facing time constraints, incomplete information, or misaligned incentives, does not translate symptoms into a definitive diagnosis.

This underdiagnosis pattern can also be interpreted through Arrow's framework. Time and resource constraints that drive doctors to order tests may lead them, under pressure to move on to the next patient, to treat the current symptom without further investigation. Uncertainty in medicine cannot be easily eliminated with technology. Despite abundant imaging, doctors often avoid rapid diagnosis, and patients may cycle between tests, resulting in a nonspecific code.

Another reason could be quality issues: Başar et al. (2019) found that a significant portion of MRI scans in Türkiye require repeats due to inadequate quality artifacts, incorrect technique, or inexperienced readers. If the initial imaging is inconclusive, the diagnostic process continues, leading to repeated scans and therefore delays. As

Arrow (1963) puts it, "When there is uncertainty, information or knowledge becomes a commodity. Like other commodities, it has a production cost." This highlights the limitations and costly production of information: more imaging provides more data, but not necessarily more certainty.

In summary, time and resource constraints, along with the complexity of the disease, result in underdiagnosis, reflecting the uncertainty Arrow emphasized in 1963. Physicians acting as agents of uncertainty often use existing MRI scans to narrow information gaps, explaining overuse. However, the incompleteness of information and the difficulty of accurately identifying some diseases lead to underdiagnosis. Arrow's theory helps us understand why increasing the amount of service when information remains asymmetric and incomplete does not guarantee optimal outcomes. MRI use in Türkiye is a case in point: abundant scans yield diminishing returns on diagnostic yield, as seen in the persistent presence of R-coded diagnoses.

Figure 9 summarizes the analysis of the thesis based on Arrow’s theoretical framework. Anchored in Arrow’s (1963) economics of uncertainty, the diagram integrates agency and capacity to explain why overutilization and underdiagnosis can coexist. When illness outcomes are uncertain and information is asymmetric, the patient (principal) relies on the physician (agent), creating scope for moral hazard amplified by insurance and volume-oriented payment (pay-for-performance, fee-for-service). These incentives increase the demand for diagnostic resolution, but the ability to convert tests into information is bounded by capacity along three pillars: human (physicians and auxiliary staff), institutional (hospital role, centralization/referral intensity, bed turnover, scanner use per machine), and technological (MRI availability, utilization, and quality).

Accordingly, when incentive-driven volume meets constrained or misallocated capacity, the marginal informational yield of additional imaging falls: congestion, rushed workflows, and uneven expertise increase diagnostic noise, producing repeat MRI (overutilization) while simultaneously raising the probability of R-codes (underdiagnosis).

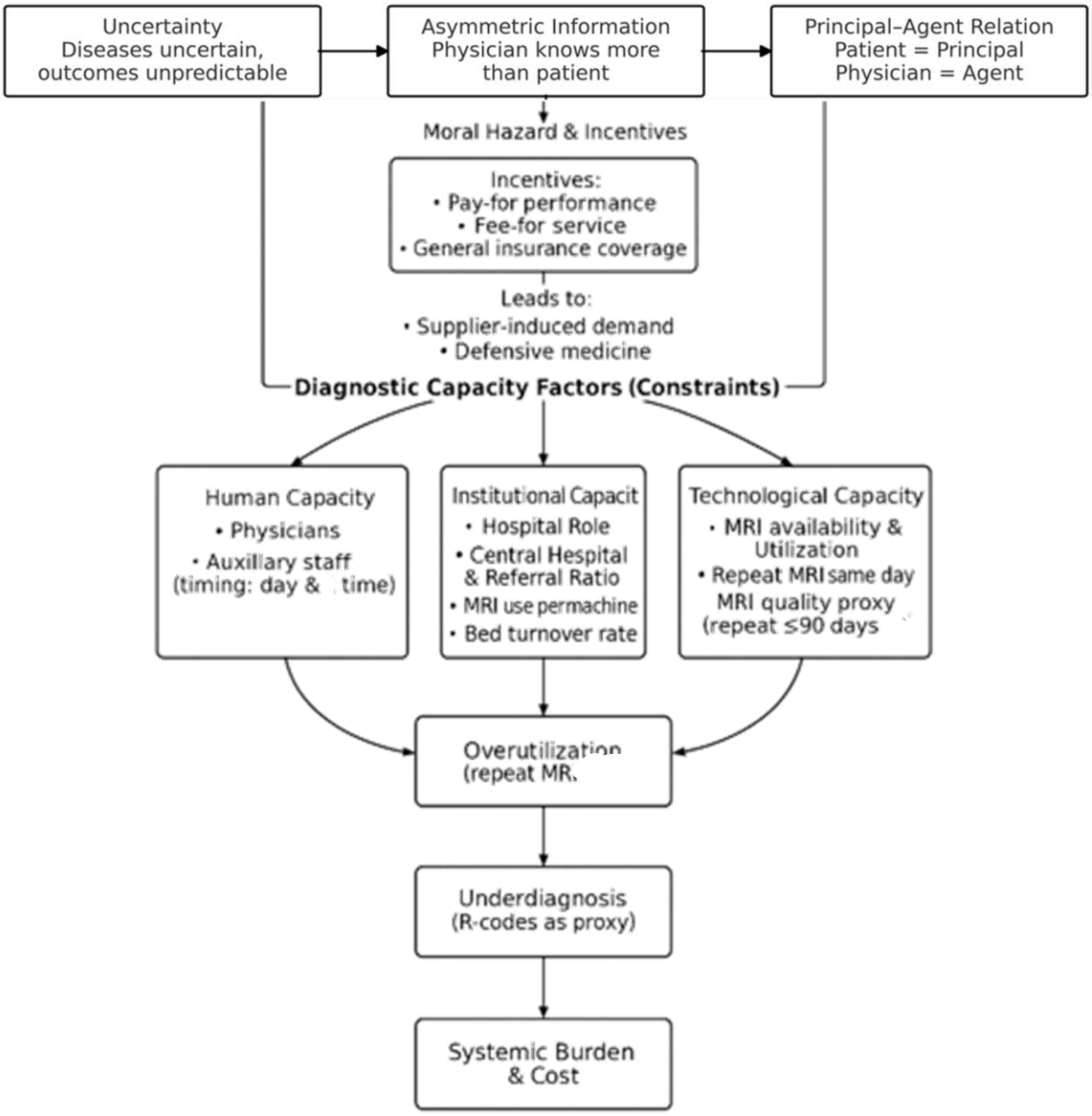

Figure 9. Diagnostic Capacity under Arrow's Uncertainty

Source: Author's own illustration.

Thus, under Arrow's framework, capacity does not merely limit quantity; it mediates the translation of tests into reliable information, shaping equilibrium diagnostic precision. Empirically, the model predicts that stronger volume incentives and tighter human/institutional/technological constraints are associated with (i) higher repeat-MRI rates and (ii) higher R-code probabilities, thereby linking incentive design and capacity to systemic burden and cost. Thus, following Arrow's framework, the study will test empirically whether systemic capacity constraints, human, institutional, and technological, raise the probability that a consultation ends in an R-code diagnosis, that is, *underdiagnosis*.

3.3. Literature Review

3.3.1. Supplier-Induced Demand and Financial Incentives

Arrow's information asymmetry opens the door for what health economists term '*supplier-induced demand (SID)'* in the literature (Evans, 1974). Physicians act as agents for patients, but they may have a financial or institutional incentive they may recommend more tests or visits than are clinically necessary. In fee-for-service and dual-payment systems, healthcare providers may respond to financial incentives by increasing the quantity of services provided (Clemens and Gottlieb, 2014; Zabrodina et al., 2020). This behaviour has been particularly observed in the case of diagnostic imaging, where technological infrastructure allows for high efficiency and profitability. Numerous empirical studies in the literature demonstrate SID in healthcare systems (McGuire, 2000). For example, Brekke et al. (2017) analysed how general practitioners in Norway responded to a change in reimbursement and found that increasing fee-for-service payment led to a significant increase in the number of patient visits while reducing the time spent per visit. This suggests a shift toward quantity over quality and illustrates how volume-based payment can motivate more service delivery.

On the demand side, the classic theory of moral hazard in health insurance (Pauly, 1968) suggests that patients may consume more healthcare services than they need when faced with little or no cost for services. Türkiye's healthcare system offers a case study of how financial incentives can increase demand for MRIs. Under the Health Transformation Program, most MRI services are covered by public insurance, and physicians are rewarded for higher service output. In public hospitals, radiologists are typically paid under a performance-based supplementary payment scheme, measured by the number of procedures performed. Radiologists are paid for each imaging test they interpret, which naturally creates an incentive to prescribe more tests (Başar et al., 2019). The OECD (2017) reported that MRIs in Türkiye are "systematically prescribed to patients with various health conditions," clearly leading to overuse. From a SID perspective, healthcare providers and hospitals in Türkiye have financial incentives to maximize MRI scans, potentially leading to a supply-driven utilization that exceeds clinical justification.

In imaging technologies, SID is most common in fee-for-service systems or self-referral settings. There is research in the literature supporting SID in diagnostic imaging. For example, Zabrodina et al. (2020) showed that hospitals in Switzerland responded to tighter inpatient budgets by ordering more outpatient MRI/CT scans after a reimbursement change. They achieved this through repeated examinations, highlighting how financial incentives can drive overuse. Additionally, a study in Japan found that hospitals increased MRI scans per patient when they experienced revenue shortfalls from other services. These cases suggest that when providers benefit from performing more tests, they may order additional imaging, potentially leading to overuse of MRIs. Provider-driven demand also applies to underdiagnosis: ordering more MRIs does not guarantee better diagnostic results; this may reflect overuse driven by incentives.

3.3.2. Defensive Medicine and Malpractice Pressure

Another driver of imaging overuse is *defensive medicine*, the practice of ordering tests or treatments using imaging technologies as a defensive tool to protect physicians from potential malpractice liability due to fear of missing a diagnosis. Brownlee et al. (2017) identified defensive medicine as a significant driver of overuse. They noted that it is associated with aggressive diagnostic testing, particularly in litigious environments such as the United States. In a survey of physicians across multiple countries, most physicians cited medico-legal concerns as a reason for ordering unnecessary tests (Brownlee et al., 2017). Imaging technologies are a common target for defensive applications. For example, in emergency departments, physicians may order an MRI or CT scan for a minor head injury to avoid missing a brain haemorrhage and risking subsequent litigation. Tung et al. (2018) studied emergency department imaging overuse and found that defensive imaging influences physician behaviour in numerous studies. The emergency room environment reinforces the defensive impulse because of high uncertainty and lack of follow-up.

In Türkiye, malpractice has evolved over the past two decades, increasing the prevalence of defensive medicine. Legal reforms in 2004 and 2012 expanded patients' ability to sue or seek compensation for medical errors (Başar et al., 2019, p. 260). While Türkiye's malpractice litigation rates are lower than those in the United States, they are rapidly increasing (Özmen et al., 2015; Arıkan et al., 2017). Başar et al. (2019) examine this trend and note a "significant move toward more defensive medicine" in Türkiye following these legal changes. As malpractice lawsuits have increased, Turkish physicians have increasingly adopted precautionary measures, including the liberal use of diagnoses to document comprehensive care (Başar et al., 2019). Non-invasive and readily available imaging tests are often the first line of defence. In summary, defensive medicine offers a compelling explanatory framework, alongside financial incentives, for why MRI overuse occurs: Even without a profit motive, physicians may overuse MRIs to protect themselves, contributing to the overall high volumes we observe.

3.3.3. Patient Expectations and Demand Factors

On the demand side, patient expectations may also contribute to MRI overuse. In many healthcare systems, patients equate comprehensive care with advanced diagnostics; an MRI may be viewed as the “best” or most definitive test, and some patients may insist on an MRI even when it is not medically necessary. Clinician surveys have found that patient demand or expectation is a frequently cited reason for requesting unwarranted imaging. A qualitative synthesis found that patients and clinicians commonly view diagnostic imaging as essential to locate the source of non-specific low back pain and that clinicians sometimes order imaging to manage patients’ expectations or reduce perceived medico-legal risk, while patients often underestimate the harms of unnecessary tests (Sharma et al., 2020). Survey evidence aligns with this: 54.3% believed imaging was necessary for the best LBP care, and 48.0% believed everyone with LBP should be imaged (Jenkins et al., 2016). In a large Danish cohort, roughly one-third of patients believed imaging was important for LBP management, and 26.1% wanted imaging at the initial consult; wanting imaging independently increased the likelihood of receiving a referral. These beliefs were more common with longer symptom duration, prior imaging, and lower education (Jenkins et al., 2022). Together, this evidence shows how patient expectations—seeking reassurance or proof of a “real” problem—can nudge clinicians toward imaging and contribute to potential MRI overuse when access is easy and perceived risks are low. Within the global overuse debate, this phenomenon is part of a broader "more is better" mentality in medicine that is driving utilization. Over-medicalization and the expanding definition of disease also fuel patient demand for testing, as people increasingly seek early diagnosis for peace of mind (Brownlee et al., 2017, p. 162).

In the Turkish healthcare context, high patient expectations likely play a significant role in MRI use. Over the past two decades, Türkiye's healthcare reforms have significantly improved access, creating a culture where patients can consult specialists directly without the need for primary care gatekeeping. This self-referral system (Başar et al., 2019) means that if a patient requests an MRI, they can potentially "shop around" until a doctor orders it. Importantly, MRIs in public hospitals are free to the patient, removing any financial disincentive to request or accept a scan. Clinicians in Türkiye have noted that patient satisfaction is a key criterion, and refusing a requested test can carry the risk of a complaint. Furthermore, Türkiye's significant increase in imaging capacity has conditioned both doctors and patients to trust scans. Agarwal et al. (2010) noted that as imaging has become more accessible, clinicians have become accustomed to using it as part of their routine work. A study conducted in Türkiye found that the majority of participants believed an MRI was harmless and should be performed *'just to be sure,'* illustrating this widespread mindset. In addition, Kayadibi et al. (2021) surveyed patients referred for USG, CT, or MRI in Eastern Türkiye to assess knowledge and attitudes toward imaging. Notably, among MRI patients, nearly half of patients had had at least one test before. 42% of patients requested MRI scans from their doctors, and nearly half of the patients who had MRIs believed they could be cured with these treatments. This theme reinforces that healthcare utilization is not solely supply-driven; understanding the demand-side context in Türkiye strengthens our theoretical framework for why MRI overuse may occur.

3.4. Hypothesized Drivers of Underdiagnosis

Chapter 1 of the thesis provides an overview of the capacity constraints in the healthcare system. Based on this and existing literature, the study proposes a conceptual framework whereby capacity strains lead to higher rates of underdiagnosis. Shortly, more "unspecified" diagnoses result from a reduction in diagnostic quality caused by overworked workers or the system. Here, the study explains each hypothesis and dimensioned factor in detail:

3.4.1. Human Capacity Strain: Daily Patient Load and Auxiliary Staff Gaps

The first hypothesis of the study, based on uncertainty theory, is that constraints in human capacity, encompassing both physicians and auxiliary staff, increase the likelihood of underdiagnosis. In Türkiye, physicians have less time for each case due to the large number of patients they see throughout the day and, therefore, are more likely to complete examinations without reaching a definitive diagnosis. This might result in an early closing of the case without any clarity because of rushing the examination, lack of time to wait for test results, or just cognitive fatigue. Previous research supports this concern: Overworked clinicians are likely to make more errors or omissions. For instance, research found that trainees working extremely long weeks (80+ hours) made five times more diagnostic errors than those working moderated hours (Landrigan et al., 2004). In addition, according to a recent study by Kasalak et al. (2023), radiologists who interpret 21% more work than their average workload are more likely to make diagnostic errors. On the other hand, human factors research also emphasizes fatigue and burnout as consequences of overload that impair diagnostic reasoning (Rodriguez-Jareño et al., 2014; Wise, 2022).

The availability of auxiliary staff is an integral part of human capacity and strongly shapes diagnostic success. This study argues that the influence of support staff is most visible through temporal variation in service delivery. Accordingly, hospitals often operate at limited resources during evenings, nights, or weekends; fewer staff are available, many specialized consultative services are unavailable, and certain diagnostic tests are on hold. Consequently, if a patient presents at 2 a.m. with a complex condition, comprehensive care may not be provided. This can lead to a tentative or nonspecific diagnosis pending further investigation: *an underdiagnosis*. Empirical research confirms this pattern, often described as the “weekend effect” or “after-hours effect.” For example, one study found that diagnostic errors were 12%

more common during night shifts than during day shifts in a hospital setting (Rodriguez-Jareño et al., 2014). Night-shift radiologists have been shown to make more interpretation errors due to fatigue, particularly during the second half of the night shift (Hanna et al., 2018; Patel et al., 2020). Moreover, physicians working very long shifts (24+ hours) experience cognitive fatigue similar to being legally intoxicated, significantly increasing the likelihood of missed diagnoses (Link & Sack, 2009). Therefore, the study expects the underdiagnosis rate to increase during off-peak times, specifically during night shifts, early mornings, and possibly on weekends/holidays.

In summary, the thesis hypothesizes a positive relationship between workload strain and underdiagnosis, and it will analyse when workload exceeds a physician's normal capacity, errors and underdiagnoses increase. Thus, what the thesis seeks to prove: If a doctor sees an exceptionally large number of patients in a single day, the proportion of these visits ending in R codes should increase.

3.4.2. Technological Strain: Overutilization and Quality Challenges

The second hypothesis of the study is that overutilization of MRI diagnostic technologies might lead to inefficient diagnostic processes. The study addresses two related issues: the quantity and quality of MRI imaging technologies. First, if a hospital's imaging rates are significantly higher than the national average, this may indicate a tendency to order scans impulsively, without careful clinical correlation. This excessive reliance on imaging may lead clinicians to stray from clinical reasoning and issue R codes when scans do not provide an answer. The second factor is imaging quality: hurried scans, insufficient procedures, or bad interpretation might lead to inconclusive studies that don't help with diagnosis. In such cases, patients may undergo repeat scans because the first image was interpreted incorrectly or unclearly.

From this point of view, the study suspects that hospitals with high repeat imaging rates or very high scan counts per machine will have higher rates of underdiagnosis

The literature on overutilization of diagnostic technologies supports this view: an estimated 20–50% of advanced imaging in OECD countries may be clinically inappropriate or of low value (Hendee et al., 2010; OECD, 2017; Kjelle et al., 2022). Such low-throughput imaging can delay true diagnosis by diverting time and resources or producing incidental findings that confuse the picture (Hendee et al., 2010).

Another study found that indiscriminate MRI scans for low back pain often find abnormalities unrelated to pain, which does not bring the clinician any closer to a true answer (Brinjikji et al., 2015). Therefore, more is not always better, and in this context, we hypothesize that extremely high MRI utilization may be associated with more underdiagnosis, not less, especially if some of these MRIs are repeated scans due to indeterminate results. The study will then use the hospital's repeat scan rate, an indicator of "low-quality imaging practice," as a proxy. In summary, the technological capacity strain hypothesis is that diagnostic inefficiencies or overutilization may paradoxically correlate with a higher rate of unresolved diagnoses.

3.4.3. Institutional Strain: Hospital Tier and Resource Constraints

The first part of the thesis classified hospital roles according to tiers for the Turkish health system. Another hypothesis of the thesis assumes that lower-tier hospitals will have higher underdiagnosis rates due to resource constraints. Accordingly, a small regional hospital, such as Type D *(Basic District Hospital*) or Type E *(Integrated District Hospital)*, may not have on-site specialists and may have limited diagnostic infrastructure. In such settings, when a complex patient arrives, the physician may reach the limit of their capabilities and resort to an R code with the plan to "refer to a higher-level facility" or request follow-up elsewhere.

Previous findings suggest significant quality differences in diagnostic capabilities between small community hospitals and larger tertiary care centres (OECD, 2017). For example, the OECD (2017) noted that a lack of specialized training in primary care in mental health services in Japan led to the underdiagnosis of mild to moderate disorders until reforms were implemented. This exemplifies how limited expertise at lower levels contributes to the underdiagnosis of conditions. The data in the study are predicted to show similar patterns: Type C/D/E hospitals with smaller capacity will exhibit higher R code frequencies than Type A university or research hospitals. Accordingly, underdiagnosis should show a decreasing trend as hospital capacity and resources increase. This hypothesis aligns with the intuitive idea that capacity involves human expertise and equipment: where these are lacking, accurate diagnoses are more difficult to achieve.

3.4.4. Patient Characteristics: Age, Gender, Severity of Disease

While the focus of the study is on health system factors, it is also hypothesized that certain patient characteristics may impact diagnostic complexity and, consequently, underdiagnosis. Older patients with multiple comorbidities may present with vague, overlapping symptoms, making it difficult to establish a single, unifying diagnosis (Marengoni et al., 2011). Given that complex disease symptoms are associated with age and disease severity, these factors are hypothesized to potentially increase underdiagnosis because older patients or more complex cases are inherently more difficult to diagnose.

According to some studies, women may actually be underdiagnosed for specific diseases: for example, cardiovascular disease in women is often missed early because symptoms are atypical or nonspecific, leading to delays in diagnosis (Keteepe-Arachi, 2017; Johnson et al., 2021). However, in the context of the study, preliminary data-based findings suggest that being female is associated with a lower likelihood of being separated by R code. Patient gender will be treated as a control variable here, and its causality will not be explored in depth. The key is to be able to adjust for patient case mix when examining system-level factors; thus, the conceptual model assumes that capacity constraints increase underdiagnosis rates when holding constant patient differences. This sets the stage for our empirical tests.

Overall, under diagnostic capacity factors and theoretical framework, several systemic factors contribute to the intensity of diagnostic test use in Türkiye. This is reported in the Turkish Radiology Association's report as: First, the limited number of physicians relative to patient demand reduces the time available for adequate clinical examination. Second, pressures to ensure patient satisfaction and meet rising expectations regarding healthcare services encourage higher rates of imaging. Third, the number of radiologists per patient remains insufficient, placing additional strain on diagnostic processes. Fourth, excessive, uncontrollable, and often irrefutable patient demands for imaging studies lead to utilization even in cases where such tests are not medically necessary. Finally, defensive medicine practices play a critical role: physicians increasingly rely on imaging methods perceived as rapid and reliable substitutes for clinical examination, partly to mitigate the risk of patient dissatisfaction or even violence in the event of perceived diagnostic shortcomings (Turkish Radiology Association, 2018, p.5). Taken together, these drivers of test intensity align closely with the hypotheses of the study. Under the constraints on human capacity, institutional dynamics, and technological reliance echo both the theoretical arguments in the literature. In the subsequent chapters, these dynamics will be empirically tested to demonstrate how diagnostic capacity factors shape underdiagnosis and uncertainty within the Turkish healthcare system.

CHAPTER 4

DATA AND METHODOLOGY

4.1. Data Sources and Description

The empirical analysis in this thesis relies on three complementary datasets obtained from the Ministry of Health (MoH), Republic of Türkiye, under official access agreements. First, the *MoH Hospital Utilization Data (2012–2016) – Selected Nationals Subset* is a patient-level dataset covering reimbursed hospital visits with detailed information on demographics, application dates, hospital and physician identifiers, ICD-10 diagnostic codes, and procedures, enabling longitudinal analysis of diagnostic transitions and underdiagnosis. Second, the *MOH Hospital Utilization Data (2019–2023) – Selected Nationals Subset* provides hospital-level information on imaging utilization (MRI, CT, USG), allowing the assessment of recent trends in diagnostic capacity and MRI efficiency. Third, a *2019–2023 patient-level MRI dataset* of approximately 50,000 applications is employed to investigate repeat MRI behaviour, diagnostic quality, and patient heterogeneity in greater depth. Together, these datasets capture both historical patient-level dynamics and contemporary hospital- and patient-level utilization patterns, forming the empirical foundation of the thesis.

4.1.1 Patient-Level Hospital Data (2015–2016)

The primary dataset is a comprehensive dataset of patient-level hospital records from 2015–2016. The dataset contains 2,648,448 individual hospital visit observations and covers 577 different healthcare facilities, including public and private hospitals with different administrative types and roles. Each record corresponds to a patient visit and contains detailed information about patient demographics, admission information, diagnoses, procedures, and outcomes.

Patient Profile: Patients ranged in age from 0 to 98, with a mean age of around 40. There is a slight female predominance in the sample, with approximately 65% of visits made by women. Most admissions were relatively short: the average length of stay was approximately 4.5 days, suggesting that many hospital visits were brief; a small number had very long stays, up to a maximum of 516 days.

Diagnoses and Outcomes: Each record contains an ICD-10 diagnosis code for the patient's condition. For the analysis of underdiagnosis, a certain proportion of cases were coded in the range R00–R99. In the dataset, 7.5% of hospitalizations end with an R-code, suggesting a nonspecific or symptom-based outcome and thus serving as a potential proxy variable for underdiagnosis in the study. Specific diagnostic codes are present in about 92.5% of the remaining cases. The study defines a binary variable R_Code_Prob 1 if the diagnosis is an R code, 0 otherwise, to identify cases of underdiagnosis.

Treatments and Procedures: The data also contains rich information about the procedures and services performed during each hospital stay. Each aspiration has a unique procedure name, and these procedures are categorized into procedures such as MRI scans, CT scans, laboratory tests, surgeries, etc. This information is used to derive variables such as whether the patient received an MRI scan during their visit. As a result, '*Repeat MRI Same Day'* is created as a key variable for the research. It was set to 1 if the patient had multiple MRIs done on the same day during a single admission, and 0 otherwise. This captures situations where an MRI exam needs to be repeated on short notice.

Hospital and Doctor Details: Each record includes details on the related physician and is connected to a specific hospital. The data details the time period in which the patient visited the hospital, including day, month, year, and time. From the time information and doctor IDs, the doctor's total patient visits for that day were calculated. Specifically, $doctor\ total\ visits$ is the number of patients treated by the same doctor on a given day. Although this varies greatly, a doctor in the sample sees around 75 people every day on average. While some physicians see over 100 patients a day, others see only a handful. Each entry includes hospital information as well.

For instance, $hope$ identifies the category of the hospital, while $central\ hospital$ indicates whether the hospital is a central high-level hospital. In addition, Hospital bed capacity and utilization metrics are also combined with patient-level data using general hospital-level information for 2015-2016. One significant variable here is $Bed\ Turnover\ Ratio$, which measures how often beds are occupied and vacated on average. The average bed turnover ratio in the data set is approximately 87 per year. The data also include $Referral\ Degree$ indicator, which reflects the hospital's degree of referral. In practice, $Referral\ Degree$ is calculated as the percentage of patients arriving from outside the province or region where the hospital is located; higher values indicate that the hospital serves as a regional referral centre. In this sample, the referral index ranges from approximately 0% for local hospitals to approximately 100%, highly specialized centres, with an average of around 30%.

The 2015-2016 patient-level dataset will form the basis of this thesis on how imaging utilization relates to diagnostic outcomes. The study will then describe additional datasets from 2019-2023, which are used to create instrumental variables and additional covariates.

4.1.2. Hospital-Level Imaging Utilization Data (2019–2023)

To capture current long-term trends and hospital-level patterns in imaging utilization, the study uses a second dataset containing hospital-level statistics from 2019 to 2023. This dataset contains 23,245 observations, each representing one hospital year. The dataset includes data on the total number of patient admissions per hospital and the total number of imaging scans, broken down by imaging technology: Magnetic Resonance Imaging (MRI) scans, Computed Tomography (CT) scans, and Ultrasound (USG) scans. Key variables in this dataset are:

$Total\ applications$: total number of patient visits to the hospital in that year.

$Total\ MRI\ Scans$: total number of MRI examinations performed at the hospital in that year.

$Total\ CT\ Scans$: total number of CT examinations in that year.

$Total\ USG\ Scans$: total number of ultrasound examinations in that year.

$Hospital\ Type$: It indicates the type of hospital, such as primary healthcare facilities, second and third level public healthcare facilities, other private, medical center, private hospital, foundation, and university hospitals.

The study demonstrates the significant growth in MRI utilization in Türkiye and the high utilization rate per device, as shown by national statistics in Chapter 1. Here, the 2019-2023 dataset provides a macro-level, up-to-date view of imaging utilization. Accordingly, this hospital-level dataset shows these trends continuing: the total number of MRI scans is increasing annually. On average, a hospital performs approximately 15,000 MRI scans per year, but this varies depending on hospital size and type. In addition, using the above variables, the study calculates an MRI utilization rate for each hospital based on the ratio of total admissions to the number of MRI scans.

4.1.3. Patient Follow-Up MRI Data (2019–2023)

The third dataset contains a total of 50,920 observations of patient-level MRI utilization. This dataset includes patients who undergo MRI scans at the hospital between 2019 and 2023, as well as their repeat MRI examinations. For each patient, the data record includes the patient's demographic information, including age, gender, and state of residence, the hospital where the MRI was performed, and the dates of the first and second MRI, if applicable. Critically, it includes a binary variable called $repeat\ in\ 90\ days$ which indicates whether the patient had a repeat MRI within 90 days of the first MRI. In other words, $repeat\ in\ 90\ days = 1$ if the patient's second MRI occurred within 3 months of the first MRI, and 0 otherwise. The dataset also counts the number of MRI procedures for that patient during that period.

This dataset allows the study to quantify how frequently patients require additional imaging over a short period of time. According to the data, approximately 12-13% of patients who underwent an MRI had a second MRI within 90 days. Comparing this repeat imaging rate, some studies have found repeat imaging rates in the 10-20% range for specific modalities across various countries. Repeat scans may occur for legitimate clinical reasons, but studies in the literature suggest that a significant portion are likely due to quality issues with the initial scan. The study generated a hospital-level low MRI quality variable by calculating repeat MRI rates from this dataset. This variable is then integrated into the 2015-2016 patient-level dataset.

4.2. Descriptive Statistics and Visualizations

Below, the thesis summarizes the descriptive statistics and visualizations proposed in the study for each of the three datasets. Through summary statistics, the study captures central tendencies and variations in patient characteristics, hospital types, and diagnostic practices. These descriptive insights not only illustrate the scale and structure of the datasets but also guide the subsequent econometric models by highlighting areas where diagnostic uncertainty and underdiagnosis are most likely to arise.

4.2.1. Patient-Level Hospital Data (2015–2016)

This dataset includes individual hospitalizations between 2015 and 2016, capturing whether patients underwent MRI, CT, or USG examinations, their diagnoses, and the characteristics of the hospitals where they received care. The descriptive statistics and visualizations are designed to provide a comprehensive overview of diagnostic uncertainty and imaging utilization across multiple dimensions of the data. Specifically, the analyses cover the age distribution of patients and their probabilities of receiving an R-code, highlighting potential age-related vulnerabilities to underdiagnosis. They further examine the gender distribution of patients and R-code probabilities, allowing us to assess whether diagnostic uncertainty systematically differs by gender. Revisit patterns are investigated both by gender and by hospital type,thereby illustrating how institutional and demographic factors interact with diagnostic processes. In addition, temporal distributions of admissions are linked with R-code assignments, providing insight into how day-of-week and time-of-day factors, often proxies for auxiliary staff capacity, shape diagnostic uncertainty. The section also addresses resolution patterns of R-codes and diagnostic transitions, tracing how patients initially assigned an uncertain diagnosis progress toward more definitive outcomes. Finally, the descriptive analyses include an exploration of same-day repeat MRI patterns, which serve as a proxy for diagnostic quality and efficiency, and provide a first indication of supplier-induced demand. Together, these descriptive statistics and visualizations establish the empirical foundation for the econometric models that follow, ensuring that patterns of diagnostic uncertainty are contextualized before causal analysis.

Firstly, Table 2 summarizes all variables used in the empirical analysis below. Accordingly, on average, hospitals recorded 97,561 applications per year, with wide variation from 1 to nearly 4.9 million, reflecting differences in hospital size. MRI and CT utilization also showed large disparities: while mean MRI scans per hospital were around 14,569 and CT scans 16,000, maxima exceeded 2.2 million. These results highlight both the expansion of diagnostic imaging capacity and the uneven distribution of services across hospitals and regions.

Table 2. Summary Statistics of Patient Level Data (2015-2016)

	(1)				
	Observations	Mean	SD	Min	Max
R_Code_Prob	2648448	0.08	0.26	0	1
Central Hospital	2648448	0.10	0.31	0	1
Day Category	2648448	1.17	0.38	1	2
Hospital Name	2648448	278.63	196.82	1	577
Patient Gender	2648448	1.65	0.48	1	2
Patient Age	2648448	39.67	21.62	2	110
Hospital Type	2648448	2.11	1.29	1	8
Number of Beds	1861167	595.64	254.78	50	1627
Number of MRI Devices	1861167	1.37	0.52	1	4
Number of MRI Scans	1861167	42002.76	29096.56	123	122436
MRI Scans per Machine	1861167	28957.96	17111.35	123	72898
Number of CT Devices	2005012	1.80	0.94	0	5
Number of CT Scans	2005012	54855.33	37184.68	0	208457
CT Scans per Machine	2003455	30569.34	19450.61	0	89166
Number of MAM Devices	1830768	1.25	0.68	0	4
Number of MAM Scans	1830768	5850.55	5417.11	0	38532
Table 2 (continued)					
MAM Scans per Machine	1813904	2000.28	4399.30	0	23695
Time Category	2648448	0.93	0.82	0	2
Doctor's total daily visits	2648448	74.82	63.00	1	568
Bed Turnover Rate	2464396	87.23	34.67	0	278

Table 2. (continued)

ICD severity weight	2648448	38.99	59.35	-23	395
Severity Score	2648448	43.33	62.47	-23	558
Low MRI Quality	1399301	0.79	0.40	0	1
Referral Degree	797457	29.69	24.58	0	100
Repeat MRI same day	2648448	0.01	0.11	0	1
Observations	2648448				

Figure 10 displays the overall age distribution of patients in the 2015–2016 dataset. More than half of all encounters are concentrated among young adults aged 18–39 (52.3%). Middle-aged adults (40–64) represent 23.3% of the sample, while the elderly (65+) account for 16.1% and children (0–17) for 8.3%. This composition reflects the demographic profile of health service utilization, with younger adults generating the largest share of visits.

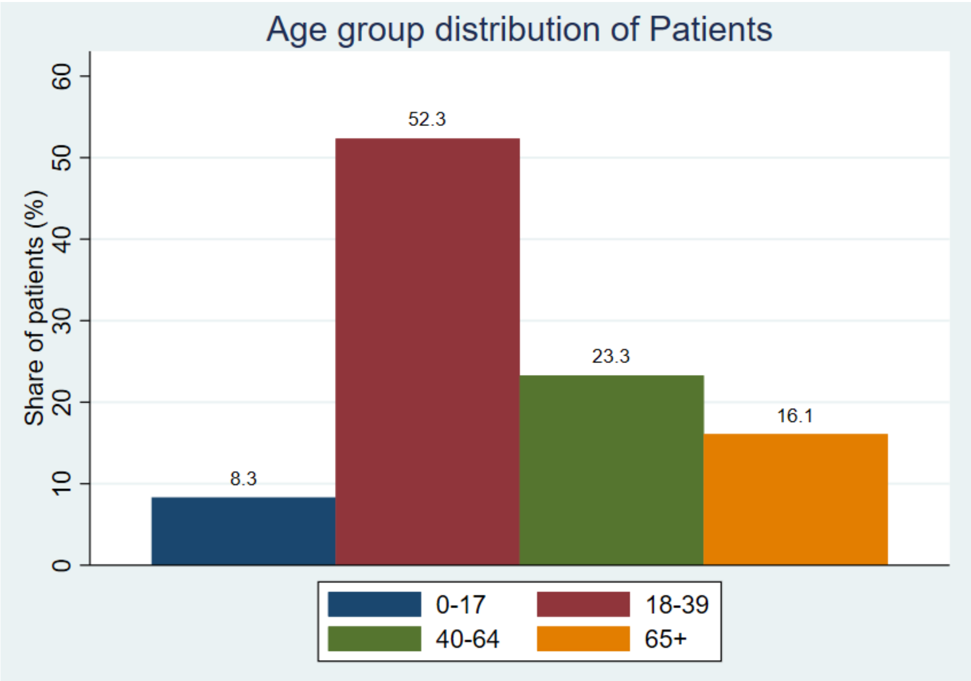

Figure 10. Patient Age Groups (2015-2016)

Figure 11 presents the probability of receiving an R-code within each age group. Here, it can be seen that diagnostic uncertainty is not evenly distributed across age. Both the youngest (0–17) and oldest (60+) groups face the highest R-code probabilities, at 11.1% and 11.4% respectively. By contrast, young adults (18–39) have the lowest risk, with only 5.0% of encounters coded as R, while middle-aged patients (40–59) occupy an intermediate position at 9.0%. This U-shaped distribution underscores that diagnostic uncertainty is concentrated at the extremes of the age spectrum.

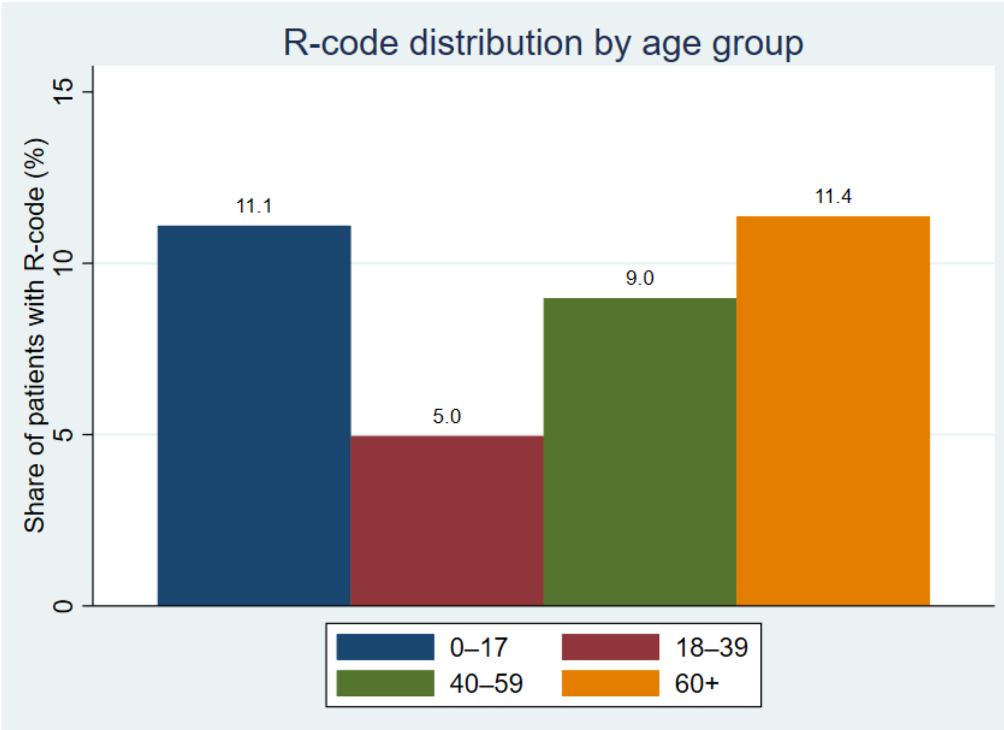

Figure 11. R-Code Distribution by Age Group

The juxtaposition of Figures 10 and 11 highlights an important distinction between capacity and risk. Although young adults dominate overall patient volume, their likelihood of receiving an R-code is relatively low. Conversely, pediatric and geriatric patients contribute fewer encounters overall but are significantly more likely to experience diagnostic uncertainty. Approximate decomposition shows that, in absolute terms, young adults still account for a substantial share of all R-coded encounters due to their large numbers, but the burden of risk is disproportionately borne by children and the elderly.

Figure 12 shows the overall gender distribution of patients in the 2015–2016 dataset. Female patients constitute nearly two-thirds of all encounters (64.9%), while male patients account for 35.1%. This imbalance reflects a well-documented pattern in healthcare utilization, whereby women make more frequent use of medical services due to reproductive health needs, preventive care, and greater health-seeking behaviour.

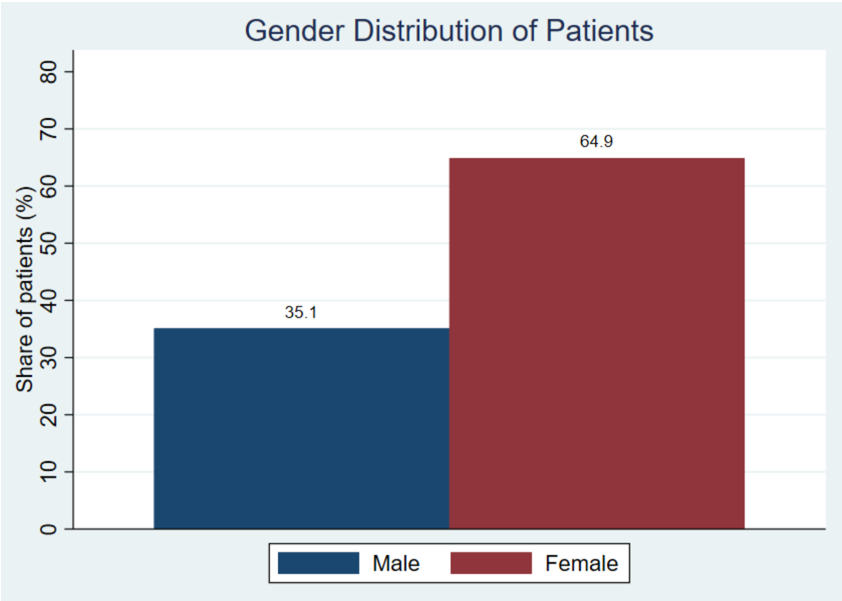

Figure 12. Gender Distribution of Patients

Figure 13 presents the share of patients with R-codes by gender, capturing the distribution of diagnostic uncertainty. Despite their lower representation in the dataset, male patients exhibit a markedly higher probability of receiving an R-code (10.6%) compared to female patients (5.9%). This finding indicates that diagnostic uncertainty disproportionately affects men, even though women dominate the overall patient population. Prior research suggests that men are more likely to delay seeking healthcare, which often results in more advanced or ambiguous symptom presentations at the point of care (Galdas et al., 2005; Courtenay, 2000). This pattern may contribute to higher rates of R-code assignment, as physicians face greater difficulty categorizing atypical or late-stage symptoms. At the same time, women—despite higher service utilization—are often overrepresented in certain diagnostic categories, such as musculoskeletal or mental health conditions, due to both biological and socialized care-seeking patterns (Kuhlmann & Annandale, 2012). From a systemic perspective, this gender gap highlights the importance of incorporating patient heterogeneity into evaluations of diagnostic performance and designing interventions that address both underdiagnosis in men and misclassification risks in women.

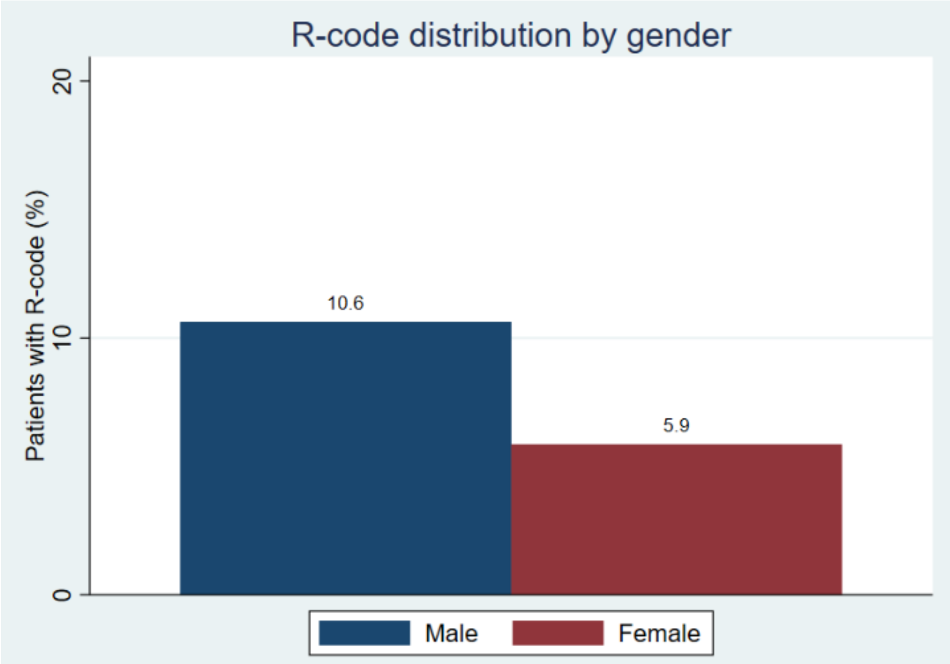

Figure 13. R-Code Distribution by Gender

Figure 14 shows the likelihood of patients revisiting after their first encounter with any doctor. Male patients are more likely to return, with a revisit probability of 10.8%, compared to 7.7% among female patients. Although women constitute the majority of overall patients in the dataset, men are more likely to return shortly after their initial encounter. This suggests that men, despite their lower baseline utilization, generate relatively greater short-term repeat demand on the system. One possible explanation is that men's health-seeking behavior tends to be more reactive and problem-driven, meaning that when they do engage with the healthcare system, they may present with conditions that require additional follow-up (Galdas et al., 2005; O'Brien et al., 2005). Another interpretation is that diagnostic uncertainty at the first encounter, as reflected by higher male R-code rates (Figure 13), could lead to greater revisit probabilities. This reinforces the link between diagnostic ambiguity and downstream healthcare demand, showing how underdiagnosis or unclear initial evaluations can manifest in short-term system re-entries. Furthermore, gendered communication dynamics between physicians and patients may contribute to differences in follow-up visits, with studies suggesting that male patients often receive less detailed consultation and reassurance, increasing the likelihood of return visits (Tannen, 1990; Verbrugge, 1985).

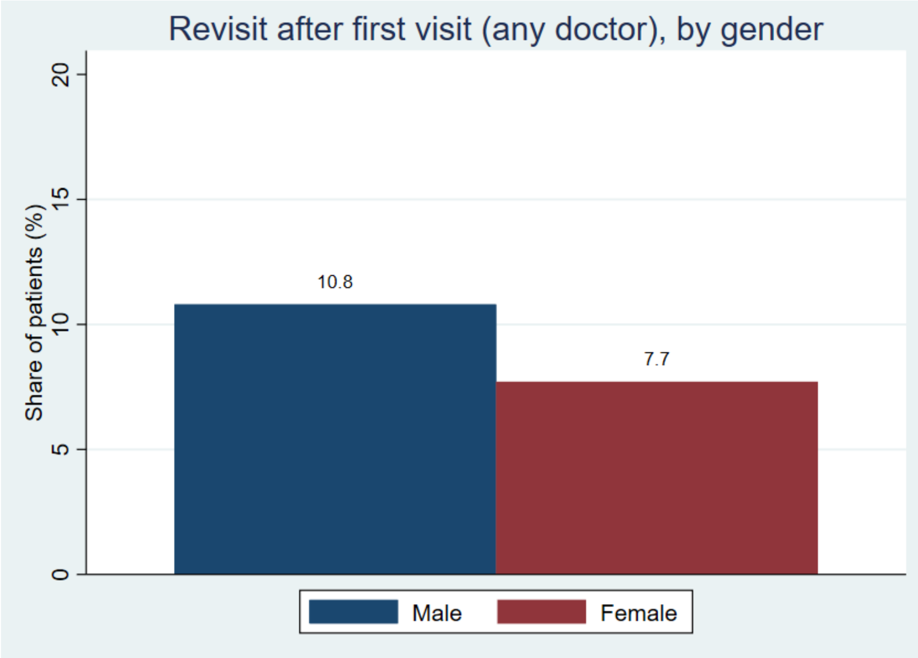

Figure 14. Revisit after first visit by gender

Figure 15 illustrates the distribution of revisit counts by gender. A large majority of both male and female patients did not revisit within the observation period, accounting for 76.7% of men and 81.3% of women. Among those who did revisit, however, gendered differences are evident. Men are more likely to return once (14.6% vs. 10.9% among women) or two to three times (5.3% vs. 3.6%). By contrast, women are slightly more likely than men to return four or more times (4.2% compared to 3.3%). These patterns suggest that men's follow-up demand is more concentrated in the short term, whereas women—although less likely to revisit initially—generate a subset of high-frequency repeat users. This aligns with broader evidence that women's greater health consciousness and chronic condition management often translate into more sustained healthcare engagement (Verbrugge, 1985; Bertakis et al., 2000). Conversely, the higher one-off or moderate revisits among men may reflect unresolved diagnostic uncertainty from initial encounters, as indicated by their higher R-code probabilities (Figure 13), driving short-term returns rather than long-term continuity of care. The gender divergence in revisit intensity underscores the importance of differentiating between short-term diagnostic clarification demand and longer-term chronic or preventive care demand when evaluating system strain (Kuhlmann & Annandale, 2012).

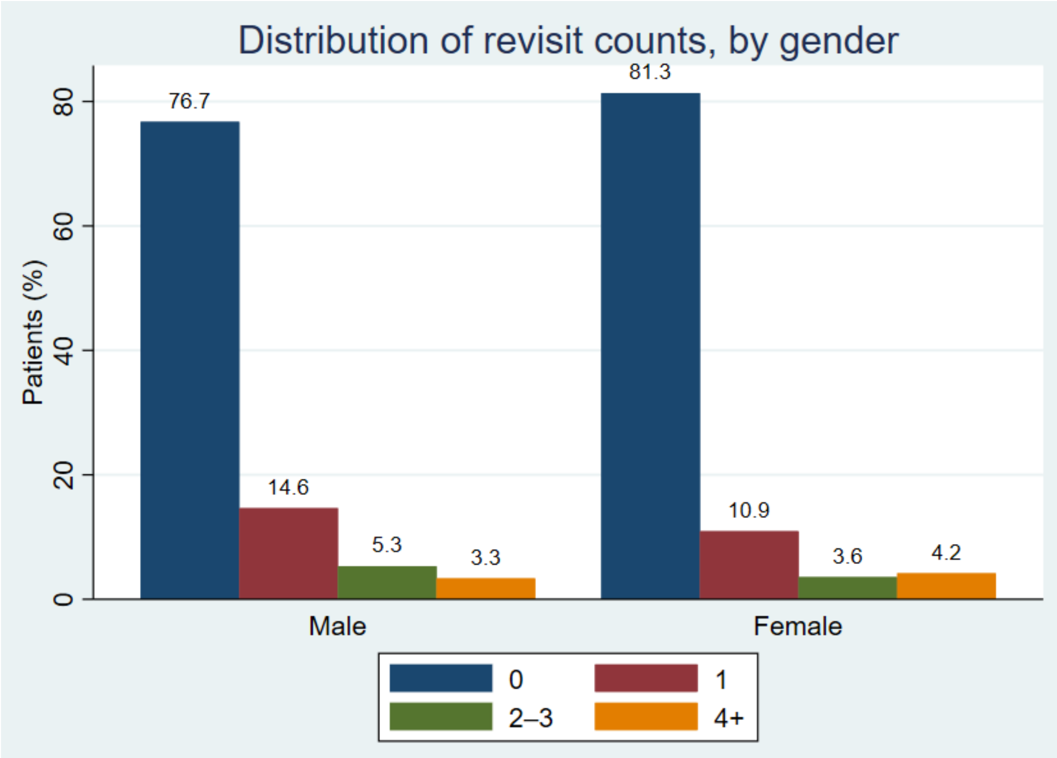

Figure 15. *Distribution of revisit counts, by gender*

Figure 16 further demonstrates that, conditional on revisiting, women accumulate more repeat visits on average (3.54) than men (2.28). These results highlight a dual pattern: men are more prone to episodic, short-term repeat visits, whereas women who do revisit tend to engage in longer-term healthcare utilization.

This aligns with existing literature suggesting that women not only access healthcare more frequently but also sustain longer episodes of care due to their greater involvement in managing chronic conditions, preventive care, and reproductive health needs (Bertakis et al., 2000; Green & Pope, 1999).

By contrast, men's interactions with the healthcare system are often concentrated around acute or unresolved episodes, consistent with their higher rates of diagnostic uncertainty at first visits (Figure 13) and early repeat encounters (Figure 14). The gendered divergence between episodic versus sustained utilization underscores the importance of distinguishing between demand generated by immediate diagnostic clarification and demand linked to continuous care pathways.

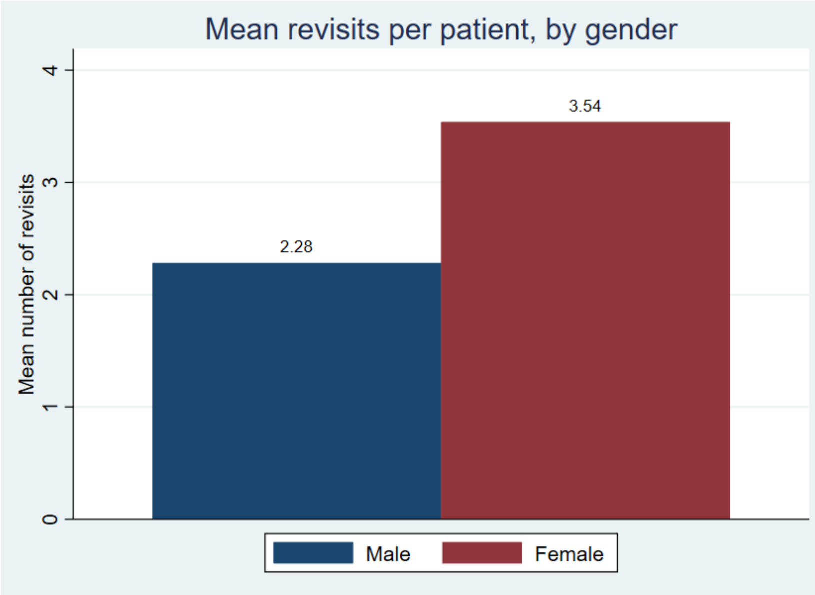

Figure 16. Mean Revisits per patient, by gender

Taken together, these findings suggest that revisit behaviour is not randomly distributed but follows gendered patterns that shape how diagnostic and institutional capacities are engaged. Men contribute disproportionately to short-term workload strain, while women contribute disproportionately to sustained demand. Both dynamics interact with human, institutional, and technological capacity, thereby influencing the extent of diagnostic uncertainty within the Turkish healthcare system.

Figure 17 shows the overall distribution of patients across hospital types. The vast majority of patient encounters take place in high-capacity hospitals: A2-type institutions (*tertiary-level general hospitals)* account for 50.7% of admissions and A1-type institutions (*teaching and research hospitals*) for 29.2%, together representing almost 80% of the dataset. By contrast, lower-tier hospitals, B (*provincial general hospitals)*, C (*district-level general hospitals)*, and D (*small district hospitals)* types, receive only 12.1%, 4.3%, and 0.8% of patients, respectively. The share of E1 (*comprehensive integrated district hospitals*), E2 (*small integrated facilities*) type hospitals and "Other" facilities is marginal. This distribution underscores the centralization of diagnostic and treatment capacity within Türkiye's health system. The concentration of patients in A1 and A2 hospitals reflects their role as referral and teaching institutions, where advanced diagnostic infrastructure and specialized physicians are predominantly located.

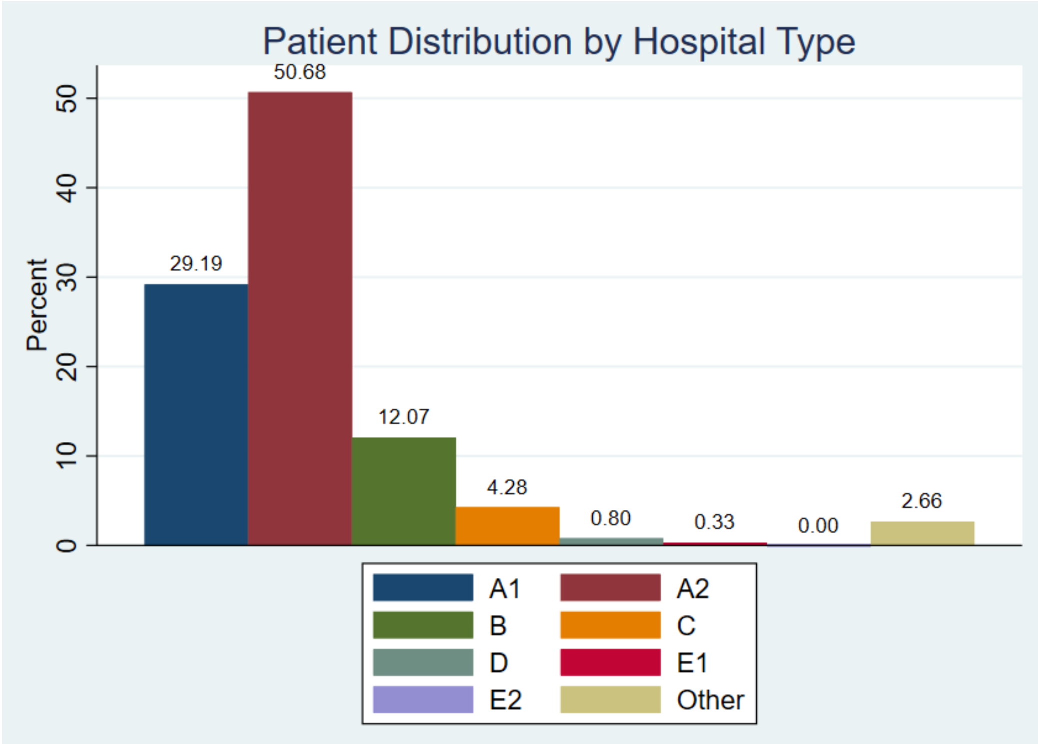

Figure 17. Patient Distribution by Hospital Type

Figure 18 presents the distribution of R-code diagnoses by hospital type. A2 Type hospitals, despite hosting the largest share of patients, have the lowest rate of R-code assignment (6.2%), while A1 hospitals record a similar but slightly higher rate (9.0%). By contrast, smaller and peripheral institutions demonstrate much higher levels of diagnostic uncertainty: D-type hospitals (19.3%) and E1/E2 hospitals (28–29%) show disproportionately high R-code shares relative to their patient volume.

This gradient suggests that institutional capacity plays a central role in diagnostic precision, with better-resourced and higher-tier hospitals able to minimize uncertainty through access to more advanced technology and specialized staff.

Conversely, the elevated uncertainty observed in lower-tier hospitals likely reflects resource constraints, limited diagnostic infrastructure, and heavier reliance on generalist physicians. These findings underline how structural inequalities across hospital types translate directly into disparities in diagnostic accuracy and patient care.

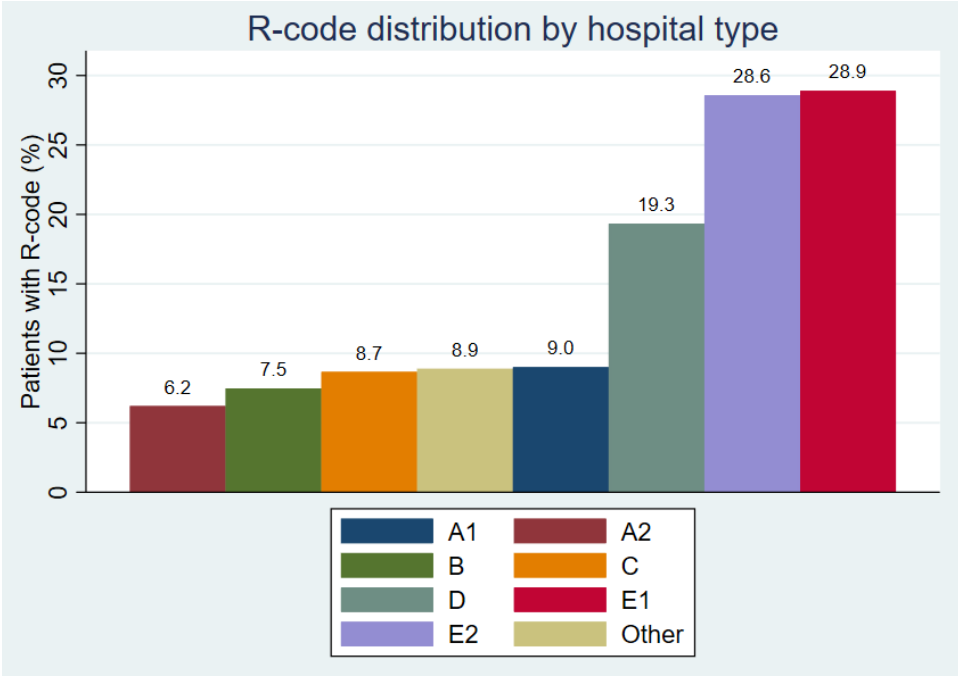

Figure 18. R-Code Distribution by Hospital Type

This divergence between patient concentration and diagnostic uncertainty underscores the uneven distribution of institutional and technological capacity. High-tier hospitals appear better equipped to manage diagnostic demand, supported by specialized physicians, auxiliary staff, and advanced imaging technologies. In contrast, peripheral or lower-tier facilities face limited diagnostic infrastructure and staff shortages, leading to greater reliance on provisional R-codes. From the perspective of human capacity, the results suggest that physician expertise and auxiliary staff support are more concentrated in A1/A2 hospitals, mitigating underdiagnosis risks despite higher patient volumes. Conversely, lower-tier hospitals, often with fewer specialists and less reliable support systems, appear more prone to diagnostic uncertainty even under smaller workloads.

In summary, these descriptive findings reveal a dual dynamic in the Turkish health system: central hospitals face volume-driven strain but achieve relatively lower diagnostic uncertainty, while peripheral institutions handle fewer patients yet exhibit disproportionately high rates of underdiagnosis. This structural imbalance motivates the thesis's econometric models, which explicitly test how human, institutional, and technological capacity interact to shape diagnostic precision across hospital settings.

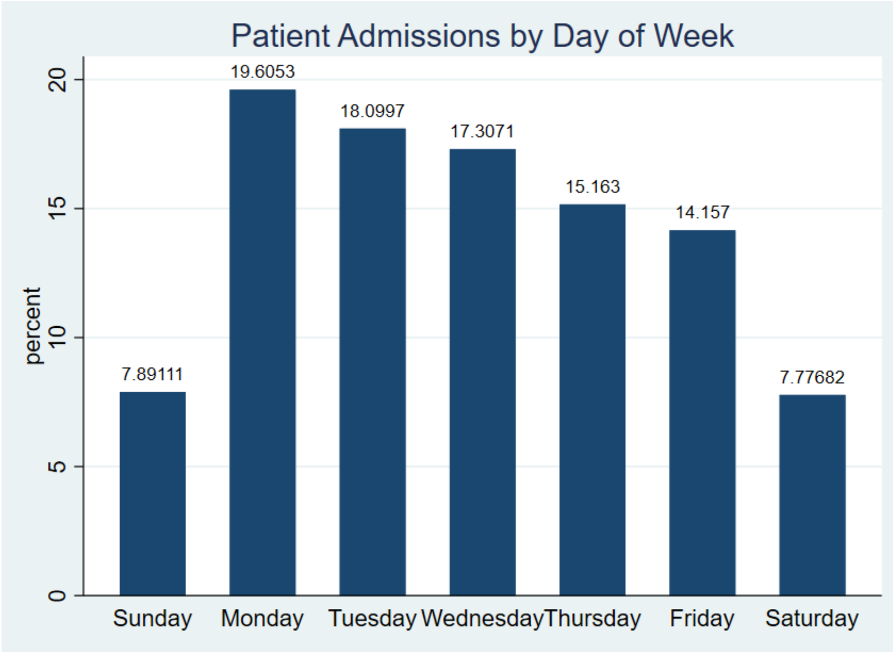

Figure 19. Patients Admissions by Day of Week

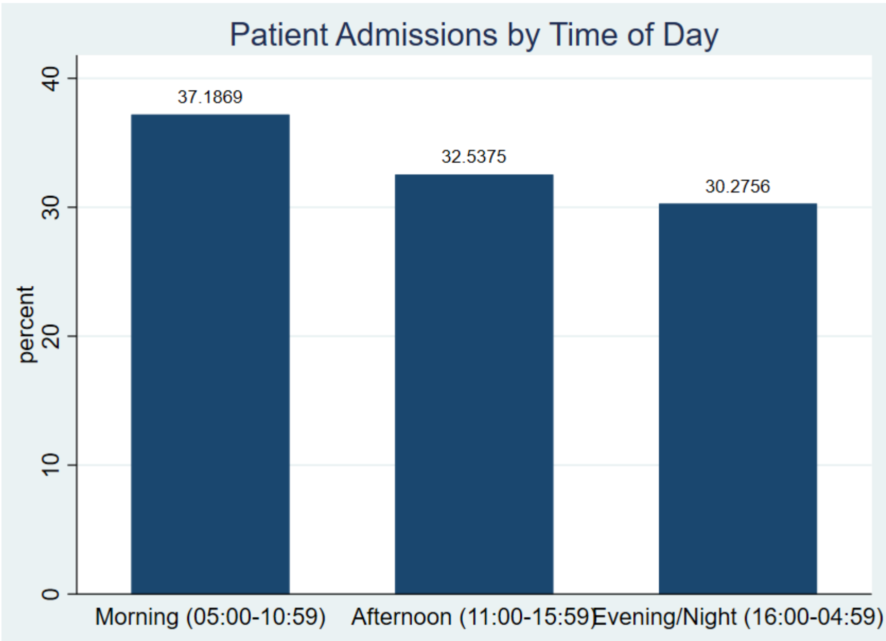

Figure 20. Patient Admissions by Time of Day

Figures 19 and 20 show that patient admissions are unevenly distributed across both the week and the day. Mondays concentrate the highest share of admissions (19.6%), followed by Tuesdays (18.1%) and Wednesdays (17.3%), with volumes tapering off toward Friday (14.2%) and dropping sharply on weekends (7.8% each on Saturday and Sunday). By time of day, admissions peak in the morning hours (37.2%) and decline progressively into the afternoon (32.5%) and evening/night (30.3%). These descriptive findings highlight systematic workload asymmetries for physicians and auxiliary staff, illustrating how human capacity constraints vary across temporal dimensions.

Figure 21 shows the overall share of patients receiving an R-code in the 2015–2016 dataset. R-codes, which represent uncertain or provisional diagnoses, account for 7.5% of all patient visits, while the remaining 92.5% received definitive diagnostic codes. Although the proportion of R-codes appears modest, its magnitude is significant in a large patient population and underscores the presence of diagnostic uncertainty within the healthcare system.

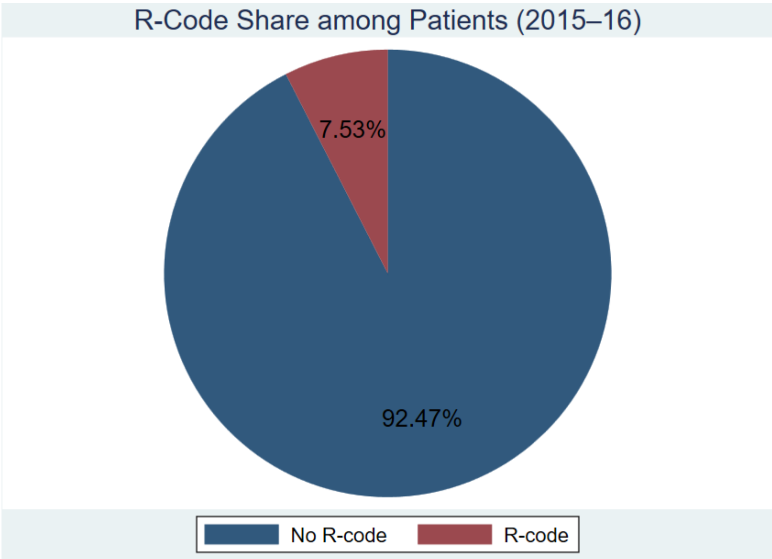

Figure 21. R-Code Shares among Patients (2015-2016)

Figures 22 and 23 examine the distribution of R-codes across time. By day of the week, R-code probabilities are lowest on Mondays (7%) and rise progressively throughout the week, reaching 9% on weekends (Sundays and Saturdays). This pattern indicates that diagnostic uncertainty is more prevalent when institutional and staffing capacities are reduced, as is the case during weekends. The midweek decline may reflect higher staffing levels and greater availability of specialist consultations, while the weekend increase suggests that constrained human capacity—in terms of both physicians and auxiliary staff—contributes to higher reliance on uncertain diagnostic classifications.

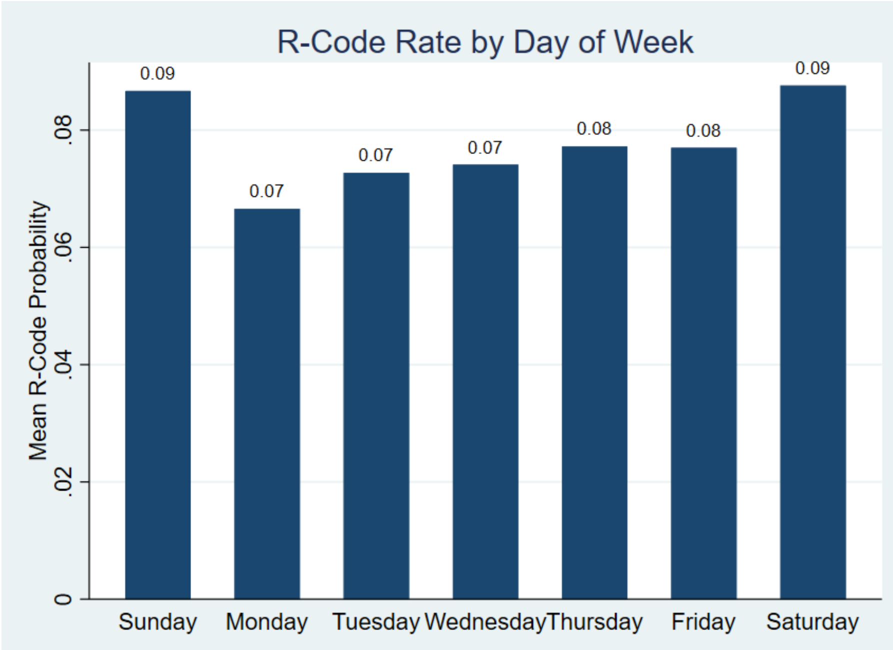

Figure 22. R-Code Rate by Day of Week

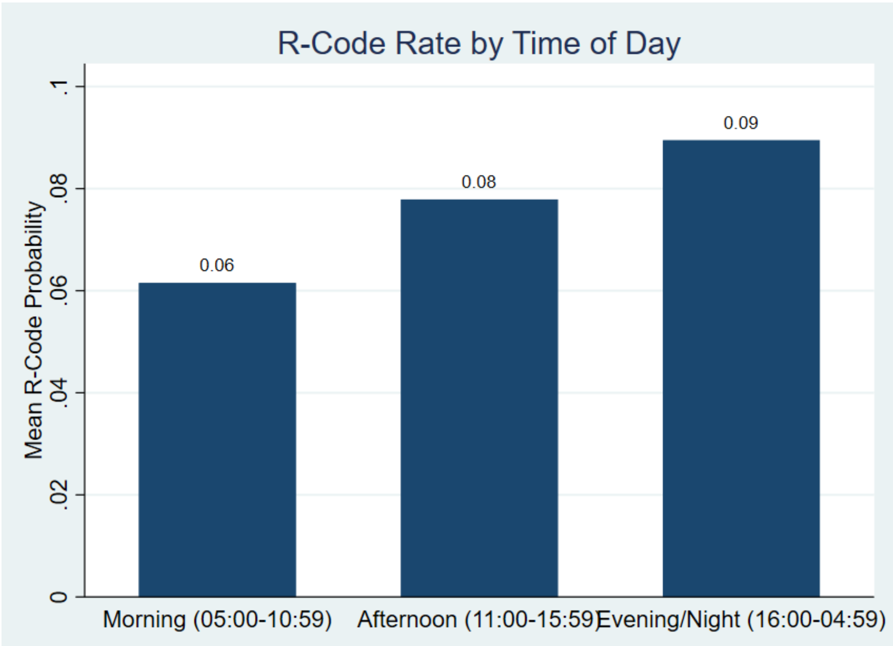

Figure 23. R-Code Rate by Time of Day

A similar pattern emerges by time of day (Figure 23). Morning admissions have the lowest R-code rate (6%), while the probability increases to 8% in the afternoon and peaks at 9% during evening and night hours. This distribution reinforces the central role of human capacity strain in shaping diagnostic quality. Physicians face rising fatigue and reduced decision-making precision as the day progresses, while support staff availability is often reduced during evening and night shifts. Together, these dynamics elevate the risk of diagnostic uncertainty at later hours.

Figure 24 shows revisit behaviour among patients initially assigned an R-code. The vast majority (98.7%) of these patients did not appear again in the dataset within the observed period, while only 1.3% revisited on a different day. Among those who revisited, 64.2% subsequently received a definitive diagnosis, whereas 35.8% were again classified under an R-code. This pattern indicates that while diagnostic uncertainty is occasionally resolved in follow-up encounters, a substantial share of revisiting patients continues to circulate within ambiguous diagnostic categories. The persistence of R-codes upon return highlights inefficiencies in the diagnostic process, where additional encounters do not always provide greater clarity. This outcome is consistent with findings in the literature that limited institutional resources, overreliance on non-specialist physicians, and variability in imaging quality can prolong diagnostic ambiguity rather than resolve it (Azari, 1996; Özen et al., 2022). At the same time, the very low overall revisit rate points to barriers in continuity of care, potentially reflecting patient drop-out, weak referral pathways, or underdeveloped follow-up mechanisms.

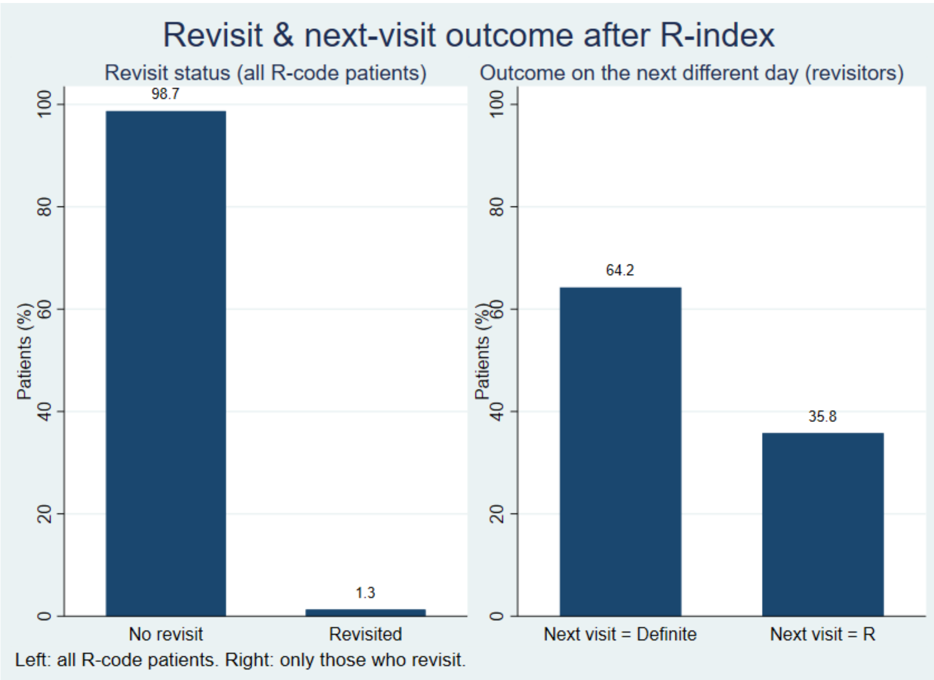

Figure 24. Revisit & Next Visit Outcome after R-Code

Figure 25 further disaggregates resolution timing. Among patients with an R-code, only a very small fraction transitioned to a definitive diagnosis within 7 days (1.5%), 8–30 days (2.5%), 31–90 days (2.3%), or even 91–180 days (1.3%). A negligible share resolved after more than 180 days (0.9%). Strikingly, over 91% of R-code cases remained unresolved within the follow-up horizon. This points to structural issues in the diagnostic pathway: either provisional codes are left uncorrected in administrative data, or patients do not return for clarification, reflecting both institutional shortcomings in follow-up protocols and human capacity strain leading to incomplete documentation. The dominance of unresolved cases also suggests that uncertainty may be "absorbed" into routine patient management, with ambiguous coding persisting rather than being actively corrected. Such persistence has been highlighted in studies showing how systemic overload, insufficient diagnostic resources, and limited incentives for coding accuracy can perpetuate provisional classifications (Baker, Atlas, & Afendulis, 2008; Smith-Bindman et al., 2008).

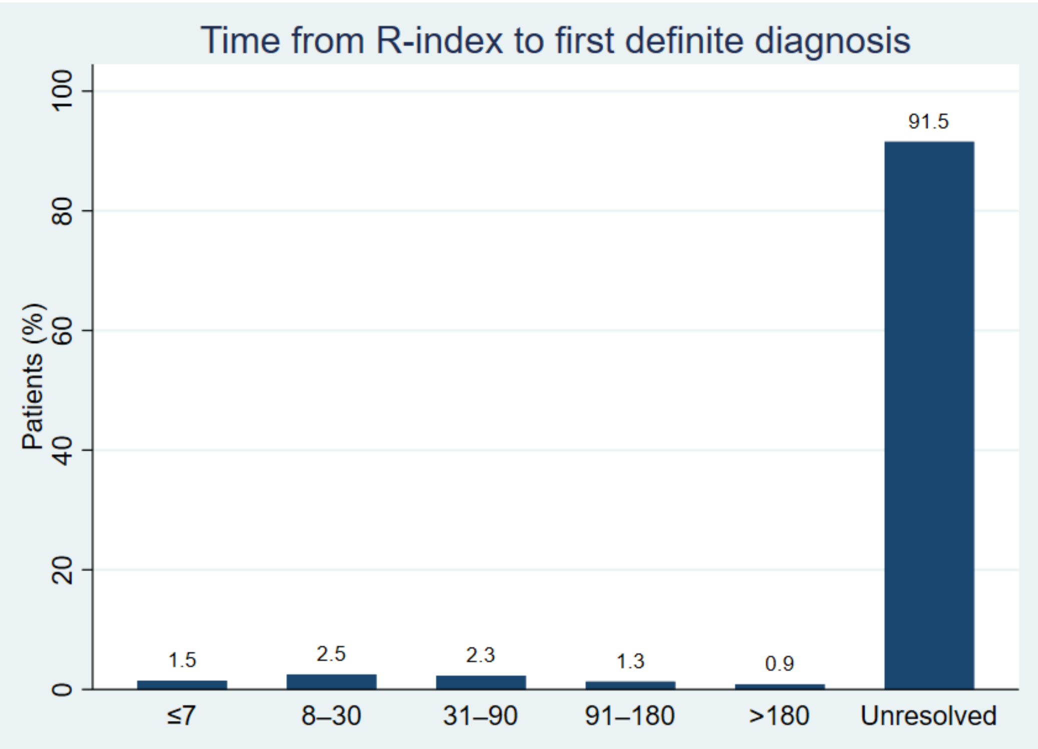

Figure 25. Time from R-index to first definite diagnosis

Figure 26 examines the ICD categories into which R-code patients transition once a definitive diagnosis is established. Among those who do convert to a non-R diagnosis, the first definite ICD chapters are concentrated in circulatory (I ≈16%) and respiratory (J ≈16%) diseases, followed by digestive (K ≈14%), factors influencing health status/encounters (Z ≈8%), genitourinary (N ≈8%), musculoskeletal (M ≈6%), pregnancy/childbirth (O ≈5%), and malignant neoplasms (C ≈5%), with smaller shares across endocrine (E), nervous (G), hematologic/benign neoplasms (D), skin (L), injury (S/T), and others. These transitions are clinically plausible, as R-codes frequently mask nonspecific symptoms (e.g., chest pain, general malaise, dizziness) that may later be clarified into chronic or acute conditions within these ICD domains. The concentration in such categories underscores the clinical complexity behind provisional coding and its link to technological capacity, as imaging and laboratory support are often required to move from symptom-based to disease-specific diagnoses.

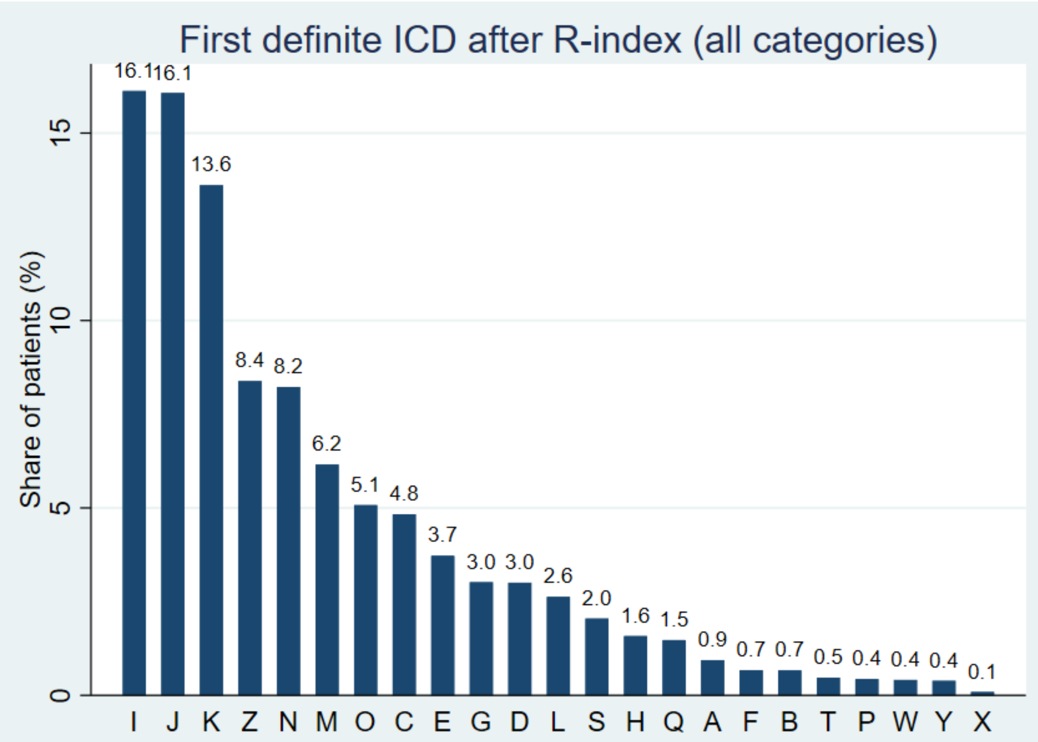

Figure 26. *First Definite ICD Code after R-Index*

In summary, these descriptive findings show that diagnostic uncertainty (R-codes) is both persistent and unevenly resolved. The extremely low revisit rates and high unresolved share highlight gaps in follow-up care and systemic inefficiencies in documentation. Where resolution does occur, it is disproportionately tied to conditions that require resource-intensive diagnostics. These insights provide an empirical foundation for the thesis's econometric modelling, in which R-code assignment is treated as an outcome of capacity strain across the human, institutional, and technological pillars.

Figure 27 presents the rate of same-day repeat MRI by hospital type. The incidence of repeat imaging is highest in the "Other" hospital category (2.9%) and in A1 hospitals (2.5%). In contrast, repeat rates are substantially lower in A2 hospitals (0.9%) and B hospitals (0.4%). This variation suggests that institutional and technological capacity are critical determinants of diagnostic quality. Higher repeat rates in A1 hospitals, large referral centers with advanced technology, may reflect both higher patient complexity and greater strain on MRI infrastructure. Conversely, lower repeat rates in A2 and B hospitals may indicate either lower patient complexity or underutilization of available imaging capacity.

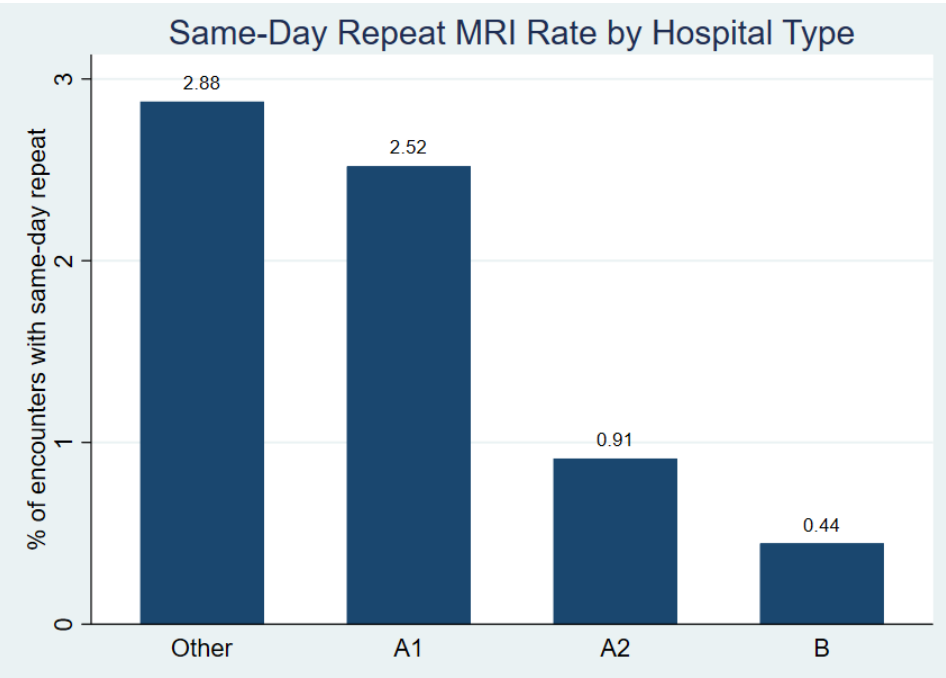

Figure 27. Same-Day Repeat MRI by Hospital Type

Figure 28 explores the diagnostic distribution of same-day repeat MRI encounters by ICD groups. The results reveal that repeat imaging is most frequently associated with circulatory system diseases (I, 15.9%), symptoms and ill-defined conditions (R, 13.2%), musculoskeletal disorders (M, 10.7%), and nervous system diseases (G, 10.5%). These categories collectively account for a substantial share of repeat scans and are typically conditions where imaging precision is vital for diagnosis. The prominence of R-codes in this distribution underscores the link between diagnostic uncertainty and repeated imaging, suggesting that same-day repeats function as a corrective mechanism in contexts of ambiguous clinical presentation.

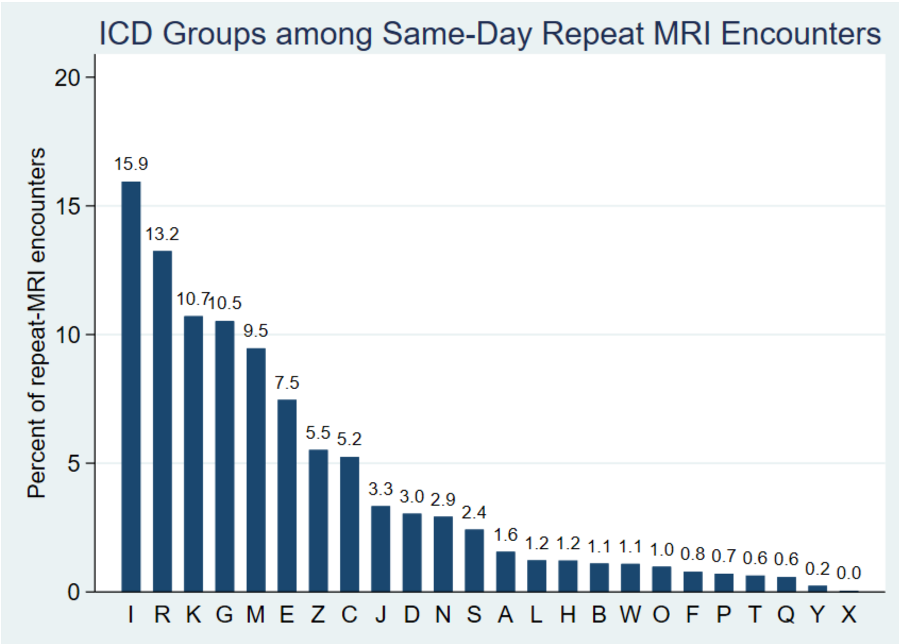

Figure 28. ICD Groups among Same-Day Repeat MRI Encounters

Taken together, these patterns highlight the interplay of technological capacity and human capacity in shaping diagnostic outcomes. High repeat rates in certain hospitals point to strains on machine quality or scheduling efficiency, while the concentration of repeats in specific ICD groups reflects clinical challenges and decision-making under uncertainty. Repeated imaging not only increases costs and resource use but also signals potential inefficiencies and risks of underdiagnosis in the diagnostic process.

4.2.2. Hospital-Level Imaging Utilization Data (2019–2023)

Table 3 reports the descriptive statistics of hospital-level imaging capacity indicators. On average, hospitals performed 14,569 MRI scans, 16,000 CT scans, and 22,000

USG exams annually, though the ranges are extremely wide, with maximum values exceeding two million for both MRI and CT. These large disparities also highlight the uneven distribution of imaging workloads across hospitals. Such variation again reflects structural differences between peripheral hospitals and central institutions, underscoring the imbalances in diagnostic capacity within the healthcare system.

Table 3. Descriptive Statistics for Hospital-Level Imaging Utilization (2019–2023)

	(1)				
	Obs	mean	sd	min	max
Year	23245	2021.533	1.457459	2019	2023
Total Applications	23245	97561.02	282668.1	1	4907701
MRI Scans	7306	14569	50447.99	1	2202343
CT Scans	7435	19223.8	61041.88	1	2437267
USG Scans	8047	31471.39	94290.24	1	2546659
Hospital Type	23245	3.777845	1.806254	1	9
Hospital ID	23245	5414.141	2509.781	1	9599
MRI Utilization Rate	7306	4.382448	9.997797	.0002172	420.2327
USG Utilization Rate	8047	8.457502	15.11001	.0003892	594.6362
CT Utilization Rate	7435	4.360801	9.695454	.0004108	340.3199
Observations	23245				

Note: Utilization Rates represent scans per 100 hospital applications

Figure 29 maps MRI utilization per 100 outpatient applications by province in 2023. Utilization is heterogeneous across space, with clusters of high and very high use concentrated in several western and northern provinces, while low to medium use is more common in parts of the interior and southeast. This spatial pattern is consistent with the concentration of A1/A2 type hospitals, which typically house greater scanner capacity and subspecialist radiology services.

From the perspective of the diagnostic capacity pillars, firstly, on the side of technological capacity, provinces with higher utilization likely have greater scanner density and longer operating hours, enabling more scans per application. Under institutional capacity, metropolitan referral networks and stronger gatekeeping/triage may channel complex cases toward high-capacity provinces, elevating MRI per application. Conversely, provinces with lower rates may rely more on alternative modalities (US/CT) or face scheduling bottlenecks that depress MRI access. For the human capacity, higher utilization provinces imply larger radiology teams and auxiliary staff, but also heightened workload intensity, which can affect reporting times and image quality.

Thus, the observed clustering is consistent with evidence of spatial accumulation of diagnostic technologies in Türkiye (Sonğur & Top, 2016) and mirrors international findings that high-capacity regions tend to disproportionately attract investment and patient flows, thereby reinforcing inequalities in diagnostic access (Huang et al., 2023).

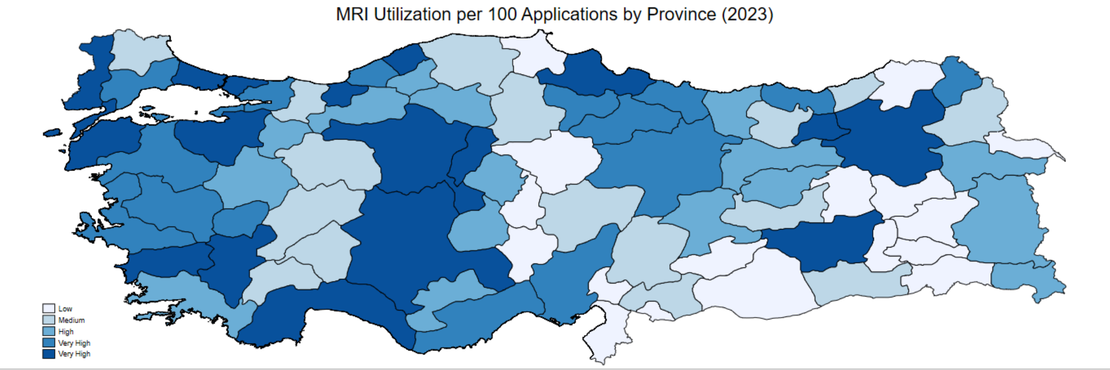

Figure 29. MRI Utilization per 100 Applications by Province (2023)

Figure 30 depicts the direction of change in MRI utilization over 2019–2023. The map reveals a fragmented picture: some provinces show increases, others decreases, and many no significant change, highlighting persistent regional gaps. These divergent paths reflect the unequal distribution of diagnostic capacity. Increases often occur in provinces catching up to national averages, while declines in higher-baseline areas point to efficiency limits and substitution pressures.

The COVID-19 period amplified these inequalities: although all provinces faced temporary contractions, recovery trajectories diverged, with stronger regions restoring volumes faster while weaker ones lagged. Overall, the evidence suggests that MRI utilization in Türkiye is shaped less by a coordinated national trajectory than by regional disparities in capacity and resilience, leaving patients' access to diagnostic certainty heavily dependent on geography and raising concerns about both equity and efficiency in the health system.

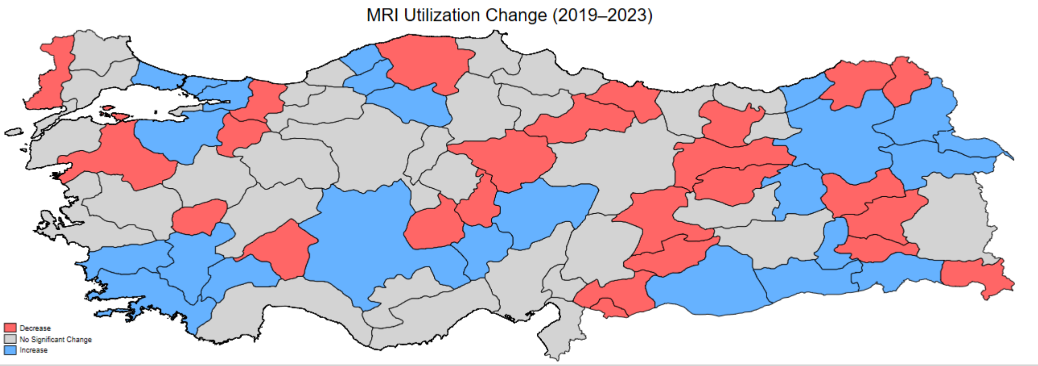

Figure 30. MRI Utilization Change (2019-2023)

Taken together, the spatial evidence underscores both concentration and divergence in MRI utilization. The 2023 map demonstrates the centralization of high MRI use in provinces with stronger technological and institutional endowments, while the 2019–2023 change map points to uneven convergence: some underserved areas expanded MRI activity, others contracted or stagnated.

These descriptive patterns motivate the hospital-level models in Chapter 4, where we examine how repeat MRI rates and diagnostic outcomes vary with (i) scanner availability and utilization intensity (technological capacity), (ii) hospital mix and provincial referral networks (institutional capacity), and (iii) staffing proxies (human capacity).

Figure 31 illustrates yearly average imaging scans per hospital across modalities (MR, CT, USG). USG remains the most widely performed modality, with average volumes ranging between 27,000 and 35,000 scans per hospital annually. CT scans show a sharp increase in 2020, likely reflecting pandemic-driven diagnostic needs, followed by a modest decline and stabilization. MRI volumes are consistently lower than CT and USG but exhibit steady growth, rising from approximately 12,000 in 2019 to nearly 19,000 in 2023. These trends confirm that MRI capacity has expanded in recent years, though it remains secondary to USG and CT in terms of raw volume.

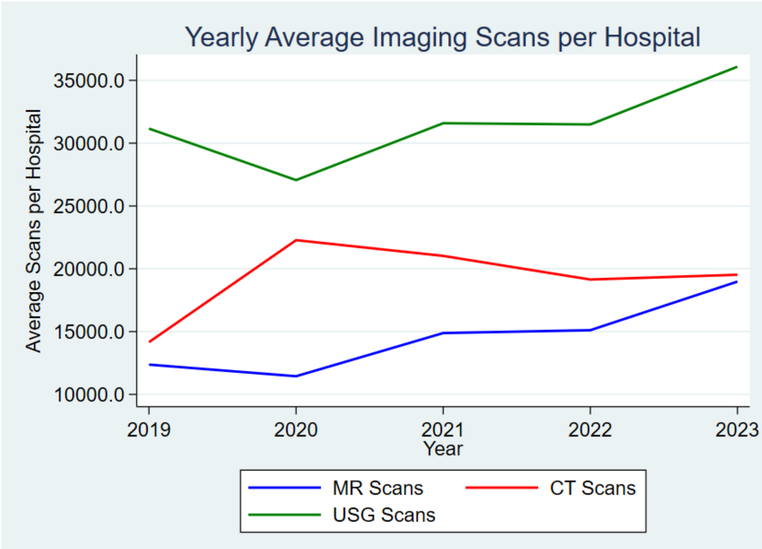

Figure 31. Yearly Average Imaging Scans per Hospital

Figure 32 depicts mean MRI utilization per hospital, measured as scans per 100 applications. Utilization increased from 3.9 in 2019 to a peak of 4.7 in 2021, dropped sharply in 2022 (pandemic backlog and reallocation effects), and then rose again to 4.8 in 2023. This trajectory demonstrates the sensitivity of diagnostic demand to external shocks, as well as the recovery of MRI use in the post-pandemic period.

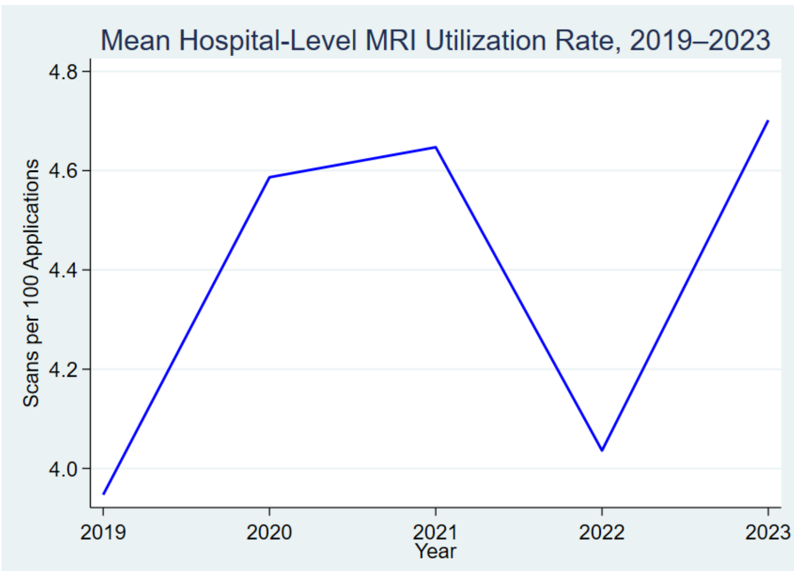

Figure 32. Mean Hospital Level MRI Utilization Rate, 2019-2023

Figure 33 disaggregates MRI utilization rates by institution group. University hospitals consistently display higher utilization, averaging 8–10 scans per 100 applications, reflecting their role as referral centres handling complex diagnostic cases. Private hospitals, medical centres, and secondary public facilities operate at much lower rates (around 3–6 scans per 100 applications), suggesting more restrained or selective MRI use. Interestingly, “Other Private” hospitals experienced a dramatic drop from 30 scans per 100 applications in 2019 to about 13 in 2021, followed by a recovery after 2022, highlighting instability in private sector utilization patterns. These differences underscore the role of institutional capacity, where ownership and governance structures directly shape MRI demand intensity.

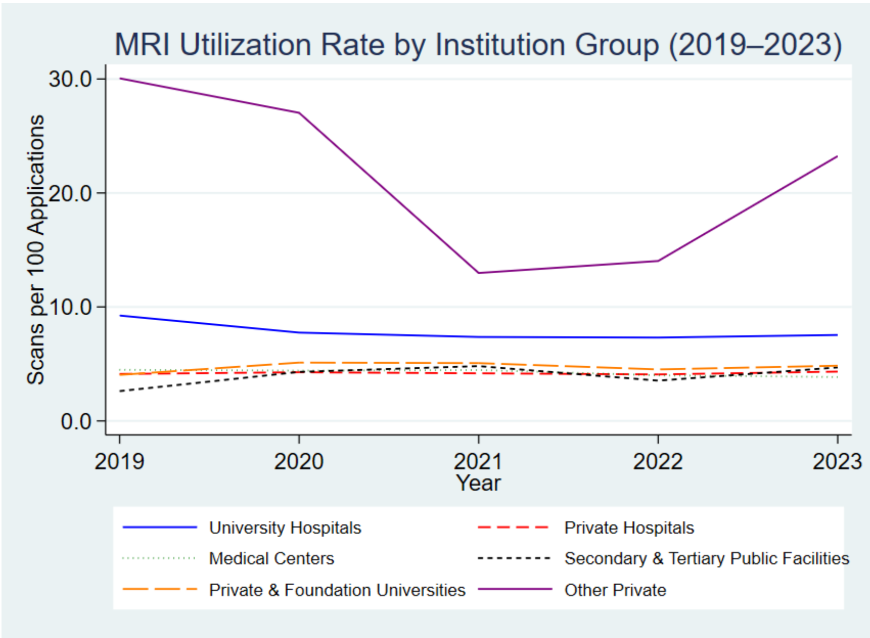

Figure 33. MRI Utilization Rate by Institution Group (2019-2023)

Figure 34 examines MRI utilization across NUTS-1 regions. East Marmara consistently reports the highest utilization, rising above 7 scans per 100 applications in 2021 and exceeding 9 by 2023. This suggests a concentration of diagnostic demand in industrialized and densely populated provinces such as Kocaeli and Sakarya, where both technological capacity and patient demand are high. By contrast, Central Anatolia, the Black Sea regions, and Southeast Anatolia maintain lower utilization levels (around 3–5 scans per 100 applications), reflecting more constrained access. Istanbul, although the largest metropolitan hub, exhibits only moderate utilization, which may be explained by higher patient volumes diluting per-application scan rates.

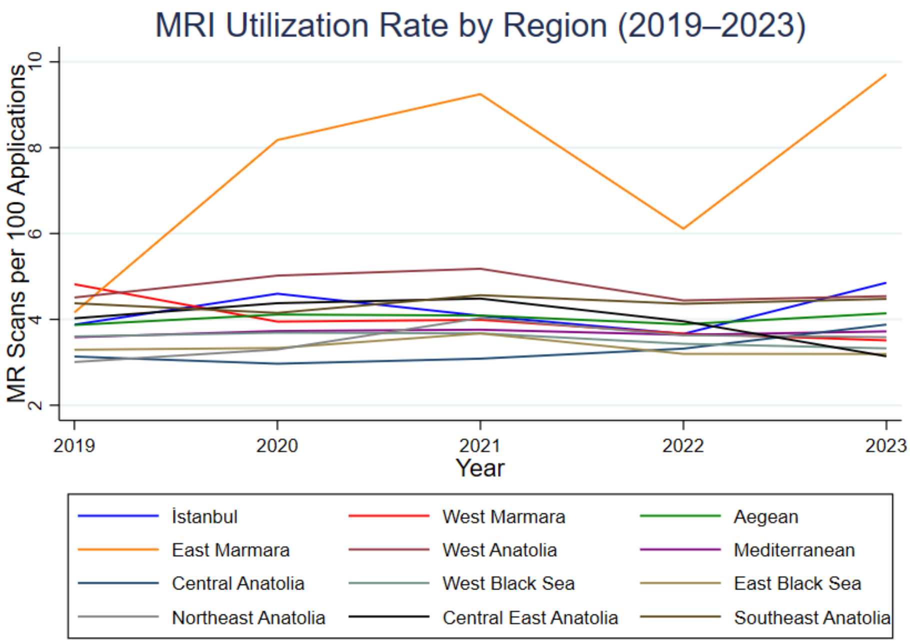

Figure 34. MRI Utilization Rate by Region (2019-2023)

Together, these descriptive findings reveal that technological capacity (machine stock and utilization intensity) and institutional structures jointly determine MRI activity across Türkiye. Regional disparities further demonstrate that access and intensity are unevenly distributed, raising equity concerns. The growth of MRI use from 2019–2023 reflects both supply-side expansion and shifting patient demand, with the pandemic leaving clear but temporary disruptions. These hospital-level patterns will be directly incorporated into subsequent econometric models, where MRI utilization per hospital serves as a proxy for technological capacity, and institutional and regional groupings are used to capture systemic heterogeneity in diagnostic performance.

4.2.3. Patient-Level MRI Imaging Utilization Data (2019–2023)

Table 4 reports descriptive statistics for a hospital–year panel spanning 2019–2023 (N=50,920). It presents patient-level MRI application data, which forms the basis for constructing the low MRI quality indicator. By tracking repeat MRI examinations within patients, the dataset allows measurement of repeat rates, serving as a proxy for imaging quality.

Table 4. Summary Statistics of Patient-Level MRI Follow Up Data (2019-2023)

	(1)				
	Observations	mean	sd	min	max
Patient Age	45297	43.66936	17.511	0	105
Repeat in 90 Days	50920	.1244501	.3300976	0	1
Number of MRI Procedures	50920	.1787313	.5666963	0	15
Hospital Type	50920	3.16469	2.657601	1	8
Patient Gender	45297	1.530388	.4990812	1	2
Year	50920	2019.453	1.026728	2019	2023
Hospital ID	50920	451.4114	322.7701	1	1181
Observations	50920				

Figure 35 illustrates the annual share of MRI patient applications between 2019 and 2023. The data are heavily concentrated in 2019, which alone accounts for nearly 80% of total observations. Subsequent years show smaller proportions, ranging from 6.7% in 2020 to 3.5% in 2023.

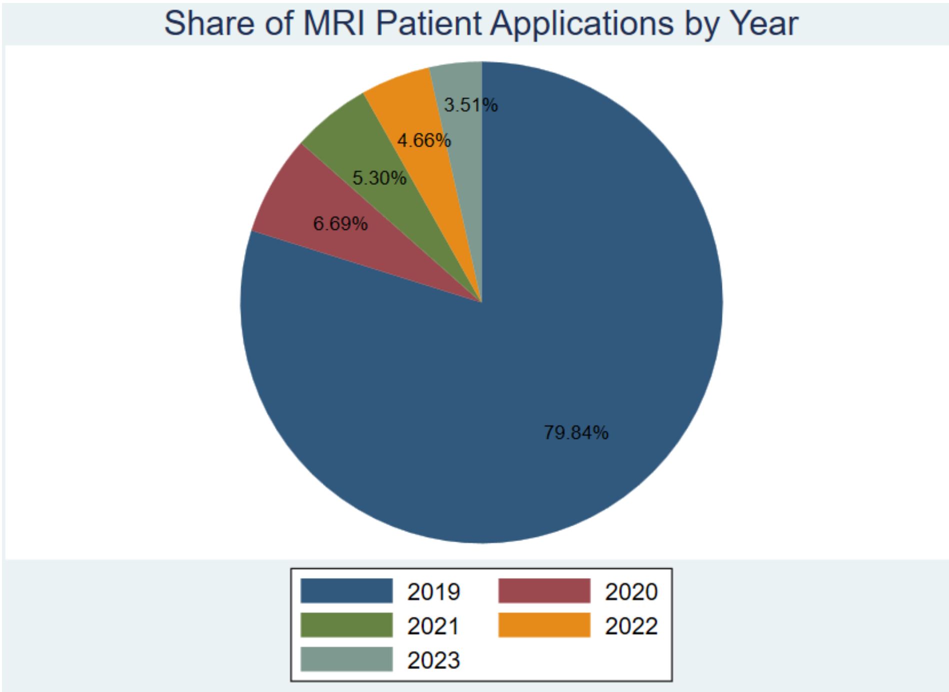

Figure 35. Share of MRI Applications by Year

Figure 36 presents the age distribution of MRI patients. The distribution follows a bell-shaped curve centred around middle adulthood, with the largest number of applications occurring between the ages of 35 and 55. This indicates that MRI demand is concentrated among working-age adults, who are most likely to present with musculoskeletal, neurological, or chronic conditions that necessitate imaging. Pediatric and elderly patients represent smaller, though still notable, proportions of total MRI users.

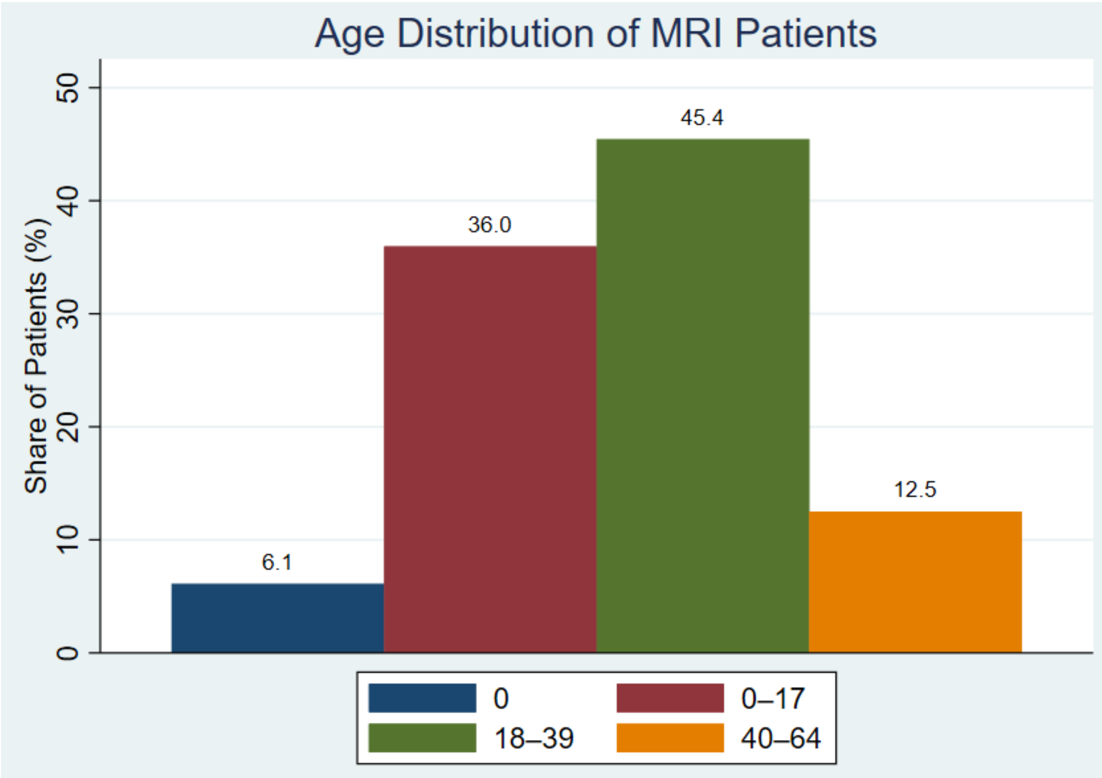

Figure 36. Age Distribution of MRI Patients

Figure 37 displays the gender distribution of MRI applications. Women account for a larger share of MRI usage than men, consistent with both international literature and the 2015–2016 patient-level findings. This pattern reflects broader gendered differences in healthcare utilization: women are more likely to seek medical care and to undergo diagnostic procedures, particularly in reproductive age groups and for chronic conditions.

The gender imbalance in MRI usage has implications for institutional capacity, as healthcare facilities must ensure equitable access and appropriate referral practices across genders. It also suggests that diagnostic demand is shaped not only by medical need but also by patterns of health-seeking behavior and physician referral practices.

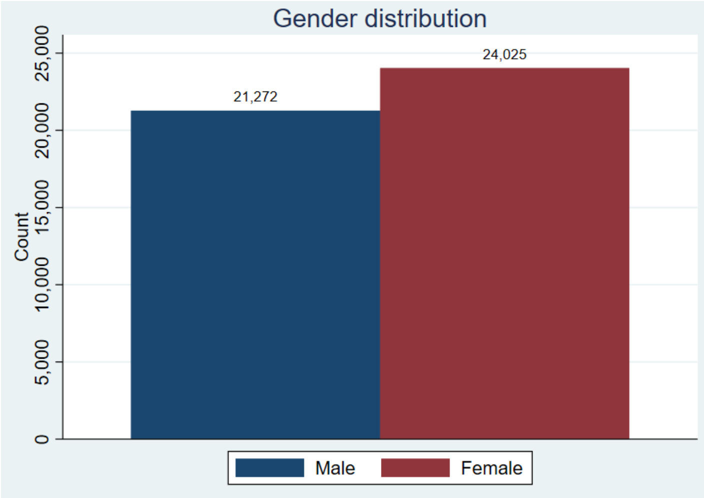

Figure 37. Gender Distribution of MRI Patients

Figure 38 reports the distribution of MRI patient counts by hospital type. Secondary and tertiary public hospitals account for the largest share of MRI patients, exceeding 30,000 cases in the dataset. Private hospitals follow with substantial volumes (around 15,000 cases), while university hospitals contribute a moderate number of patients. By contrast, medical centers, private and foundation universities, and other private institutions serve only a small fraction of MRI patients. This concentration of diagnostic imaging within public referral hospitals underscores the institutional capacity pillar, reflecting the central role of large-scale public hospitals in delivering advanced diagnostics. Private hospitals provide a complementary role, catering to a significant patient base, whereas specialized or smaller providers appear marginal in terms of patient volume.

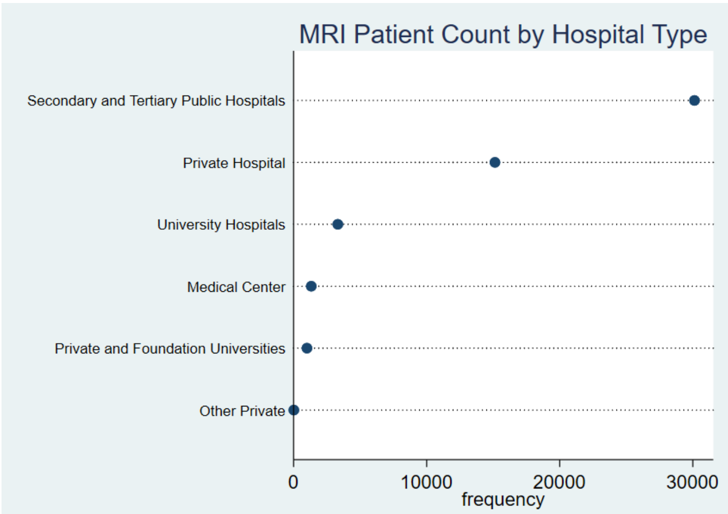

Figure 38. MRI Patient Count by Hospital Type

Figure 39 illustrates the probability of repeat MRI scans within 90 days by hospital type. Repeat rates vary considerably across institutions. “Other private” hospitals display the highest relative share of repeat MRIs, exceeding 20% of patients, suggesting potential quality concerns, patient complexity, or revenue-driven utilization incentives. University hospitals and secondary/tertiary public hospitals also show elevated repeat rates (around 10–15%), which may reflect the concentration of complex cases requiring follow-up imaging. Private hospitals and foundation universities, while serving fewer patients overall, exhibit comparatively lower repeat rates, suggesting differences in case mix or imaging protocols.

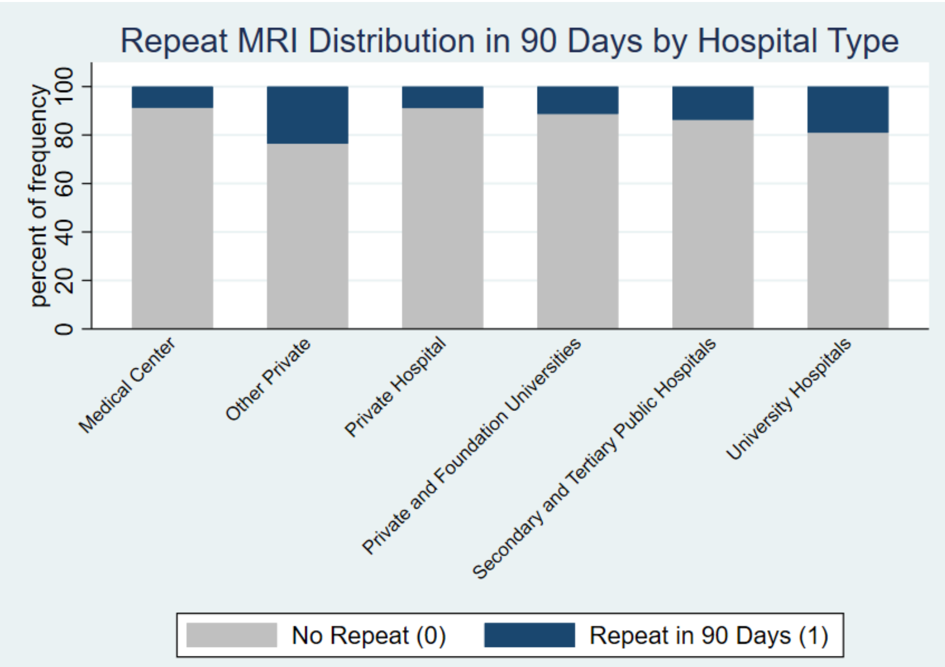

Figure 39. Repeat MRI Distribution in 90 Days by Hospital Type

Together, these results highlight two important aspects of diagnostic capacity. First, institutional concentration: the bulk of MRI activity occurs in high-capacity public and private hospitals, shaping system-wide diagnostic performance. Second, quality and efficiency variation: repeat scan rates differ markedly by hospital type, pointing to technological and organizational disparities that influence diagnostic certainty and patient management. From a policy perspective, high repeat rates in certain institutions raise concerns about efficiency, overutilization, and uneven quality standards across the healthcare system.

4.3. Key Variables and Their Construction

Now, the study describes the main variables used in the thesis, how they are defined and constructed below.

The dependent variable in the models is defined as a binary indicator of whether the patient's primary diagnosis is an ICD-10 "R-code." R-codes refer to symptoms rather than specific diagnoses. The thesis interprets an R-code result as an indicator of underdiagnosis. For each patient visit, $R\_Code\_Prob = 1$ if the primary diagnosis's ICD-10 code is in the range R00-R99, and 0 if the diagnosis is a defined disease. This variable is calculated directly from the diagnosis codes in the 2015-2016 patient data. As noted, $R\_Code\_Prob = 1$ in approximately 7-8% of visits in the sample. Using this variable allows us to quantify the probability that a patient leaves the hospital without a clear diagnosis.

This is a binary indicator indicating whether a patient received multiple MRI scans on the same day during their hospital stay. This is generated from procedure-level records in the patient dataset using multiple different MRI procedure names on the same day of hospital admission. According to this, if a patient had at least two different MRI procedure entries on the same date, $\text{Same Day Repeat MRI} = 1$. Multiple MRI scans performed on the same day at different territories indicate a quality problem in the healthcare system, suggesting poor image quality or unclear symptoms that make diagnosis challenging. This could potentially lead to an R code error. Therefore, the study included this as an independent variable to see if such an occurrence was associated with a higher probability of underdiagnosis.

Repeat MRIs performed on the same day may indicate supply-driven demand behaviour, or this could be due to time and fatigue, which make it difficult for doctors to diagnose complex cases, or it could be due to the poor image quality of the MRI machine. Thus, this variable is potentially endogenous and serves as a proxy variable, as it will be influenced by patient confusion, physician behaviour regarding potential overuse, and machine quality factors.

This variable is a hospital-level proxy for MRI practice quality generated from the 2019-2023 MRI patient-level dataset. Repeat MRIs performed within 90 days of the MRI patient dataset were used to generate the variable. repeat_90days=1. To calculate each hospital's long-term repeat MRI rate, all repeated MRIs performed between 2019 and 2023 are aggregated. This variable was then merged to the 2015-2016 analysis dataset using the hospital identifier as a time-independent hospital characteristic. For each hospital, Low Quality MRI is defined as a binary indicator, that equals 1 when the repeat rate exceeds 60%. This threshold identifies hospitals operating in low-quality MRI environments where overuse, inefficiency, or suboptimal diagnostic quality are systematically higher. Therefore, because it is measured using a later timeframe (2019-2023), it is not affected by the 2015-2016 results and can therefore be used as a proxy or external control for persistent quality differences in MRI practices.

To flexibly control for differences in hospital imaging capacity, the study includes categorical indicators of hospital MRI utilization per machine. Using hospital-level MRI utilization data from 2016, each hospital is categorized into three groups: Low, Medium, and High utilization of MRI scans per machine. Based on the 2015-2016 patient-level dataset, approximately 13% of patient visits are in low-utilization hospitals, approximately 53% in medium-utilization hospitals, and approximately 34% in high-utilization hospitals.

The severity score variable captures the clinical complexity of a patient's case by combining length of stay with a diagnosis-specific severity weight. The study first estimated the relationship between the diagnosis category and total hospital stay hours using a linear regression using the Z-category "Factors influencing health status and contact with health services." The estimated coefficients from this regression were assigned severity weights for each ICD code, and the severity weight for the initial category was set to zero. The severity score was then created by summing the patient's total length of stay and the assigned severity weight. Thus, the disease severity score quantitatively describes the complexity of the disease by combining the inherent severity of the patient's diagnosis and the length of stay.

These variable measures the number of patients seen by the treating physician on the day of the visit. This is calculated by counting all visits with the same doctor and on the same date, based on the physician ID and patient presentation date in the data. Accordingly, the daily patient count ranges from a low of 1 to a high of 568, with an average of approximately 75. In this study, it is expected that physicians with a higher patient load per day are more likely to use MRI and are more likely to miss a definitive diagnosis due to time constraints, leading to an R code.

This is a continuous variable representing hospital-level throughput. It represents the number of hospitalizations per bed per year. A higher bed turnover rate may indicate a busy hospital, where stays are short and patient volume is high. Data for the study are taken from the Ministry of Health's statistics on hospital capacity for 2015 and 2016. In our sample, there are approximately 87 hospitalizations per bed per year on average. This data is included in this study to control for hospital density and pressure because hospitals with very high turnover rates may be practicing hasty discharges, leading to more uncertain diagnoses.

The Referral degree variable represents the degree of a hospital's referral role within the Turkish healthcare system. It is an integer-coded classification ranking from low to high referral capacity. Higher values indicate hospitals are more likely to serve as tertiary referral centers. The lowest categories, such as 1, 2, and 3, represent smaller or primary care institutions, while the highest codes, such as 90 and 100, correspond to large, fully specialized hospitals. This variable allows us to control for institutional role and specialization in the analysis and capture the impact of a hospital's location in the referral network on diagnostic accuracy results.

The $hospital\ type$ variable includes grade roles A through E based on hospital capacity. This categorical variable controls for institutional differences. In Türkiye’s hospital classification system, A1 hospitals are *tertiary-level teaching and research hospitals* with specialist training programs and advanced care capacity. A2 hospitals are *tertiary-level general hospitals* that provide complex care but lack teaching status. B hospitals are *provincial general hospitals* with moderate capacity and 24/7 emergency and specialty coverage. C hospitals are *district-level general hospitals* with basic specialty services and limited inpatient capacity. D hospitals are *small district hospitals* providing core services such as deliveries, minor surgeries, and outpatient care. E1 hospitals are *comprehensive integrated district hospitals* offering limited inpatient and maternity care with some specialist support. E2 hospitals are *small integrated facilities* run by general practitioners, providing basic emergency, lab, and imaging services in low-population areas. In addition, the $central\_hospital$ variable is included; approximately 10% of the sample is in these central hospitals. Including these hospitals helps control for unobserved differences across facilities.

Other important control variables are patient age and gender. Age is a continuous variable and ranges from 0 to 98 in the sample. The mean age is 40. Gender is included as a dummy variable, and we use factor variables for $patient\_gender$ (male/female). This accounts for systematic differences in diagnostic patterns by age or gender.

To capture timing differences in diagnostic capacity, the study includes variables for day of the week and time of admission, which serve as proxies for auxiliary staff availability. $day\_category$ is an indicator of whether the admission occurred on a weekend/holiday or workday. Most admissions occur on weekdays; only about 17% are admitted on weekends/holidays when auxiliary staff levels are lower and elective admissions are rare. This controls for any reduced diagnostic staffing or different case mix on weekends.

Similarly, $time\_category$ indicates the time of day the admission occurred and was categorized into three groups: morning hours, afternoon/evening, and evening/night. Diagnostic support staff are concentrated during daytime hours, while staffing is reduced in the evenings and nights. This variable thus proxies the reduced availability of auxiliary staff and potential diagnostic bottlenecks outside regular hours.

In summary, variables include human capacity factors as patient demographics, severity, and doctor total visits daily, institutional capacity factors as hospital type, central hospital, referral degree and MRI utilization level per device and, most importantly, variables related to machine capacity as repeat MRI on same day and low-quality MRI. Having defined these variables, we now describe the empirical strategy for assessing the impact of MRI use on underdiagnosis.

4.4. Empirical Methodology

This section outlines the empirical framework used to estimate the determinants of underdiagnosis in the Turkish healthcare system, focusing on variables identified in human, institutional and machine capacity factors. Two main approaches are applied: (i) basic logistic regressions of the probability of underdiagnosis, (ii) instrumental variable (IV) models to address endogeneity in total daily physician visits and MRI quality utilization.

4.4.1. Logistic Regression for Underdiagnosis Probability

The study first estimates a logistic regression model to examine how various factors are associated with the probability of underdiagnosis. The dependent variable is $y_{ij} = R\_Code\_Prob_{ij}\ (1 = underdiagnosis, 0 = specific\ diagnosis)$. Key explanatory variables include MRI-related indicators and a full set of controls, as described above. Given the binary nature of the dependent variable, the primary modelling approach is logistic regression. For patient i in hospital j, the probability of underdiagnosis is modelled as:

$$logit\left(P\left(R\_Code\_Prob_{ij} = 1\right)\right) = \alpha + \beta_1\left(Repeat\_MRI\_Same\_Day_{ij}\right) + \beta_2(Low\_MRI\_Quality_j) + \beta_3(Doctor\_Total\_Daily\_Visits_{ij}) + \beta_4(MRI\_Utilization\_Per\_Machine_j) + \gamma^T X_{ij} + \varepsilon_{ij}$$

Where:

X_{ij}: Vector of control variables including patient age, gender, severity score, hospital type, central hospital status, and temporal indicators.

ε_{ij}: Error term, clustered at the hospital level to account for intra-hospital correlation.

The logistic regressions are estimated in three stages:

i) **Human Capacity Model**: includes supply-side indicators including repeat MRI same day, doctor total daily visits, MRI utilization level per machine and controls but excludes quality metrics.

ii) **Hospital Capacity Model**: augments the baseline with hospital capacity indicators of Referral Degree Rate and Bed Turnover Ratio, and Low MRI Quality

iii) **Machine Capacity Model**: augments the baseline by incorporating both human and hospital capacity indicators and includes machine capacity indicators as utilization levels per machine for MRI, CT, and mammography, and Low MRI Quality metric.

4.4.2. Instrumental Variable (IV) Models for Doctor Total Daily Visits and MRI Machine Quality

This chapter addresses key methodological challenges in estimating the causal impact of diagnostic system constraints on underdiagnosis. In particular, it focuses on the endogeneity of doctor' daily visits and same-day repeat MRI and presents an instrumental variable (IV) strategy to isolate their true effects. The discussion proceeds step by step: (i) introducing IV methodology using time-based instruments due to unobserved doctor quality, (ii) introducing IV methodology for repeat MRI utilization to detect machine quality.

First, the total number of physician visits, defined as the total number of patients seen by a physician on a given day, cannot directly capture physician workload, as it may be related to unobservable physician quality. Therefore, $Doctor\_Total\_Daily\_Visits_{ij}$ serves as one of the main endogenous regress in the model. It reflects variation in a physician's daily workload and can affect diagnostic accuracy. However, because more patients may be assigned to more skilled physicians, this measure of workload could potentially be related to unobservable physician quality, necessitating an IV approach in the study. To address this, the study constructs an instrumental variable using exogenous time-based variation:

$time\_category$: morning, afternoon, evening appointment slots

$day\_category$: weekday, weekend, holiday

These time variables influence how many patients a doctor sees on a given day, but are plausibly uncorrelated with a doctor's intrinsic diagnostic skill.

First-stage regression (relevance test):

$$Doctor\_Total\_Visits\_Day_{ij} = \pi_0 + \pi_1 i.time\ category + \pi_1 i.day\ category + \delta^T X_{ij} + \vartheta_{ij}$$

Second-stage regression (structural equation):

$$logit\left(P\left(R\_Code\_Prob_{ij} = 1\right)\right) = \alpha + \beta_1\left(Doctor\_Total\_Visits\_Day_{ij}\right) + \beta_2(Repeat\_MRI\_Same\_Day_{ij}) + \beta_3(MR\_Utilization\_Per\_Machine_j) + \gamma^T X_{ij} + \varepsilon_{ij}$$

Theoretically, if unobserved physician quality is positively correlated with workload, the OLS coefficient will exhibit a reverse trend. Thus, the IV estimation should isolate the true causal effect of workload, which the study expects to be positive. Endogeneity may arise because unobserved physician quality (skill/speed) affects both workload and underdiagnosis. Because Türkiye has a self-referral system, patients can make appointments with doctors of their choice. Therefore, highly skilled physicians are more likely to make more appointments and better diagnoses. Thus, a physician's workload reflects not only the time constraints they face but also latent quality characteristics. This endogenous relationship tends to bias the coefficient toward zero or even negative in the basic OLS regression. In a simple OLS model, these effects cancel each other out, leading to attenuation bias. By using time-based IVs that shift workload exogenously, the study holds quality constant and recovers the true negative diagnostic effect of increased workload.

The main hypothesis of the study is that MRI overutilization, whether due to increased capacity pressure or poor diagnostic quality, leads to higher rates of underdiagnosis and is driven by the efficiency-quality trade-off in healthcare. However, MRI utilization intensity is likely endogenous due to omitted variables such as management quality across hospitals, doctor skills, or diagnostic culture.

To identify the causal effect of MRI utilization intensity pressure on underdiagnosis, the study proposes an instrumental-variables (IV) strategy that treats high-frequency

MRI intensity as endogenous and uses Low MRI Quality (2019–2023) as an instrument. It presents a primary design using Same Day Repeat MRI as the endogenous regressor. Both are estimated with clustered standard errors at the hospital level.

Endogenous regressor: The study considers hospital-level measures of MRI utilization intensity as potentially endogenous. Specifically, repeat MRI on the same day is measured as a binary indicator equal to one if a patient received two or more MRI scans on the same day. This variable serves as a proxy for diagnostic inefficiency or scanner mismanagement, capturing instances in which limited capacity or poor image quality may lead to redundant imaging. However, this measure may also be correlated with unobserved institutional or physician characteristics, such as management quality, triage practices, or patient case-mix. Hospitals with higher MRI throughput may simultaneously differ in their diagnostic accuracy, which could bias estimates in underdiagnosis models if not properly addressed. Evidence from Türkiye shows that poor image quality is a significant driver of repeated MRI scans, with up to 34% repetition rates due to suboptimal quality (Başar et al., 2019), while international studies similarly document rising rates of repeat imaging as a signal of inefficiency and overuse in advanced health systems (Smith-Bindman et al., 2008).

Instrument: The study uses an indicator for low MRI quality measured from 2019–2023 low MRI quality as an instrument. $Low\_MRI\_Quality_{ij}$ is a binary variable capturing whether the hospital is classified as low-quality in MRI performance based on repeat scan rates in 90 days during 2019–2023.

IV requirements

1) **Relevance Condition**

Definition: $Cov(Low\_MRI\_Quality_{ij}; Repeat\_MRI\_Same\_Day_{ij} \neq 0)$

This assumption requires that hospital-level poor MRI quality measured from 2019–2023 predicts greater diagnostic inefficiency or throughput burden in 2015–2016 in the Turkish healthcare system.

This assumption is empirically testable via first-stage F-statistics (rule-of-thumb: F > 10). It reflects institutional persistence: hospitals exhibiting low diagnostic value from imaging in recent years likely exhibited similar inefficiencies in earlier years due to stable capacity, coordination, or professional habits.

2) Exclusion Restriction

Definition: $(Low\_MRI\_Quality_{ij} \perp \varepsilon_{ij} \;\; X_{ij})$

The instrument must affect the outcome (R_Code_Prob)only through the endogenous variable (same day repeat MRI), conditional on X_{ij}. It should have no direct effect on underdiagnosis.

Where X_{ij} represents a set of control variables that account for patient, hospital, and capacity-related characteristics. Patient demographics are captured through age and gender, while hospital characteristics include the type of hospital and whether it is designated as a central hospital. Human capacity controls incorporate the day and time categories, which proxy for variations in physician availability and workload across different periods.

However, a potential threat to exclusion is that poor MRI quality may itself be directly associated with diagnostic error, beyond its effect on imaging intensity. To address this concern, the instrument is interpreted not as a random shock but as a capacity-channel instrument, justified by the persistence of low-quality imaging environments and institutional practices. In addition, the empirical strategy incorporates rich hospital and time controls to absorb systematic variation in patient mix, institutional roles, and temporal shocks, thereby strengthening the credibility of the identification approach.

Second Stage (Structural Equation): Let $R\_Code\_Prob_{ij} \in \{0,1\}$ be the underdiagnosis indicator. The structural equation is:

$$logit(P(R\_Code\_Prob_{ij} = 1)) = \alpha + \beta(Repeat\_MRI\_Same\_Day_{ij}) + \gamma^T X_{ij} + \varepsilon_{ij}$$

First Stage Equation:

$$Repeat\_MRI\_Same\_Day_{ij} = \pi_0 + \pi_1 Low\_MRI\_Quality_{ij} + \theta^T X_{ij} + \mu_{ij}$$

We expect $\pi_1 > 0$: hospitals flagged as low-quality exhibit higher MRI intensity via more scans per doctor-day and/or more repeats.

The estimation strategy relies on two-stage least squares (2SLS) with a linear probability specification. This approach serves as a useful benchmark, as it not only facilitates the interpretation of results in terms of marginal effects but also directly provides weak-instrument robust diagnostics, including the Kleibergen–Paap FFF-statistic and the Anderson–Rubin test.

The identification strategy first requires relevance, meaning that the instrument should strongly predict the endogenous regressor in the first-stage equation. To assess this, the study reports several diagnostic statistics: the Kleibergen–Paap rk Wald FFF-statistic, where values above 10 are generally taken as evidence of strong instrument relevance; the partial R-squared, which indicates the proportion of variation in the endogenous regressor explained by the instrument; and the Kleibergen–Paap LM statistic, which serves as an underidentification test.

Secondly, the exclusion restriction requires that low MRI quality affects underdiagnosis only through its impact on MRI intensity. Although this assumption cannot be directly tested, the study supports its validity through two key arguments:

i) Rich covariates (X_{ij}): patient severity, demographics, hospital type, central hospital, day/time, year; these absorb alternative pathways from $Low\_MRI\_Quality_{ij}$ could directly influence underdiagnosis.

ii) Structural persistence: MRI quality observed in 2019–2023 is unlikely to directly affect 2015–2016 outcomes, aside from its correlation with MRI overutilization.

This chapter has presented a structured IV strategy for addressing endogeneity in physician workload and MRI utilization. The theoretical and empirical steps outlined here allow us to credibly estimate the causal impact of diagnostic capacity on underdiagnosis. The next chapter presents the econometric models and results from these models in detail.

CHAPTER 5

ECONOMETRIC MODELS AND RESULTS

In this section, the thesis analyses how capacity constraints at the physician, hospital, and technology levels affect the probability of underdiagnosis. The study builds econometric models to measure these effects, estimate them using logistic regressions, and address potential endogeneity using instrumental variables. All models use the dependent variable $y = R\_Code\_Prob$. The models include relevant control variables such as patient demographics, severity of disease, hospital type fixed effects, time-of-day indicators, and standard errors are clustered at the hospital level to account for within-hospital correlation. Below, the study presents model specifications, results, and interpretations, supported by theoretical expectations and literature.

5.1. Logistic Regression for Underdiagnosis Probability

This section presents empirical findings from logistic regressions estimating the probability of underdiagnosis. Each section below introduces a specific model, presents the regression results in a formal table, and interprets the estimated coefficients. All models use hospital-clustered standard errors and control for patient age, sex, severity, day of the week, time of day, hospital type, and referral degree status.

5.1.1. Logistic Regression 1: Human Capacity Model

This model tests whether underdiagnosis is predicted by physician-level daily patient load, same-day repeat MRI behaviour, and the MRI utilization level per device at the institutional level, controlled by patient and hospital characteristics.

The baseline specification estimates the probability that a patient encounter i at hospital j will lead to underdiagnosis $(R\_Code\_Prob = 1)$ as a function of human capacity and patient/hospital controls:

$$logit(P(R\_Code\_Prob_{ij} = 1)) = \alpha + \beta_1(Repeat\_MRI\_Same\_Day_{ij}) + \beta_2(Doctor\_Total\_Visits\_Day_{ij}) + \beta_3(MR\_Utilization\_Per\_Machine_j) + \gamma^T X_{ij} + \varepsilon_{ij} \quad (1)$$

Where:

i indexes patient encounters

j indexes hospitals

X_{ij} is a vector of control variables, including patient age, gender, severity score, hospital type, central hospital status, and day/time categories.

This model is estimated with hospital-clustered standard errors using a sample of over 2.6 million observations from 577 hospitals.

Table 5 presents the logistic regression of the probability of underdiagnosis on physician capacity level. Several important patterns emerge in this human capacity model. First of all, the coefficient for '*physician total daily visits'* is negative and statistically significant ($\beta_2 = -0.003$). This suggests that physicians who see more patients per day are associated with lower underdiagnosis rates. However, the thesis aims to expect that overworked doctors may make more errors and increase underdiagnosis due to time and resource constraints, as explained in chapter 3 using Arrow's uncertainty theory.

In addition, relevant to the purpose of this thesis, there are well-documented studies in the literature. Ivanovic et al. (2023) found in their empirical study that longer interpretation times, higher shift volumes, and weekend interpretations increase the risk of errors. Therefore, as noted in Chapter 4, the study hypothesizes that the negative coefficient is due to physicians who can handle more patients being more skilled or working in better-resourced environments, leading to fewer missed diagnoses. While higher patient volumes may increase underdiagnosis, the physician workload in Model 1 incorporates an unobservable endogenous variable, such as physician skill, making the coefficient negative.

Table 5. Logistic Regression- Human Capacity Model

	(1) R_Code_Prob b	se	p
Human Capacity Factors			
Doctor’s Total Daily Visits	-0.003**	0.002	0.035
Patient Age	0.009***	0.002	0.000
Patient Gender			
Reference: *Male*			
Female	-0.396***	0.096	0.000
Severity Score	0.005***	0.001	0.000
Day Category			
Reference: *Workday*			
Vacation Day	0.254***	0.057	0.000
Time Category			
Reference: *Morning (05:00-10:59)*			
Afternoon (11:00-15:59)	0.267***	0.048	0.000
Evening/Night (16:00-04:59)	0.580***	0.066	0.000
Institutional Capacity Factors			
Central Hospital	-0.413	0.342	0.227
Hospital Type			
Reference: *A1: Teaching and research hospitals*			
A2: Tertiary-level general hospitals	-0.523*	0.271	0.054
B: Provincial general hospitals	-0.309	0.283	0.276
C: District-level general hospitals	0.048	0.313	0.878
D: Small district hospitals	0.888***	0.271	0.001
E1: Comprehensive integrated district hospitals	1.162***	0.241	0.000
E2: Small integrated facilities	1.569***	0.297	0.000
Other	0.015	0.423	0.972
Machine Capacity Factors			
Same-Day Repeat MRI	0.311**	0.150	0.038
MRI Utilization Level per Machine			
Reference: *Low*			
Medium	0.333	0.286	0.244
High	0.190	0.310	0.541
Constant	-2.937***	0.302	0.000
Log Pseudolikelihood	-669334.403		
Pseudo R^2 (McFadden)	0.054		
Observations	2648448.000		

Dependent variable: R_Code_Prob (Probability of Underdiagnosis)
Clustered standard errors at hospital level (577 clusters)
Significance Levels: *** p < 0.01 ** p < 0.05 * p < 0.1

The results for the other variables in the physician capacity model are as expected. The indicator for '*Same-Day Repeat MRI'* is positive and significant $(\{\beta_1 = 0.311\}, \mathrm{p} = 0.038)$, indicating that repeating a patient's MRI on the same day increases the likelihood of underdiagnosis. The corresponding odds ratio, β_1 is $e^{0.311} \approx 1.3$, indicates that patients who undergo repeat MRI scans on the same day have an approximately 37% higher probability of underdiagnosis. The need for multiple repeat MRI scans on the same day raises the research question of the paradoxical dichotomy of diagnostic accuracy between human expertise and machine capacity, or quantity and quality. What are the two factors that create this paradox? Is the initial scan inconclusive because the machine provides poor image quality? Or are physicians overworked under time and resource constraints, leading them to overtest due to the complexity of the disease? This paradoxical dichotomy, which can be caused by both conditions, ultimately increases underdiagnosis, which can lead to a delay in the diagnostic process.

Patient factors also show significant effects in the model. '*Patient Age'* has a positive coefficient $(0.0086, p < 0.001)$ indicating that older patients are more likely to be underdiagnosed. In this context, while the marginal effect is small for each additional year of age, it becomes significant when there is a large age difference. There are studies in the literature that support this. In their empirical study on patient age and gender differences, Liu et al. observed that older patients had higher diagnostic errors than younger patients across different specialties (Liu et al. 2022). Skinner et al. also concluded in their empirical study that "Increasing age is associated with lower diagnostic accuracy, attributed to atypical presentations and a greater number of comorbidities" (Skinner et al., 2016).

On the other hand, the '*patient gender'* indicator in Table 5 is negative at $-0.3958, p < 0.001$, indicating that female patients have 33% lower probability of underdiagnosis than males $(e^{-0.396} \approx 0.67)$. This finding is also supported by literature: Male elderly patients experienced a medical error rate in general surgery and emergency departments that was almost 4% greater than that of female patients (Liu et al. 2022). Although finding a gender difference in the study was not expected, this gender difference is noteworthy and thus reflects that demographic factors may be associated with differences in error rates and thus may influence diagnostic outcomes. Contrary to this result, there are studies in the literature that gender inequalities persist for women in healthcare by showing that female patients are subject to delayed diagnosis more than males. However, some studies indicate that male patients experience more medical errors than females: For example, Liu et al. found that male elderly patients experience approximately 4% higher rates of medical errors than female patients in general surgery and emergency departments (Liu et al. 2022). Finding a gender difference is not one of the objectives of this study, but the gender difference found in a comprehensive dataset is striking. This reflects the strong control effect of demographic factors on diagnostic precision and the potential impact on diagnostic outcomes.

The *'severity score'* has a positive coefficient of $0.0048, p < 0.001$, indicating that more severe or complex cases are more likely to result in underdiagnosis. This result is directly related to uncertainty in Arrow's theoretical framework, on which the study is based. As the complexity of a disease increases, so does its uncertainty, which can make it difficult for physicians to make a correct diagnosis. In addition, Skinner et al (2016) has also indicated that diagnostic errors are more common in complex cases with multiple comorbidities or acute presentations.

Visiting time indicators also show strong and significant effects in the model. Compared to baseline visits in the early morning (4:00 AM–10:59 AM), visits in the afternoon (11:00 AM–3:59 PM) increase the probability of underdiagnosis by approximately 0.267 $(\{p < 0.001\})$, while visits in the evening/night (4:00 PM–5:00 AM) increase it even further, reaching a coefficient of 0.580 $(\{p < 0.001\})$. This corresponds to approximately 30% higher (afternoon) and 78% higher (evening/night) probabilities of underdiagnosis compared to the same patients seen in the morning. This finding strongly aligns with the uncertainty framework that physicians working under time and resource constraints can lead to fatigue and diagnostic errors. As the day progresses, clinicians may become fatigued, impairing their diagnostic judgment and thoroughness. Numerous studies in the literature support this: stress, fatigue, increased workload, and night shifts have been found to be contributing factors to increased diagnostic error rates. They also reported that medical errors occur more frequently during night or rotating shifts than during day shifts (Gates et al., 2018; Muzio et al., 2019).

Furthermore, within the day category, weekends/holidays show a positive coefficient $(0.2536, p < 0.001)$, indicating a higher probability of diagnosis on holidays compared to weekdays. This is consistent with the "weekend effect" in healthcare. Accordingly, weekends differ from weekdays in terms of the resources available, the care procedures, and the availability of specialists (Becker, 2007; Bion et al. 2021). Thus, patients hospitalized on weekends may not receive adequate care due to fewer staff or less service availability. In this context, according to the results of Model 1, being examined on weekend/holidays leads to an approximately 29% higher probability of not being diagnosed$\left(e^{\{0.254\}} \approx 1.29\right)$. This finding is consistent with previous research showing that there is a higher probability of an error occurring during the weekends (Becker, 2007; Piras et al. 2024). Overall, the time and day effects illustrate how that auxiliary staff availability affect human capacity factors (personnel, fatigue) and thus can affect diagnostic performance.

Finally, the hospital type indicators as the institutional capacity control variable reveal significant institutional differences in Model 1. The baseline category in the model is Type A1, *teaching and research hospitals*, the highest-tiered hospitals. The results show that most lower-tiered hospitals have a higher underdiagnosis rate compared to A1. Type D (*Small district hospitals)* $(0.888, p < 0.001)$, E1 $(1.161, p < 0.001)$, and E2 (*Small integrated facilities)* $(1.569, p < 0.001)$ hospitals have positive and significant coefficients. These values indicate that, holding patient case mix constant, a Type E2 hospital is 4.8 times more likely to be underdiagnosed than a Type A1 hospital $\left(e^{\{1.569\}} \approx 4.8\right)$. In contrast, Type A2 (*Tertiary-level general hospitals*) hospitals have lower underdiagnosis rates than A1 $(-0.523, p = 0.05)$, indicating that they outperform the highest-tiered reference category. In addition, type B is negative, while Type C and "Other" are positive but not significant in the model. Additionally, the Central Hospital dummy variable is negative (-0.413) but not statistically significant in this model.

These models strongly suggest that smaller, less resourced hospitals have more difficulty making accurate diagnoses, while larger, more specialized hospitals achieve better diagnostic integrity. This is consistent with general expectations: larger hospitals generally have more specialized staff, better diagnostic equipment, and diagnostic processes, thus resulting in fewer underdiagnoses. On the other hand, smaller hospitals may lack certain specialist services or up-to-date technology, making diagnosis confirmation difficult. Vaughan et al (2018) raised concerns that smaller hospitals face difficulties investing in acute medicine diagnostic services and maintaining 24/7 specialist coverage. The study results empirically support these concerns; patients in small community hospitals (Types D and E) are much more prone to underdiagnosis events. From another perspective, this disparity highlights an equity issue: patients' diagnostic confidence depends in part on where they receive care, a point noted by health system researchers who are concerned that those living in less urban areas may receive “systematically lower quality care,” including less accurate diagnoses (Gaughan et al, 2020). In summary, the human capacity model in Table 5 highlights several important influences on underdiagnosis: repeat imaging needs, patient age and gender, timing of visit, and hospital characteristics. The unexpected negative sign on doctor workload hinted at underlying biases, which we explore next using an IV approach.

5.1.2. Logistic Regression 2: Institutional Capacity Model

The model here expands on hospital-level capacity factors to examine how they influence underdiagnosis. In addition to institutional capacity controls such as hospital type, central hospital status, and per-hospital MRI utilization level in Model 1, the study also included Bed Turnover Rate and Referral Grade as indicators of hospital capacity stress in Table 6. The study also explicitly included the Low MRI Quality (2019-2023) indicator to account for differences in machine quality and retained the physician visits and repeat MRI variables to control for these effects.

$$logit\left(P\left(R\_Code\_Prob_{ij} = 1\right)\right) = \alpha + \beta_1\left(Repeat\_MRI\_Same\_Day_{ij}\right) + \beta_2(Doctor\_Total\_Visit\_Days_{ij}) + \beta_3(MRI\_Utilization\_Per\_Machine_j) + \beta_4(Referral\_Rate_j) + \beta_5(Bed\_Turnover\_Rate_j) + \gamma^T X_{ij} + \varepsilon_{ij} \quad (2)$$

Table 6. Logistic Regression- Institutional Capacity Model

	(1) R_Code_Prob b	se	p
Human Capacity Factors			
Doctor's Total Daily Visits	-0.0056	0.0032	0.077
Patient Age	0.0084***	0.0013	0.000
Patient Gender			
Reference: *Male*			
Female	-0.2578**	0.0961	0.007
Severity Score	0.0042***	0.0011	0.000
Day Category			
Reference: *Workday*			
Vacation Day	0.2456	0.1319	0.063
Time Category			
Reference: *Morning (05:00-10:59)*			
Afternoon (11:00-15:59)	0.2789***	0.0627	0.000
Evening/Night (16:00-04:59)	0.9060***	0.1035	0.000
Institutional Capacity Factors			
Central Hospital	-0.6987*	0.3255	0.032
Referral Degree	0.0038	0.0038	0.309
Bed Turnover Rate	-0.0056	0.0065	0.389
Hospital Type			
Reference: *A1: Teaching and research hospitals*			
A2: Tertiary-level general hospitals	-0.8375***	0.1912	0.000
B: Provincial general hospitals	-0.7564***	0.2023	0.000
Machine Capacity Factors			
Same-Day Repeat MRI	0.6894	0.3856	0.074
Low MRI Quality	0.6892*	0.3237	0.033
MRI Utilization Level per Machine			
Reference: *Low*	0.0000	.	.
Medium	0.7462**	0.2724	0.006
High	1.1713*	0.5390	0.030
Constant	-3.2427***	0.4140	0.000
Log Pseudolikelihood	-166730.0349		
Pseudo R^2 (McFadden)	0.0607		
Observations	574843.0000		

Dependent variable: R_Code_Prob (Probability of Underdiagnosis)
Clustered standard errors at hospital level (577 clusters) , *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

The results are similar to human capacity model (Table 5) but also show several significant differences. First, the coefficient for Same-Day Repeat MRI is positive and larger in this model ($\beta_1 = 0.689$ $\{p = 0.074\}$) The Institutional Capacity model indicates that requiring a repeat MRI is associated with a higher probability of underdiagnosis $e^{\{0.689\}} \approx 2.0$. The variable for Poor MRI Quality (2019–2023) is positive and significant $0.689\ p = 0.033$, indicating that patients imaged in hospitals with poor-quality MRI equipment are more likely to be underdiagnosed. Thus, the importance of quality in accurate diagnosis is highlighted. This is consistent with research expectations: while advanced imaging provides clearer results, older scanners can produce ambiguous images, which can lead to physicians missing diagnoses and underdiagnosing them. Because radiologic equipment has a limited lifespan, failures and a decline in image quality are inevitable. In particular, equipment older than 10 years is no longer state-of-the-art and has to be changed since it may malfunction or break down, which could create delays in patient diagnosis and care (European Society of Radiology, 2014). These findings are essentially empirical confirmation that up-to-date imaging quality is crucial for reducing diagnostic errors. They also provide a quantitative perspective on the harms caused by older equipment.

One key difference in Table 6 is the effect of MRI utilization levels per device. While MRI utilization per device is not significant in Model 1, here the coefficient for Medium MRI utilization per device was 0.746 (p=0.006) and the coefficient for High MRI utilization per device was 1.171 (p=0.030), both positive and significant. This suggests that facilities with medium or high MRI utilization per device experienced a higher probability of underdiagnosis compared to hospitals with low utilization. In fact, a hospital with a high level of MRI utilization would be expected to have fewer underdiagnoses. The opposite finding in the model results suggests overuse of imaging modalities. High MRI utilization may indicate defensive or indiscriminate screening practices that do not necessarily lead to better diagnostic outcomes. The literature on overuse of diagnostic tests has found that more testing can sometimes lead to noise or even harm but does not improve baseline outcomes (Müskens et al. 2022). Findings suggest that more is not always better for diagnosis. Performing more MRIs does not necessarily reduce underdiagnosis by detecting more missed diagnoses; on the contrary, it may increase the risk of underdiagnosis. Conversely, hospitals with high MRI utilization may be those experiencing capacity constraints, using scans as a shortcut, or handling complex cases. Therefore, high MRI utilization and high underdiagnosis due to case complexity coexist. This is consistent with evidence that over-testing can expose patients to harm without providing a clear benefit (Ali et al. 2024).

While looking specifically at hospital capacity factors, '*bed turnover rate'* is negative (-0.0056) but not statistically significant (p=0.389). This indicates that, controlling for other factors, there is no significant relationship between a hospital's bed turnover rate and underdiagnosis. The study expected that a very high bed turnover rate might increase underdiagnosis due to time pressures on diagnostic evaluation; however, it failed to find a significant result. Studies in emergency department settings have also found that emergency department crowding or occupancy does not significantly alter diagnostic decision-making (Chou et al. 2022). In this model, we do not see a significant hospital-level capacity effect on bed turnover. However, we caution that this may be due to sample size or limited variability. In addition to this, the *'Referral Degree'* status $(0.0038, \mathrm{p} = 0.309)$ is not significant. Here, it can be interpreted as hospital's case complexity or network centrality has also no clear effect on underdiagnosis.

Looking at the results for the other institutional capacity control variables, the coefficient for Central Hospital status is negative and significant $(-0.699, p = 0.032)$. This suggests that being located in a central hospital reduces the risk of underdiagnosis, and it is consistent with expectations; central hospitals likely have more specialists and diagnostic resources, resulting in more accurate diagnoses. In addition, under hospital type, both Hospital Type A2 and Type B are negative and highly significant $(-0.838\ and\ -0.756, respectively, p < 0.001)$. This suggests that secondary healthcare facilities, types A2 and B, have significantly lower underdiagnosis rates than the highest-tier tertiary healthcare facilities (type A1). The result is noticing that it is consistent with the finding that hospitals with higher MRI usage rates show higher underdiagnosis rates.
Finally, the effects for Vacation Day, Afternoon, and Evening/Night across day and time categories remain positive and significant $(0.246, 0.279, 0.906, \mathrm{respectively, all\ p} \leq 0.1)$ The coefficient for Evening/Night (0.906) is larger here than in model 1, indicating an almost 2.5 times higher

probability of underdiagnosis during night shifts compared to morning shifts ($e^{0.906} \approx 2.47$). This strongly reflects the risk of underdiagnosis during the night shift documented in the literature.

In summary, the institutional capacity model (Table 6) highlights the importance of technology and practice patterns: low MRI quality and high MRI utilization are associated with significantly more underdiagnosis. In contrast, the overall measure of hospital density (bed turnover) showed no effect, but institutional excellence (centralized, A2/B types) clearly reduced diagnostic error rates. These findings suggest that the provision of modern diagnostic equipment and avoiding overreliance on tests are important for diagnostic success.

5.1.3. Logistic Regression 3: Machine Capacity Model

The machine capacity specification extends the baseline model by incorporating both operational intensity and structural quality measures.. Here, the key explanatory variables capture technological load per scanner (MRI, CT, mammography utilization), scanner quality, and repeat status alongside institutional controls. The specification allows testing whether higher equipment utilization intensity increases the likelihood of underdiagnosis, measured by assignment of an R-code.

$$logit\left(P(R\_Code\_Prob_{ij} = 1)\right) = \alpha + \beta_1(Repeat\_MRI\_Same\_Day_{ij}) + \beta_2(Low\_MRI\_Quality_j) + \beta_3(Doctor\_Total\_Visits\_Day_{ij}) + \beta_4(MR\_Utilization\_Per\_Machine_j) + \beta_5(CT\_Utilization\_Per\_Machine_j) + \beta_6(MAM\_Utilization\_Per\_Machine_j)\ \beta_7(Referral\_Rate_j) + \gamma^T X_{ij} + \varepsilon_{ij}$$

Table 7. Logistic Regression- Machine Capacity Model

	(1) R_Code_Prob b	se	p
Human Capacity Factors			
Doctor's Total Daily Visits	-0.0052	0.0030	0.083
Patient Age	0.0098***	0.0018	0.000
Patient Gender			
Reference: *Male*			

Table 7. (continued)

Female	-0.2239**	0.0855	0.009
Severity Score	0.0040***	0.0012	0.001
Day Category			
Reference: *Workday*			
Vacation Day	0.2741*	0.1161	0.018
Time Category			
Reference: *Morning (05:00-10:59)*			
Afternoon (11:00-15:59)	0.3203***	0.0653	0.000
Evening/Night (16:00-04:59)	0.9371***	0.0999	0.000
Institutional Capacity Factors			
Central Hospital	-0.5672**	0.2094	0.007
Referral Degree	0.0032	0.0055	0.566
Hospital Type			
Reference: *A1: Teaching and research hospitals*			
A2: Tertiary-level general hospitals	-0.5795*	0.2696	0.032
B: Provincial general hospitals	-0.7729**	0.2508	0.002
Machine Capacity Factors			
Same-Day Repeat MRI	0.6956	0.3793	0.067
Low MRI Quality	0.8471**	0.2997	0.005
MRI Utilization Level per Machine			
Reference: *Low*			
Medium	1.3676*	0.6009	0.023
High	1.7636	0.9600	0.066
BT Utilization Level per Machine	0.0000	.	.
Reference: *Low*			
Medium	0.9975	0.6581	0.130
High	-1.1562**	0.4479	0.010
MAM Utilization Level per Machine	0.0000	.	.
Reference: *Low*			
Medium	-1.0335***	0.1925	0.000
High	-1.2367***	0.2135	0.000
Constant	-2.5386***	0.2246	0.000
Log Pseudolikelihood	-188045.3426		
Pseudo R² (McFadden)	0.0635		
Observations	625301.0000		

Dependent variable: R_Code_Prob (Probability of Underdiagnosis)
Clustered standard errors at hospital level (577 clusters) , *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

Machine Capacity Model (Table 7) presents a logistic regression model focusing on machine capacity and quality factors in predicting underdiagnosis. This model includes BT and MAM machine utilization levels, and the sample size (N = 625,301, 49 hospital clusters). The key factors here are MRI utilization levels per device, along with the same-day repeat MRI indicator and the Low MRI Quality indicator, which reflects the impact of imaging-related capacity issues. Additionally, physician visits and other control variables are included in the model to isolate machine effects.

The Machine Capacity Model examines whether machine capacity and imaging quality predict underdiagnosis, controlling for physician workload, institutional tier, patient mix, and time effects. The results in Table 7 largely overlap with the observed results in Table 6. Two machine-side indicators are central. First, the low-quality MRI proxy (derived from 2019–2023 repeat behaviour) is strongly associated with underdiagnosis $(\beta = 0.847, p = 0.005, OR = 2.33)$ hospitals operating in low-quality imaging environments have more than double the odds of issuing an R-code. Second, Similar to the results in Table 6, same-day repeat MRI is positive and significant $(\beta = 0.696, p = 0.067, OR = 2.00)$, consistent with the idea that Hospitals having low-quality MRI equipment and indeterminate scans show more missed diagnoses and thus contribute to underdiagnosis.

At the per-device MRI utilization level, relative to low utilization, medium and high utilization are positive and significant $(\beta = 1.368, p = 0.023, OR = 3.93; \beta = 1.764, p = 0.066, OR = 5.83)$.

Impact of high utilization is even larger, indicating that these hospitals are approximately 5.83 times more likely to experience underdiagnosis than hospitals with low utilization. The high MRI utilization in hospitals is likely explained by the defensive medicine or patient expectations inherent in the literature on which the study is based. Tertiary care facilities, including city hospitals, see large numbers of patients, and physicians, under time constraints and fatigue, may substitute MRI for clinical judgment. Radiologists may have less time for comprehensive interpretation per scan, which may negatively impact the quality and interpretation of the resulting images. Thus, this pattern is consistent with a capacity-strain mechanism: beyond a threshold, rising scans-per-machine are associated with steeper odds of underdiagnosis, even after adjusting for human and institutional factors. By contrast, higher CT and mammography utilization levels are associated with lower odds of R-codes (e.g., mammography high: $\beta = -1.237, p < 0.001, OR = 0.29$),. Here, Mammography and CT are governed by stricter and more standardized quality assurance systems than MRI. For example, mammography is regulated under laws such as the Mammography Quality Standards Act (MQSA), while CT accreditation programs require dose monitoring and equipment checks to ensure consistent quality (U.S. Congress, 1992; ACR, 2017; Feigin, 2023). By contrast, MRI quality is harder to standardize because image outcomes depend on many technical factors and vary more across hospitals and devices (Owusu et al., 2021; Hagiwara et al., 2020). As a result, while CT and mammography tend to deliver more consistent results under high use, MRI is more prone to inefficiency when volumes rise.

Institutional and human capacity controls behave as expected. Central hospitals exhibit lower underdiagnosis $(\beta = -0.567, p = 0.007, OR = 0.57)$, and A2 (β = −0.579, p = 0.032, OR = 0.56) and B $(\beta = -0.773, p = 0.002, OR = 0.46)$ hospitals outperform the A1 baseline, evidence of institutional resilience. Time pressures are pronounced: vacation days $(\beta = 0.274, p = 0.018, OR = 1.32)$, afternoons $(\beta = 0.320, p < 0.001, OR = 1.38)$, and evening/night $(\beta = 0.937, p < 0.001, OR = 2.55)$ all raise the likelihood of an R-code, consistent with off-peak bottlenecks and reduced staffing. On patient mix, age $(\beta = 0.0098, p < 0.001, OR\ per\ year = 1.01)$ and severity $(\beta = 0.0040, p = 0.001)$ increase underdiagnosis, while female patients have lower odds than males $(\beta = -0.224, p = 0.009, OR = 0.80)$. The daily physician-visits regressor is small and marginally negative $(\beta = -0.0052, p = 0.083, OR \approx 0.995\ per\ additional\ visit)$

In conclusion, the machine capacity model (Table 7) supports the thesis mechanism that both imaging quality (as proxied by repeat behaviour) and MRI utilization level per machine are independently linked to higher underdiagnosis. In Arrow’s terms, expanding quantity (machine capacity and volume) without protecting quality (protocols, reading time, QA) can shift uncertainty from access to accuracy, a paradoxical dichotomy makes visible. From these three capacity models, it follows that underdiagnosis is a multifactorial problem: human capacity factors, including physician workload, decision fatigue, patient age, gender, disease severity, and proxies for auxiliary staff availability (day of week and time of admission); institutional capacity factors, including hospital roles, imaging utilization levels, and reference quality; and technology capacity.

5.2. IV Model Regression for Underdiagnosis Probability

5.2.1. IV Model of Human Capacity: Physician Workload and Auxiliary Staff Availability

The thesis applies the two-stage IV strategy described in Chapter 4 to isolate the causal effect of physician capacity on underdiagnosis as measured by daily patient load. According to Table 8, the first-stage regression results (1) show that the instrumental variables (Afternoon (11:00-15:59), Evening/Night (16:00-04:59), and Vacation Day) significantly increase the number of patients per physicians, as validated by the first-stage results. Afternoon visits increase by 6.7 $(p < 0.001)$, evening/night visits increase by 10 $(p < 0.001)$,, and vacation days increase by 8 $(p = 0.006)$. The joint F-statistic for excluded instruments is 11.88 $(p < 0.001)$, which is above the conventional threshold of 10 (Staiger & Stock, 1997); however, it is slightly lower than the Stock–Yogo 5% critical value (13.91) for a single endogenous regressor with three excluded instruments. This implies that the instruments, despite being moderately strong, are in proximity to the borderline, and they entail weak IV-robust inference. The Sanderson–Windmeijer test affirms the instrument relevance by the same $F = 11.88$.

Table 8. IV Model of Human Capacity

	(1) Doctor total visits (daily)			(2) Underdiagnosis (R-code=1)		
	b	se	p	b	se	p
Time Category	0.0000	.	.			
Reference: *(05:00-10:59)*						
Afternoon (11:00-15:59)	6.6884***	1.6950	0.000			
Evening/Night (16:00-04:59)	9.9118***	1.6714	0.000			

Table 8. (continued)

Day Category						
Vacation Day	7.9656**	2.8973	0.006			
Doctor total visits (daily)				0.0031**	0.0011	0.004
Patient Age	-0.0080	0.0815	0.922	0.0006	0.0004	0.125
Patient Gender						.
Female	12.6003**	3.9370	0.001	-0.0681***	0.0150	0.000
Severity score	-0.0420	0.0407	0.302	0.0005***	0.0001	0.000
Repeat MRI Same Day	29.3273***	3.6637	0.000	-0.0689	0.0369	0.062
MRI per Device						
Reference: *Low*						
Medium	14.9174	9.7463	0.126	-0.0289	0.0359	0.421
High	17.8203	9.4207	0.059	-0.0412	0.0382	0.281
Central Hospital	0.9704	11.3742	0.932	-0.0347	0.0483	0.473
Hospital Type						
Reference: *A1*						.
A2	-19.0472	14.1443	0.178	0.0267	0.0537	0.619
B	-35.1836*	14.6807	0.017	0.0931	0.0521	0.074
C	-26.5712	17.3585	0.126	0.0856	0.0665	0.198
D	-48.238***	12.5732	0.000	0.2531***	0.0498	0.000
E1	-68.348***	11.8773	0.000	0.3974***	0.0613	0.000
E2	-66.641***	11.4820	0.000	0.4176***	0.0531	0.000
Other	-7.5875	10.2416	0.459	0.0238	0.0418	0.569
Constant	63.6740***	10.3112	0.000	-0.1607**	0.0596	0.007
Observations	2648448.0			2648448.0000		
R-squared	0.0641			-0.5507		
Adj. R-sq.	0.0641			-0.5507		
F-statistic	61.1268			95.7032		

Standard errors clustered by hospital
Instruments (excluded): time category, day category
Kleibergen–Paap rk Wald F (weak-ID): 11.87948561973063
Kleibergen–Paap LM (underidentification): 17.49057554484897

The identification confidence can be enhanced with further diagnostics. The Kleibergen–Paap LM statistic verifies the identification of the model with a value of 17.49 $(p = 0.0006)$. The Kleibergen–Paap rk Wald F statistic is aligned with strength under clustered errors, with a value of 11.88. The Cragg–Donald Wald F statistic is colossal (6,967), and suggests very strong identification under the i.i.d. assumption, though clustered inference constitutes the KP statistic the appropriate benchmark. Therefore, the instruments are reasonably strong, and the weak IV-robust tests can be relied on for the inference. Principally, the Anderson–Rubin

Wald test provides the weak IV-robust evidence that the instrumental variable has explanatory power in the structural equation since it rejects the null of no effect at the 1% level $(F = 9.94, p < 0.001)$.

Several hospital-type indicators exhibit strong negative coefficients in the first stage. The doctors working at Type B, D, E1, and E2 hospitals are subject to substantially fewer visits per day than those in the reference category. For instance, Type D hospitals have nearly 48 fewer visits per doctor $(p < 0.001)$, and Type E hospitals have 66–68 fewer $(p < 0.001)$. These results manifest the structural and institutional distinctions that are smaller or specialized facilities having fewer patients per physician.

The instrumented coefficient on doctor visits is 0.0031 $(se = 0.0011, p = 0.004)$ in the second stage. This indicates that each extra patient seen per doctor per day increases the underdiagnosis probability by 0.31 percent. Specifically, a 10-patient increase raises the underdiagnosis risk by more than 3%, a significant rise considering the baseline diagnostic error rates.

Other covariates behae as expected. Repeat MRI same day attains a negative but only marginal effect $(p = 0.062)$, likely showing the complex interactions between repeat pattern and diagnostic accuracy. The severity score is positively significant $(p < 0.001)$, indicating that the greater the underdiagnosis risk more severe the patients face. Women have significantly lower underdiagnosis risk $(coef = -0.068, p < 0.001)$. Hospitals of Type D, E1, and E2 hospitals are significantly more prone to higher underdiagnosis compared to the baseline category, consistent with institutional heterogeneity in diagnostic capacity.

Capacity pressure impairs diagnostic accuracy as strongly asserted by the Doctor–Demand–Stress IV results. The IV approach isolates exogenous variation in workload, unlike naïve OLS estimates that are subject to bias due to selective sorting, self-referral, or unobserved congestion. The causal effect aligns with the economic models of physicians working under time pressure (Clemens and Gottlieb, 2014) and empirical findings that fast-paced working environments undermine diagnostic care (Smith-Bindman et al., 2008; Baker et al., 2008). Our results are also aligned with the imaging efficiency literature in Türkiye. Başar et al. (2019) and Martella et al. (2023) assert that repeated procedures and resource constraints hurt the diagnostic quality. In policy terms, these findings suggest that scheduling practices, physician staffing, and holiday surge management are critical levers for reducing diagnostic error. In conclusion, scheduling practices, physician staffing, and holiday surge management are major factors for improving the diagnostic process.

5.2.2. IV Model 2: MRI Quality and Repeat MRI Instrument

The study proposes to approximate the diagnostic quality of institutions to mitigate these problems from the hospital-level repeat MRI rates of 2019-2023. The resultant factor is denoted as $Low\_MRI\_Quality_{ij}$. This factor is capable of predicting the institutional diagnostic quality thanks to the stable work culture and the equipment qualifications. It is strongly correlated to the repeat MRIs, but remarkably stays exogenous to the individual patient outcomes. In the earlier observations. This enables us to focus on the core factors such as long-run equipment quality and institutional work culture.

First-stage results: According to Table 9, the first-stage regression illustrates the strength and efficiency of the proposed factor. The coefficient of 0.029 $(p < 0.001)$ implies that the poor-diagnosis institutions in later years already exhibited higher repeat MRI rates in the study period.

Table 9. IV Model of MRI Quality

	(1) Same-Day Repeat MRI			(2) Underdiagnosis (R-code=1)		
	b	se	p	b	se	p
Low MRI Quality	0.0294***	0.0046	0.000			
Same-Day Repeat MRI				1.7206	0.9931	0.083
MRI Utilization per Device						
Reference: *Low*						
Medium	-0.0111**	0.0040	0.005	0.0815**	0.0266	0.002

Table 9. (continued)

High	-0.0125*	0.0056	0.027	0.1195*	0.0504	0.018
Hospital Type						
Reference: *A1:*	0.0000	.	.	0.0000	.	.
A2	-0.0059	0.0032	0.069	-0.0724***	0.0190	0.000
B	-0.0065	0.0035	0.064	-0.0587**	0.0216	0.007
Central hospital	0.0135	0.0093	0.147	-0.0904	0.0463	0.051
Bed Turnover Rate	-0.0006***	0.0001	0.000	0.0005	0.0003	0.068
Referral Degree	-0.0001**	0.0001	0.008	0.0006	0.0003	0.086
Doctor's Total Daily Visits	0.0002*	0.0001	0.024	-0.0007*	0.0003	0.019
Patient Age	0.0001	0.0001	0.329	0.0006***	0.0002	0.000
Patient Gender Reference: *Male*	0.0000	.	.	0.0000	.	.
Female	-0.0003	0.0029	0.928	-0.0190	0.0113	0.094
Severity score	0.0001*	0.0000	0.013	0.0002*	0.0001	0.037
Day Category						
Reference: *Workday*						
Vacation Day	0.0059	0.0057	0.303	0.0157	0.0198	0.427
Table 9 (continue)						
Time Category						
Reference: *Morning (05:00-10:59)*						
Afternoon (11:00-15:59)	0.0012	0.0033	0.727	0.0177	0.0096	0.067
Evening/Night (16:00-04:59)	0.0011	0.0040	0.789	0.0770***	0.0193	0.000
Constant	0.0298***	0.0072	0.000	-0.0159	0.0220	0.471
Observations	574843.00			574843.0000		
R-squared	0.0241			-0.2639		
Adj. R-sq.	0.0241			-0.2640		
F-statistic	31.6478			27.7419		

Standard errors clustered by hospital
Instruments (excluded): Low MRI Quality
Kleibergen–Paap rk Wald F (weak-ID): 39.70868351768426
Kleibergen–Paap LM (underidentification): 4.476514762018013
Clusters: 49 RMSE (2SLS): .3260682208406964

The F-statistic for excluded instruments is 39.71 $(p < 0.001)$, significantly greater than the conventional Staiger–Stock threshold of 10. Furthermore, the Kleibergen–Paap Wald F of 39.71 exceeds the 10% maximal IV size threshold (16.38), as well as the 5% threshold. This confirms that weak instruments are not a significant concern.

The Kleibergen–Paap LM statistic $(\chi^2 = 4.48, p = 0.034)$ addresses the concerns on possible underidentification, indicating the identification of the model. The Cragg–Donald Wald F statistic further confirms the relevance under i.i.d. assumptions with an extremely large value $(\approx 1{,}196)$. These statistics together prove the power and the validity of the Low MRI Quality.

Repeat MRI is also predicted by several covariates. High bed turnover decreases repeat MRIs $(coef = -0.0006, p < 0.001)$, implying that more efficient institutions can avoid repeating the diagnostic process. Additionally, repeats correlate with severity and the doctor visits, which reflect the complexity and congestion effects. Medium and high utilization per device reduces repeats, illustrating the marvel of economies of scale and standardized imaging protocols.

The effect of the repeat MRI on underdiagnosis is revealed by the IV estimation. The Repeat MRI Same Day coefficient is 1.72 $(se = 0.99, p = 0.083)$, which can be interpreted as each additional repeat increases the probability of underdiagnosis by 1.72 percent for compliers whose repeat behavior is driven by machine quality. The magnitude is substantial and clinically relevant, although marginal significance requires 10%.

Weak-IV-robust tests lead to mixed evidence. The Stock–Wright LM test is significant $(\chi^2 = 4.66, p = 0.03)$; however, the Anderson-Rubin Wald test does not reject the null at the 10% level $(F = 2.65, p = 0.11)$. These results shall be evaluated with caution, considering the small number of hospital clusters (49). Yet, the positive association is robust because of the strength of the first stage and the consistent estimates across different specifications.

In line behaviour is observed with other covariates. The evidence of the Doctor IV model suggests that higher throughput worsens diagnostic quality since doctor visits enter negatively $(coef = -0.00074, p = 0.019)$. As the underdiagnosis is more likely for more severe patients, the severity score has a positive effect $(p = 0.037)$. Type A2 and B hospitals are exposed to a lower underdiagnosis risk. The effects of fatigue and institutional quality show themselves in the significant rise of underdiagnosis probability $(coef = 0.077, p < 0.001)$ during the evening/night shifts.

Accordingly, the major factors for diagnosis failures are revealed by the MRI-Quality IV results: technology and personnel. The factor machine quality clearly illustrates that hospitals with poor MRI utilization both repeat more scans and are subject to higher rates of underdiagnosis.

Literature also supports our findings. Chen et al. (2012) report that repeat imaging is often caused by poor machine quality or inadequate imaging protocols. Smith-Bindman et al. (2008) raises concern about inefficiency by pointing out the more frequent imaging without proportional quality improvement. Jahanmehr et al. (2019) posit that low-quality equipment contributes to issues in the diagnostic process. In addition, the results are aligned with the inefficiency and quality-adjusted capacity literature. Overdiagnosis and repeat imaging indicate both resource waste and clinical harm, as argued by Brodersen et al. (2018). This harm translates directly into higher underdiagnosis, making machine quality a cause of diagnostic failure.

Thus, the MRI-IV analysis emphasizes the importance of the investment in diagnostic infrastructure and personnel well-being. The physician workload (Doctor IV) and the MRI repeat behaviour can capture the personnel and technological quality deficits persisting over time. The models imply that better workforce management and upgrading imaging environments are both required to mitigate underdiagnosis.

To summarize:

i) The underperforming institutions can be identified by monitoring the repeat MRI rates as a quality indicator.
ii) Poor machine quality and diagnostic errors can be remedied with systematic equipment upgrades and maintenance,
iii) Both fatigue and poor-quality imaging exacerbate risk; thus, evening and night shifts require additional measures.

In conclusion, the MRI Quality IV demonstrates that repeat imaging increases underdiagnosis risk, driven by technological and institutional deficits. This complements the Doctor IV model, showing from a broad perspective that human and technological bottlenecks jointly impair the diagnostic process.

5.3. Summary of Key Findings

To conclude Chapter 5, the main conclusions and their implications are summarized below:

5.3.1. Human Capacity Factors

Human capacity factors are based on the number of patients a physician examines daily, and also look at the impact of patient demographics and disease severity on diagnostic results.Baseline logistic regression found that physicians who see a high number of patients per day make fewer underdiagnoses. However, the IV model, constructed with instrumental variables time and day categories, suggests that physician workload creates an endogenous effect; it found the opposite result: higher physician workload diagnostic accuracy decreases, resulting in underdiagnosis. This result yielded the expected result from the first hypothesis of human capacity strain. In addition, this finding is consistent with numerous studies in the literature showing that physicians under heavy workloads often shorten consultation times due to time pressure, and in settings where decision fatigue increases the likelihood of negligence, resulting in diagnostic errors (Weber and Schneider, 2014; Ivanovic et al., 2023; Rosen et al., 2024).

In addition, auxiliary staff availability further moderates human capacity. First, later hours of the day, particularly evening and night shifts outside of work hours, when doctors are likely facing fewer staff, suggest that increased time pressure and decision-making fatigue lead to decreased diagnostic accuracy and increased R-code diagnoses. This is consistent with studies indicating increased medical error rates due to fatigue accumulation during late hours. Similarly, weekends and holidays also increase the likelihood of underdiagnosis. Because hospitals operate with fewer staff on these days, individual workload increases. Again, the likelihood of R-codes increases with time capacity and fatigue. This is consistent with research in the healthcare literature on the "weekend effect," which suggests that service delivery deteriorates under resource constraints.

Patient demographic characteristics also have a significant impact on diagnostic accuracy. Diagnostic accuracy decreases in older patients and as disease severity increases, leading to underdiagnosis. An interesting finding in the study is that female patients experience fewer cases of underdiagnosis compared to males. In conclusion, human capacity factors have revealed that physician decision-making processes are not isolated but depend on fatigue, workload, and patient characteristics.

5.3.2. Institutional Capacity Factors

Institutional capacity factors represent hospital type and central hospital variables, representing the structural organization of hospitals, and the MRI utilization rate per device, bed turnover, and referral degree rate, representing efficiency.

The results of the logistic and IV models first revealed that the probability of R codes differs across hospitals with structural differences. Smaller or less specialized hospitals, such as D, E1, and E2 hospitals, generally have fewer physician visits per physician, yet they exhibit a higher probability of R codes compared to higher-tier hospitals. This highlights the role of organizational complexity and specialization in shaping diagnostic behaviour. Furthermore, MRI utilization rates vary across hospitals. However, baseline models indicate that the utilization rate per device is not the sole determinant of underdiagnosis in situations where hospital capacity variables are not fully integrated. Bed turnover and referral grade rate, combined with other variables such as the number of patients per physician, MRI quality, and repeated MRIs, are insufficient to explain underdiagnosis in baseline models. However, the IV model showed that hospitals with high turnover rates experienced increased underdiagnosis. Again, underdiagnosis increased as referral grade increased.

5.3.3. Machine Capacity Factors

The two key variables representing machine capacity are the number of repeat MRIs performed on the same day and a low-quality MRI indicator, derived from the repeat MRI status of patients who underwent scans between 2019 and 2023. The IV model results indicate that when MRI quality is poor, the frequency of same-day repeat scans increases, thereby raising the probability of an R-code assignment. In addition, high MRI utilization per device further elevates the likelihood of an R code. These findings support a technological strain mechanism: poor-quality MRI environments resulting from excessive use generate more repeats and amplify diagnostic uncertainty.

Crucially, the fact that MRI quality measured in 2019–2023 significantly predicts R-code assignments in 2015–2016 suggests a persistent structural weakness in the healthcare system rather than a short-term fluctuation. Economically, this persistence reflects allocative inefficiency, resources are consumed through redundant imaging without commensurate improvements in diagnostic accuracy. Moreover, the results empirically validate the concerns raised by the Turkish Radiology Association in its 2018 report, which pointed to physician shortages, insufficient radiologist capacity, uncontrollable patient demand, and defensive medicine as systemic drivers of excessive test intensity. The present findings demonstrate that these institutional and behavioral pressures are not merely descriptive observations but are quantifiable and persistent contributors to diagnostic inefficiency in Türkiye.

CHAPTER 6

CONCLUSION

This thesis has investigated the challenge of increasing diagnostic precision in the Turkish health system by examining the determinants of underdiagnosis under three factors of diagnostic capacity constraint: human, institutional, and technological. To this end, both logistic regressions and instrumental variable (IV) models were analysed across a nationwide patient-level dataset during the 2015-2016 period to disentangle the confounding factors of underdiagnosis. Thus, the study uncovers a fundamental paradox within Türkiye’s diagnostic system: while the expansion of capacity enhances access to diagnostic services, it may concurrently undermine diagnostic accuracy, suggesting that underdiagnosis arises from system-level constraints and incentives.

Chapter 2 examines the structural and financial reforms implemented in the Turkish healthcare system since the Health Transformation Program and the capacity expansions implemented in response to these reforms. Structural changes under the program included the introduction of a universal health insurance system, a performance-based payment system for physicians, and the expansion of roles in the private sector. Performance-based payment systems aimed to increase efficiency but often forced clinicians to see more patients in their limited time. These reforms focused on the breadth of care rather than the depth of care, leading to capacity imbalances in diagnostic capacity at the human, mechanical, and institutional levels. Then, those diagnostic capacity factors are defined according to the Turkish healthcare system, and changes in capacity from 2002 to 2023 are statistically demonstrated. In this context, Türkiye's diagnostic capacity demonstrates a paradoxical trend. While the number of doctors per capita has been increasing over the years, the intense workload of doctors indicates that this capacity remains insufficient.

Chapter 3, grounding the study in Arrow's uncertainty theory of medical care framework, provides an understanding of how underdiagnosis in Türkiye is analysed under four distinct capacity constraints. In this context, it first acknowledges that diseases are inherently uncertain and that this uncertainty creates an asymmetrical relationship between the doctor (agent) and the patient (principal), where patients lack sufficient information about diagnosis and treatment and therefore must rely on the doctor's judgment. Second, within this asymmetrical relationship, physician decision-making may be influenced by time and resource constraints, financial incentives, or concerns about missed diagnoses and malpractice lawsuits. In this context, physicians request more MRI scans, but under time and resource constraints, overuse of MRI technologies may exacerbate underdiagnosis rather than diagnosis. Particularly in situations where patients are fully covered by insurance, mutual demands between physicians and patients may increase, creating a moral hazard in the system.

This study developed a multidimensional empirical strategy in Chapter 4, using comprehensive nationwide healthcare data (2015-2016 and 2019-2023). First, underdiagnosis is defined as R-coded diagnoses (ICD-10 codes R00–R99, indicating signs, symptoms, or abnormal findings without a definitive diagnosis). In the Turkish healthcare system, ICD-10 codes are mandatory in diagnostic reports. Therefore, clinicians actively use ICD codes in their examinations since these codes are used only when a healthcare provider cannot or does not make a specific diagnosis. Then, three different logistic regressions are set up to understand human, institutional, and machine capacity on underdiagnosis. Then, two complementary instrumental variable (IV) models were designed: one to address endogeneity in physician workload, and another to isolate the role of machine quality in repeat imaging. These methods allowed for causal inference beyond the limitations of standard regression models.

The empirical analysis in Chapter 5 yields several consistent findings: First, human capacity factors show that higher daily patient visits increase the probability of underdiagnosis. This is consistent with evidence on decision fatigue and cognitive overload in medical care. Second, Time-of-day and vacation effects matter. Underdiagnosis increases in afternoon/evening shifts and during holiday periods, in line with literature on “weekend effects” and resource shortages. These patterns confirm that temporal bottlenecks exacerbate both human fatigue and machine usage pressures. Third, under the institutional capacity, hospital type, and bed turnover rates influence diagnostic precision, with higher-tier institutions performing better. MRI utilization per machine reflects organizational load balancing rather than individual machine strain, underscoring that diagnostic accuracy is shaped by both institutional efficiency and equipment availability. Finally, machine capacity shows that the MRI Quality IV model demonstrated that poor scanner quality, proxied by historical repeat MRI rates, significantly contributes to underdiagnosis.

In this context, two different IV model strategies are used: The first one is to determine workload strain at the physician level. Instruments based on time category and day category are highly predictive of workload. According to doctor IV results, the cumulative effect is substantial; an increase of 50 patients per doctor per day in Turkish hospitals raises underdiagnosis odds by more than 15 percentage points. This quantitatively confirms that overburdened physicians contribute meaningfully to diagnostic uncertainty. The second IV model is based on the low-quality MRI indicator, generated from repeat MRI visits between 2019 and 2023, strongly predicted patient repeat MRI use in the first stage in 2015-201616 $(\beta = 0.0294, SE = 0.0046, p < 0.001)$. In the second stage, repeat MRIs were positively associated with underdiagnosis $(\beta = 1.72, SE = 0.99, p = 0.083)$, suggesting that low-quality scans directly increase patients' likelihood of receiving R codes. Accordingly, repeat MRIs often arise from poor imaging quality or doctor fatigue, but the strong first-stage results confirm that machine quality is a credible driver. This highlights the importance of technology standards and maintenance.

Taken together, these two models show that both technological strain and human strain independently and significantly increase the probability of underdiagnosis. The effect sizes are nontrivial: moving from low to high workload environments increases R-code probability by double-digit margins, and operating under low imaging quality regimes produces up to a 12% higher likelihood of inconclusive outcomes. These findings strongly validate the thesis hypothesis: systemic strain in Türkiye's health system, whether in people, machines, or institutions, directly translates into higher rates of underdiagnosis. Moreover, the finding that MRI quality measured in 2019–2023 significantly predicts underdiagnosis already in 2015–2016 points to a persistent structural inefficiency in the Turkish healthcare system rather than a short-term fluctuation. This long-run imprint of poor quality resonates with the concerns raised by the Turkish Radiology Association (2018), which documented how insufficient physician and radiologist capacity, uncontrollable patient demand, and defensive medicine practices collectively drive excessive test intensity. The empirical results of this thesis thus not only quantify but also substantiate the structural problems identified by the TRA, showing that these systemic bottlenecks continue to undermine diagnostic precision over time.

This thesis makes three contributions: First, it contributes to extending Arrow's uncertainty framework by operationalizing underdiagnosis through R-code probabilities. The persistence of high R-code rates despite abundant MRI availability confirms Arrow's prediction that uncertainty in medicine cannot be eliminated simply by increasing service volume. Instead, asymmetric information and misaligned incentives lead to both overuse (supplier-induced demand for MRIs) and underdiagnosis (R-codes), which coexist as two sides of the same uncertainty coin. Thus, it enables making empirical clarity on how uncertainty manifests in imaging-intensive health systems. Second, it demonstrates endogeneity of physician workload. By using time-of-day and vacation instruments, the thesis provides robust evidence that workload is not exogenous but causally linked to diagnostic performance. Finally, by leveraging historical repeat MRI rates (2019–2023) as a proxy for scanner quality in 2015–2016, this thesis highlights the long-run institutional imprint of poor technology on diagnostic accuracy.

In this context, the thesis has shown that the rapid expansion of diagnostic imaging capacity in Türkiye, driven by the Health Transformation Program (HTP), has delivered unprecedented access to diagnostic imaging (MRI) services, but at the cost of emerging quality challenges. The econometric results demonstrate that underdiagnosis (proxied by R-codes) is systematically shaped by three capacity dimensions: human, institutional, and technological. Heavy doctor workloads and the unavailability of auxiliary staff increase the probability of underdiagnosis; peripheral hospitals with weaker institutional structures display systematically higher diagnostic uncertainty; and low MRI quality and machine overuse amplify diagnostic errors. These findings confirm that while access has been broadened, the system still struggles to convert volume into diagnostic precision.

Thus, the Turkish case illustrates a fundamental trade-off: expanding the number of machines, visits, and examinations improves access, but unless human resources, institutional processes, and technological quality keep pace, diagnostic accuracy may suffer. Simply put, more scans do not automatically translate into better diagnoses. The evidence of underdiagnosis despite massive capacity growth highlights the diminishing returns of access expansion when quality management lags behind. This is a critical lesson for health systems globally: achieving universal access to technology without embedding quality safeguards risks entrenching inefficiencies and diagnostic errors.

The analysis also shows that existing incentive structures may inadvertently worsen these trade-offs. Fee-for-service and performance-based payment schemes, while successful in expanding throughput, reward volume rather than accuracy. Providers facing high patient loads and inadequate support staff are incentivized to maximize service counts, even at the risk of diagnostic shortcuts. In Arrow's (1963) framework, this reflects how uncertainty and asymmetric information allow incentives and organizational design to shape outcomes in unintended ways. Unless carefully calibrated, provider payment systems may encourage overuse while simultaneously raising the probability of underdiagnosis.

These contributions bridge health economics, medical decision-making, and health system studies by showing that diagnostic precision emerges from the joint constraints of men and machines. The results carry several actionable implications on behalf of those reforms that should aim to realign incentives and balance access with quality. In this context, first, performance incentives must be recalibrated. Payment schemes should reward not only service volume but also diagnostic accuracy by incorporating quality metrics into performance evaluations. Second, investments in human resources are as critical as those in technology. Ensuring the presence of adequate radiology staff, technicians, and support personnel during evenings, nights, and weekends can alleviate time-pressure-induced underdiagnosis. Physician workload should be managed by reducing daily patient volumes and redesigning shift schedules to allow for adequate diagnostic focus. Third, technological quality must be upgraded and maintained. Regular maintenance, accreditation, and performance monitoring of MRI devices are necessary to prevent quality deterioration under heavy usage. Repeat MRI scans should be systematically monitored as a key performance indicator. Finally, institutional resilience should be strengthened. Smaller and peripheral hospitals must be supported through tele-radiology networks, centralized review mechanisms, and well-defined referral pathways to mitigate regional disparities in diagnostic accuracy.

Some limitations should be acknowledged. First, the analysis relies on administrative data and does not capture physician decision-making processes directly. Second, instruments, while strong, cannot fully eliminate unobserved heterogeneity across institutions. Third, underdiagnosis measured via R-codes is only one dimension of diagnostic quality. Future research could expand by linking underdiagnosis to long-term patient outcomes, studying interactions between artificial intelligence diagnostics and human/machine capacity, and conducting cross-country analyses to compare Türkiye's experience with other middle-income health systems undergoing rapid diagnostic expansion.

Turkish experience contributes to the broader global debate on the future of diagnostic systems. As universal health coverage (UHC) and the WHO's Essential Diagnostics List emphasize equitable access, Türkiye's case highlights that access alone is insufficient. Without robust attention to incentives, human resources, and technology quality, expansion risks entrench inefficiency and diagnostic gaps. Future research should extend this approach to other modalities (CT, mammography) and examine patient outcomes of underdiagnosis. Last but not least, the thesis has shown that diagnostic precision in Türkiye's health system is constrained simultaneously by human effort, machine reliability, and institutional structure. Thus, the underdiagnosis is a systemic result of capacity imbalances and exemplifies a paradox at the heart of modern healthcare: The system celebrates more machines, more scans, more throughput, yet patients ultimately need more certainty, not just more activity. In Arrow’s (1963) framework, the health system is a theatre of uncertainty, where information gaps and incentive misalignments determine outcomes. Thus, more capacity does not dissolve uncertainty; under asymmetric information and volume-linked incentives, it can reallocate uncertainty. The findings here reveal a dichotomy between human expertise and machine capacity or, put differently, between quantity and quality. The overall conclusion is that improving diagnostic accuracy requires a balanced policy response. Türkiye can achieve sustainable improvements in diagnostic care only by addressing the three dimensions of capacity together: managing workload effectively, ensuring technological quality, and strengthening institutional structures with a focus on quality rather than quantity.

REFERENCES

Agarwal, R., Bergey, M., Sonnad, S., Butowsky, H., Bhargavan, M., & Bleshman, M. H. (2010). Inpatient CT and MRI utilization: Trends in the academic hospital setting. *Journal of the American College of Radiology, 7*(12), 949–955.

Akçay, G., & Gül, Ö. (2022). Çocuk acil servisine yeniden başvuran vakaların demografik ve klinik özelliklerinin değerlendirilmesi. *Sağlık Akademisyenleri Dergisi*, *9*(3), 280-285.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*(3), 488–500. https://doi.org/10.2307/1879431

Alam, R., Cheraghi-Sohi, S., Panagioti, M., Esmail, A., Campbell, S., & Panagopoulou, E. (2017). Managing diagnostic uncertainty in primary care: a systematic critical review. *BMC Family Practice*, *18*(1), 79.

Ali, K. J., Ehsan, S., Tran, A., Haugstetter, M., & Singh, H. (2024). Diagnostic excellence in the context of climate change: a review. *The American journal of medicine*, *137*(11), 1035-1041.

American College of Radiology. (2017). *ACR CT Accreditation Program requirements.* Reston, VA: ACR.

Aran, M., & Rokx, C. (2014). *Turkey on the way of universal health coverage through the Health Transformation Program (2003-2013)* (No. 91326). The World Bank.

Arıkan, A. & S. Çınarlı & F.Ş. Aykar & A. Sayan (2017), “Attitudes of Medical Malpractice in Pediatric Surgery”, The Journal of Pediatric Research, 4(3), 117.

Armao, D., Semelka, R. C., & Elias Jr, J. (2012). Radiology's ethical responsibility for healthcare reform: tempering the overutilization of medical imaging and trimming down a heavyweight. *Journal of magnetic resonance imaging*, *35*(3), 512-517.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review, 53*(5), 941-973.

Atun, R., Aydın, S., Chakraborty, S., Sümer, S., Aran, M., Gürol, I., ... & Akdağ, R. (2013). Universal health coverage in Turkey: enhancement of equity. *The Lancet*, *382*(9886), 65-99.

Baker, L. C., Atlas, S. W., & Afendulis, C. C. (2008). Expanded use of imaging technology and the challenge of measuring value. *Health Affairs, 27*(6), 1467–1478.

Balogh et al., National Academies of Sciences, Engineering, and Medicine. (2015). *Improving diagnosis in health care*. Washington, DC: The National Academies Press. https://doi.org/10.17226/21794

Başar, D., Özen, İ. C., Öztürk, S., Gümeler, E., Akata, D., & Çiftçi, A. Ö. (2019). *The case of MRI repetitions with direct and indirect cost implications: Evidence from Hacettepe University Hospitals*. Sosyoekonomi, 27(41), 253-262.

Bernardy, M., Ullrich, C. G., Rawson, J. V., Allen Jr, B., Thrall, J. H., Keysor, K. J., ... & Mabry, M. R. (2009). Strategies for managing imaging utilization. *Journal of the American College of Radiology, 6*(12), 844-850.

Bertakis, K. D., Azari, R., Helms, L. J., Callahan, E. J., & Robbins, J. A. (2000). Gender differences in the utilization of health care services. *Journal of Family Practice, 49*(2), 147–152.

Bion, J., Aldridge, C., Beet, C., Boyal, A., Chen, Y. F., Clancy, M., ... & Lilford, R. (2021). Safety and quality of weekend care in hospital: a mixed-methods evaluation. In *Increasing specialist intensity at weekends to improve outcomes for patients undergoing emergency hospital admission: the HiSLAC two-phase mixed-methods study*. NIHR Journals Library.

Brekke, K. R., Holmas, T. H., Monstad, K., & Straume, O. R. (2017). Do treatment decisions depend on physicians' financial incentives? *Journal of Public Economics, 155*, 74-92.

Brinjikji, W., Luetmer, P. H., Comstock, B., et al. (2015). Systematic literature review of imaging features of spinal degeneration in **asymptomatic** populations. *AJNR American Journal of Neuroradiology, 36*(4), 811–816.

Brownlee, S., Chalkidou, K., Doust, J., Elshaug, A. G., Glasziou, P., & Nagpal, S. (2017). Evidence for overuse of medical services around the world. *The Lancet*, *390*(10090), 156-168.

Chassin, M. R., & Galvin, R. W. (1998). The urgent need to improve health care quality: Institute of Medicine National Roundtable on Health Care Quality. *Jama*, *280*(11), 1000-1005.

Chou, S. C., Chang, Y. S. C., Chen, P. C., Schuur, J. D., & Weiner, S. G. (2022). Hospital occupancy and its effect on emergency department evaluation. *Annals of emergency medicine*, *79*(2), 172-181.

Christensen, E. F., Larsen, T. M., Jensen, F. B., Bendtsen, M. D., Hansen, P. A., Johnsen, S. P., & Christiansen, C. F. (2016). Diagnosis and mortality in prehospital emergency patients transported to hospital: a population-based and registry-based cohort study. *BMJ open*, *6*(7), e011558.

Clemens, J., & Gottlieb, J. D. (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review, 104*(4), 1320–1349.

CMS. (2024). *ICD-10-CM official guidelines for coding and reporting FY 2024 (April 1, 2024–September 30, 2024): Updated April 1, 2024*. U.S. Department of Health and Human Services. https://www.cdc.gov/nchs/data/icd/ICD-10-CM-Coding-Guidelines-FY2024-April.pdf

Colla, C. H., Mainor, A. J., Hargreaves, C., Sequist, T., & Morden, N. E. (2017). Interventions aimed at reducing use of low-value health services: a systematic review. *Medical Care Research and Review, 74*(5), 507–550.

Emanuel, E. J., & Fuchs, V. R. (2008). The perfect storm of overutilization. *Jama*, *299*(23), 2789-2791.

Emery, D. J., Shojania, K. G., Forster, A. J., Mojaverian, N., & Feasby, T. E. (2013). Overuse of magnetic resonance imaging. *JAMA internal medicine*, *173*(9), 823-825.

European Society of Radiology (ESR) communications@ myesr. org. (2014). Renewal of radiological equipment. *Insights into Imaging*, *5*(5), 543-546.

Evans, R. G. (1974, June). Supplier-induced demand: some empirical evidence and implications. In *The economics of health and medical care: proceedings of a conference held by the International Economic Association at Tokyo* (pp. 162-173). London: Palgrave Macmillan UK.

Feigin, K. (2023). Quality assurance in mammography: An overview. *European Journal of Radiology, 161,* 110803. https://doi.org/10.1016/j.ejrad.2023.110803

Gadeka, D. D., & Esena, R. K. (2020). Quality of care of medical imaging services at a teaching hospital in Ghana: clients' perspective. *Journal of Medical Imaging and Radiation Sciences, 51*(1), 154-164.

Galdas, P. M., Cheater, F., & Marshall, P. (2005). Men and health help-seeking behaviour: Literature review. *Journal of Advanced Nursing, 49*(6), 616–623.

Gates, M., Wingert, A., Featherstone, R., Samuels, C., Simon, C., & Dyson, M. P. (2018). Impact of fatigue and insufficient sleep on physician and patient outcomes: a systematic review. *BMJ open, 8*(9), e021967.

Gaughan, J., Siciliani, L., Gravelle, H., & Moscelli, G. (2020). Do small hospitals have lower quality? Evidence from the English NHS. *Social Science & Medicine, 265*, 113500.

Graber, M. L. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety, 22*(Suppl 2), ii21–ii25. https://doi.org/10.1136/bmjqs-2012-001615

Green, C. A., & Pope, C. R. (1999). Gender, psychosocial factors and the use of medical services: A longitudinal analysis. *Social Science & Medicine, 48*(10), 1363–1372.

Gregersen, R., Villumsen, M., Mottlau, K. H., Maule, C. F., Nygaard, H., Rasmussen, J. H., ... & Petersen, J. (2024). Acute patients discharged without an established diagnosis: risk of mortality and readmission of nonspecific diagnoses compared to disease-specific diagnoses. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 32*(1), 32.

Gürsoy, G. (2014). Türkiye'de sağlık hizmetlerinin mekânsal eşitsizliği: Görünür olan ve olmayan. *Toplum ve Hekim, 29*(1), 39–45.

Hagiwara, A., Fujimoto, K., Nakazawa, M., & Aoki, S. (2020). Variability and standardization of quantitative imaging. *Magnetic Resonance in Medical Sciences, 19*(4), 283–296. https://doi.org/10.2463/mrms.rev.2019-0125

Hanna, T. N., Zygmont, M. E., Peterson, R., Theriot, D., Shekhani, H., Johnson, J.-O., & Krupinski, E. A. (2018). The effects of fatigue from overnight shifts on radiology search patterns and diagnostic performance. *Journal of the American College of Radiology, 15*(12), 1709–1716.

Hendee, W. R., Becker, G. J., Borgstede, J. P., Bosma, J., Casarella, W. J., Erickson, B. A., ... & Wallner, P. E. (2010). Addressing overutilization in medical imaging. *Radiology, 257*(1), 240-245.

Hershey, N. (1972). The defensive practice of medicine: Myth or reality? *Milbank Memorial Fund Quarterly: Health and Society, 50*(1), 69–98.

Huang, Q., et al. (2023). *Equity and efficiency of MRI services in Henan Province, China: A Gini coefficient and data envelopment analysis.*

Hutubessy, R. C., Hanvoravongchai, P., & Tan-Torres Edejer, T. (2002). Diffusion and utilization of magnetic resonance imaging in Asia. *International Journal of Technology Assessment in Health Care, 18*(3), 690–704.

Ivanovic, V., Broadhead, K., Beck, R., Chang, Y. M., Paydar, A., Biddle, G., ... & Qi, L. (2023). Factors associated with neuroradiology diagnostic errors at a large tertiary-care academic medical center: a case-control study. *American Journal of Roentgenology, 221*(3), 355-362.

Jahanmehr, N., Bigdeli, A. S., Salari, H., Mokarami, H., KhodaKarim, S., & Damiri, S. (2019). Analyzing inappropriate magnetic resonance imaging (MRI) prescriptions and resulting economic burden on patients suffering from back pain. *International Journal of Health Planning and Management, 34*(4), e1437–e1447.

Jenkins, H. J., Hancock, M. J., Maher, C. G., French, S. D., & Magnussen, J. S. (2016). Understanding patient beliefs regarding the use of imaging in the management of low back pain. *European Journal of Pain, 20*(4), 573–580.

Jenkins, H. J., Kongsted, A., French, S. D., Jensen, T. S., Doktor, K., Hartvigsen, J., & Hancock, M. (2022). Patients with low back pain presenting for chiropractic care who want diagnostic imaging are more likely to receive referral for imaging: A cross-sectional study. *Chiropractic & Manual Therapies, 30*, 16.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360.

Johnson, H. M., et al. (2021). *Missed & Delayed Diagnosis of Cardiovascular Disease in Women.* PLOS ONE.

Kasalak, Ö., Alnahwi, H., Toxopeus, R., Pennings, J. P., Yakar, D., & Kwee, T. C. (2023). Work overload and diagnostic errors in radiology. *European Journal of Radiology, 167*, 111032.

Katz, D. A., Williams, G. C., Brown, R. L., Aufderheide, T. P., & Bogner, M. (2005). Emergency physicians' fear of malpractice in evaluating patients with possible acute cardiac ischemia. *Annals of Emergency Medicine, 46*(6), 525–533. https://doi.org/10.1016/j.annemergmed.2005.06.034

Kayadibi, Y., Aladağ Kurt, S., Tekcan Şanlı, D. E., Esmerer, E., & Şanlı, A. N. (2021). Türkiye'nin Doğusundaki Hasta Popülasyonunun Radyolojik Tetkikler Hakkındaki Bilgi Düzeyi [Knowledge level of the patient population in Eastern Turkey on radiological tests]. *Journal of Anatolian Medical Research, 6*(2), 30–37.

Kessler, D. P., & McClellan, M. B. (1996). Do doctors practice defensive medicine? *Quarterly Journal of Economics, 111*(2), 353–390.

Keteepe-Arachi, T., & Sharma, S. (2017). Cardiovascular disease in women: Understanding symptoms and risk factors. *British Journal of Hospital Medicine, 78*(3), 162–165

Kılıçaslan, İ., Bozan, H., Oktay, C., & Göksu, E. (2005). Türkiye'de acil servise başvuran hastaların demografik özellikleri. *Türkiye Acil Tıp Dergisi*, *5*(1), 5-13.

Kizildag Yirgin, I., Koyluoglu, Y. O., Seker, M. E., Ozkan Gurdal, S., Ozaydin, A. N., Ozcinar, B., ... & Aribal, E. (2022). Diagnostic performance of AI for cancers registered in a mammography screening program: a retrospective analysis. *Technology in cancer research & treatment*, *21*, 15330338221075172.

Kuhlmann, E., & Annandale, E. (2012). *The Palgrave handbook of gender and healthcare*. Palgrave Macmillan.

Landrigan, C. P., Rothschild, J. M., Cronin, J. W., Kaushal, R., Burdick, E., Katz, J. T., Lilly, C. M., Stone, P. H., Lockley, S. W., Bates, D. W., & Czeisler, C. A. (2004). Effect of reducing interns' work hours on serious medical errors in intensive care units. *The New England Journal of Medicine*, *351*(18), 1838–1848.

Larson, D. B., Johnson, L. W., Schnell, B. M., Salisbury, S. R., & Forman, H. P. (2011). National trends in CT use in the emergency department: 1995–2007. *Radiology*, *258*(1), 164-173.

Lavaste, K. (2025). Overuse of medical imaging and effects of payer–provider integration: Quasi-experimental evidence from Finland. *Health Economics Review, 15*(1), 4.

Link, H., & Sack, R. (2009). Managing the effects of shift work in medicine. *AMA Journal of Ethics (Virtual Mentor), 11*(11), 876–881.

Liu, P., Yang, Y., & Cheng, J. (2022). Gender differences in medical errors among older patients and inequalities in medical compensation compared with younger adults. *Frontiers in Public Health*, *10*, 883822.

Lysdahl, K. B., & Hofmann, B. M. (2009). What causes increasing and unnecessary use of radiological investigations? A survey of radiologists' perceptions. *BMC Health Services Research*, *9*(155), 1-8.

Marengoni, A., et al. (2011). Aging with multimorbidity: A systematic review of older adults. *Journal of Clinical Epidemiology*, *64*(3), 243-256

Marshall, T. L., Milliren, C. E., Kahn, R., & Bhattacharya, J. (2022). Diagnosis Code and Health Care Utilization Patterns Among Children Hospitalized With Uncertain Diagnoses. *Pediatric Quality & Safety*, *7*(5), e711.

Martella, M., Lenzi, J., & Gianino, M. M. (2023). Diagnostic technology: Trends of use and availability in a 10-year period (2011–2020) among sixteen OECD countries. *Healthcare*, *11*(14), 2078.

McGuire, T. G. (2000). Physician agency. *Handbook of health economics*, *1*, 461-536.

Ministry of Health, Republic of Turkey (n.d.). *Hastane hizmetlerinde temel kavramlar* [PDF]. Retrieved from https://dosyasb.saglik.gov.tr/Eklenti/961/0/ekpdf.pdf

Ministry of Health, Republic of Türkiye. (2023). *MEDAS Health Statistics Database* [Data set]. Ankara: Ministry of Health. Retrieved September 15, 2025, from https://medas.saglik.gov.tr/

Ministry of Health, Republic of Türkiye. (2024). *Health Statistics Yearbook 2023*. Ankara: General Directorate of Health Information Systems, Ministry of Health. Retrieved September 15, 2025, from https://sbsgm.saglik.gov.tr/

Müskens, J. L., Kool, R. B., van Dulmen, S. A., & Westert, G. P. (2022). Overuse of diagnostic testing in healthcare: a systematic review. *BMJ quality & safety*, *31*(1), 54-63.

Muzio, D., Dionisi, M., Simone, D., Cianfrocca, E., Fabbian, F., Barbiero, G., ... & Giannetta, N. (2019). Can nurses' shift work jeopardize the patient safety? A systematic review. *European Review for Medical and Pharmacological Sciences*, *23*(10), 4507-4519.

Nesemeier, R., Jones, S., Jacob, K., Cash, E., & Goldman, J. (2023). Trends in emergency department computed tomography usage for diagnosis of peritonsillar abscess. *Ear, Nose & Throat Journal*, *102*(1), NP35-NP39.

Newman-Toker, D. E. (2014). A unified conceptual model for diagnostic errors: underdiagnosis, overdiagnosis, and misdiagnosis. *Diagnosis*, *1*(1), 43-48.

O'Brien, R., Hunt, K., & Hart, G. (2005). "It's caveman stuff, but that is to a certain extent how guys still operate": Men's accounts of masculinity and help seeking. *Social Science & Medicine*, *61*(3), 503–516.

OECD (2008). *Saglık Sistemi İncelemeleri-Türkiye*. ISBN 978–975–590–282–1-© OECD ve IBRD/Dünya Bankası.

OECD (2017). *Health at a Glance 2017: OECD Indicators*. Organisation for Economic Co-operation and Development, Paris.

OECD (2023). *Health at a Glance 2023: OECD Indicators*. OECD Publishing. (Data on diagnostic imaging utilization).

OECD. (2025). *The economics of diagnostic safety* (OECD Health Working Papers No. 176). OECD Publishing. https://doi.org/10.1787/6e0ed50b-en

Office of Technology Assessment. (1994). *Impact of legal reforms on medical malpractice costs*. Washington, DC: U.S. Government Printing Office.

Official Gazette of the Republic of Türkiye. (2006, June 16). *Regulation on additional payment in private hospitals*, Resmî Gazete, No. 26126, pp. 4–5. Retrieved from https://www.resmigazete.gov.tr/eskiler/2006/06/20060616-1.htm

Official Gazette of the Republic of Türkiye. (2022, February 10). *10 Şubat 2022 tarihli Resmî Gazete (Sayı: 31746).* Retrieved from https://www.resmigazete.gov.tr/eskiler/2022/02/20220210-1.htm

Okçan, N., & Çakar, U. (2015). Türkiye sağlık politikalarının dönüşümü: Sağlıkta Dönüşüm Programı üzerine bir değerlendirme. *Ankara Üniversitesi SBF Dergisi, 70*(1), 101–126.

Owusu, N., Park, H. W., & Krishnamurthy, P. (2021). Factors influencing daily quality assurance measurements in MRI systems. *Journal of Applied Clinical Medical Physics, 22*(11), 82–92. https://doi.org/10.1002/acm2.13468

Özen, İ. C., Başar, D., Öztürk, S., Gümeler, E., Akata, D., & Çiftçi, A. Ö. (2022). Not just about machines: Analysis of MRI quality determinants in the Turkish health system context. *The International journal of health planning and management*, *37*(2), 902-912.

Özmen, P. & S. Sahin & M. Cetin & Y.Z. Turk (2015), "A Medical Malpractice Research Conducted in Turkey, Causes and the Role of Media", Journal of Archives in Military Medicine, 3(2).

Patel, A. G., Pizzitola, V. J., Johnson, C. D., Zhang, N., & Patel, M. D. (2020). Radiologists make more errors interpreting off-hours body CT studies during overnight assignments as compared with daytime assignments. *Radiology, 297*(2), 374–379. https://doi.org/10.1148/radiol.2020201558

Path, Inc.. (2023). *2023 annual report*. PATH. https://www.path.org/who-we-are/annual-report-2023/

Piras, G. N., Baggio, U., Tortorella, V., Sorace, L., Negro, F., & Marinelli, E. (2024). The weekend factor: the effect on healthcare quality. A medico-legal, retrospective study. *La Clinica Terapeutica*, *175*(4).

Rodriguez-Jareño, M. C., Demou, E., Vargas-Prada, S., Sanati, K. A., Škerjanc, A., Reis, P. G., ... & Serra, C. (2014). European Working Time Directive and doctors' health: a systematic review of the available epidemiological evidence. *BMJ open*, *4*(7), e004916.

Rosen, M., Kilcullen, M. P., Zhang, A., Sharma, R., Xiao, C., & Bass, E. B. (2024). Fatigue and Sleepiness of Clinicians Due to Hours of Service. In *Making Healthcare Safer IV: A Continuous Updating of Patient Safety Harms and Practices [Internet]*. Agency for Healthcare Research and Quality (US).

Rylands-Monk, F. (2014, May 25). *Turkish radiology prepares to step into spotlight*. AuntMinnie.com. https://www.auntminnie.com/clinical-news/mri/article/15607927/turkish-radiology-prepares-to-step-into-spotlight

Sanderson, E., & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of econometrics*, *190*(2), 212-221.

Semin, S., Demiral, Y., & Dicle, O. (2006). Trends in diagnostic imaging utilization in a university hospital in Turkey. *International journal of technology assessment in health care*, *22*(4), 532-536.

Sharma, S., Traeger, A. C., Reed, B., Hamilton, M., O'Connor, D. A., Hoffmann, T. C., Bonner, C., Buchbinder, R., & Maher, C. G. (2020). Clinician and patient beliefs about diagnostic imaging for low back pain: A systematic qualitative evidence synthesis. *BMJ Open, 10*(8), e037820.

Singh, H., Meyer, A. N., & Thomas, E. J. (2014). The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ quality & safety*, *23*(9), 727-731.

Skinner, T. R., Scott, I. A., & Martin, J. H. (2016). Diagnostic errors in older patients: a systematic review of incidence and potential causes in seven prevalent diseases. *International journal of general medicine*, 137-146.

Smith-Bindman, R., Miglioretti, D. L., & Larson, E. B. (2008). Rising use of diagnostic medical imaging in a large integrated health system. *Health Affairs, 27*(6), 1491–1502.

Sonğur, C., & Top, M. (2016). Regional clustering of medical imaging technologies. *Computers in Human Behavior, 61*, 333-343.

Sosyal Güvenlik Kurumu (SGK). (2011). *Universal health insurance system in Turkey*. Ankara: Social Security Institution.

Sparkes, S. (2015). *Turkey: Performance-based payment reform*. World Bank Health, Nutrition, and Population Discussion Paper.

Staiger, D. O., & Stock, J. H. (1994). Instrumental variables regression with weak instruments.

Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear IV regression.

Studdert, D. M., Mello, M. M., Sage, W. M., DesRoches, C. M., Peugh, J., Zapert, K., & Brennan, T. A. (2005). Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *Journal of the American Medical Association*, *293*(21), 2609–2617.

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York: Ballantine Books.

Tatar, M., Mollahaliloğlu, S., Sahin, B., Aydin, S., Maresso, A., & Hernández-Quevedo, C. (2011). Turkey. Health system review. *Health systems in transition*, *13*(6), 1-186.

Temiz, H., & Kara, T. (2025). Classification of medical imaging technologies: results from Türkiye. *BMC Health Services Research*, *25*(1), 847.

TTB (Turkish Medical Association) (2021). 5 Dakikada Hekimlik Yapılmaz, Sağlık 5 Dakikaya Sığmaz! [Press release]. Turkish Medical Association Central Council

Tung, M., Sharma, R., Hinson, J. S., Nothelle, S., Pannikottu, J., & Segal, J. B. (2018). Factors associated with imaging overuse in the emergency department: A systematic review. *American Journal of Emergency Medicine, 36*(2), 301–309.

Turkish Medical Association Ethical Committee (2009) Hekimlerin değerlendirmesi ile performansa dayalı ödeme. Ankara: Turkish Medical Association.

Turkish Radiology Association. (2018). *Radyolojik Tetkik Yoğunluğu, Tetkik Yoğunluğundan Kaynaklanan Problemlerin Analiz ve Çözüm Önerileri*.

Turkish Radiology Association. (2018). *Türkiye'de Radyoloji Hizmetleri Durum Raporu*. TÜRKRAD Yayınları.

Turkish Statistical Institute. (2024, April 18). *Health expenditure statistics, 2023*. TurkStat. https://data.tuik.gov.tr/Bulten/Index?p=Health-Expenditure-Statistics-2023-53561

U.S. Congress. (1992). *Mammography Quality Standards Act (MQSA)*, Pub. L. No. 102–539, 106 Stat. 3547.

Ülgü, M. M., & Birinci, Ş. (2023). An analysis of the distribution of patient diagnoses based on ICD codes in Turkey between 2016 and 2022: A descriptive study. *Ankara Medical Journal, 2023*(3), 336–343.

Uyğun, U., & Yardımcıoğlu, F. (2019). Evaluation of Health Expendıtures in Turkey and EU Countries A Cluster Analysis. *Ekonomik Ve Sosyal Araştırmalar Dergisi*, *15*(3), 401-416.

Uysal, Ö. Ü. B., & Semiz, Ö. Ü. T. (2022). Sağlık Hizmetlerinde Dijitalleşme ve Geleceği.

Vaughan, L., Edwards, N., Imison, C., & Collins, B. (2018). Rethinking acute medical care in smaller hospitals. *London: Nuffield Trust*.

Verbrugge, L. M. (1985). Gender and health: An update on hypotheses and evidence. *Journal of Health and Social Behavior, 26*(3), 156–182.

Vujicic, M., Sparkes, S., & Mollahaliloglu, S. (2009). Health workforce policy in Turkey. *Washington, DC: World Bank.*

Walsh, B., Roberts, H. C., & Nicholls, P. G. (2011). Features and outcomes of unplanned hospital admissions of older people due to ill-defined (R-coded) conditions: retrospective analysis of hospital admissions data in England. *BMC geriatrics*, *11*(1), 62.

Wang, R. C., Kornblith, A. E., Grupp-Phelan, J., Smith-Bindman, R., Kao, L. S., & Fahimi, J. (2021). Trends in use of diagnostic imaging for abdominal pain in US emergency departments. *American Journal of Roentgenology*, *216*(1), 200-208.

Wen, J., Cheng, Y., Hu, X., Yuan, P., Hao, T., & Shi, Y. (2016). Workload, burnout, and medical mistakes among physicians in China: a cross-sectional study. *Bioscience trends*, *10*(1), 27-33.

Wise, J. (2022). Burnout among trainees is at all time high, GMC survey shows. *BMJ: British Medical Journal (Online)*, *378*, o1796.

Yüksel, A. (2020). Erişkin acil servise başvuran hasta profili, tanı kodları ile triaj yönünden değerlendirilmesi. *Anatolian Journal of Emergency Medicine*, *3*(2), 37-41.

Zabrodina, V., Dusheiko, M., & Moschetti, K. (2020). A moneymaking scan: Dual reimbursement systems and supplier-induced demand for diagnostic imaging. *Health Economics*, *29*(12), 1566–1585.

APPENDICES

A. TURKISH SUMMARY / TÜRKÇE ÖZET

Sağlık hizmetlerinde tanının kritik rolü önemlidir. MR ve BT gibi görüntüleme yöntemleri de dahil olmak üzere tanı, erken hastalık teşhisi ve daha iyi hasta sonuçları sağladığı için evrensel sağlık kap samına (UHC) ulaşmak için hayati önem taşımaktadır. 2023 yılında Dünya Sağlık Örgütü, yetersiz hizmet alan nüfusların genellikle zamanında hizmet alamaması nedeniyle tanı erişiminin ve kalitesinin UHC ve eşitlik için hayati önem taşıdığını vurgulamıştır (PATH, 2023). Bu bağlamda, Türkiye'nin 2003 yılında başlatılan Sağlık Dönüşüm Programı (SDP), sağlık hizmetlerine erişimi önemli ölçüde artırmıştır. Ancak reformlar, sistem kapasitesi konusunda kapasite artışı yaşanırken kalifiye konusunda endişeleri de beraberinde getirmiştir. Kişi başına düşen doktor ve tıbbi görüntüleme MRI sayısı yetersiz kalmakta, lakin arza karşı talep çok yüksek seviyelerdedir. Buradan hareketle, bu tez, Türkiye'de tıbbi görüntüleme teknolojisi içerisinde MRI seçerek yetersiz tanıyı, teknolojik, insani ve kurumsal faktörlerin hastaların kesin bir tanı konulmadan sağlık sisteminden ayrılmalarına nasıl yol açtığını araştırmaktadır. Amaç, Türkiye'nin sağlık reformları ve tanı kapasitesini üç faktör altında sentezlemek, teorik bir çerçeve sunmak ve Türkiye çapında hasta seviyesinde ve hastane seviyesindeki geniş çaplı verileri kullanarak ampirik bulguları sonuçlamaktır.

Türk Sağlık Sisteminde Tanı Kapasitesi

2003 yılından bu yana, Türkiye'nin Sağlık Dönüşüm Programı (SDP), sağlık hizmetlerini Genel Sağlık Sigortası'na (GSS) doğru yeniden yapılandırmıştır. Reformlar, birleşik bir GSS sistemi oluşturmuş ve aile hekimliği modelini 2010 yılına kadar tüm illere yaygınlaştırmıştır. Ayrıca, SDP kapsamında yeni Kamu-Özel

Ortaklığı (PPP) hastane projeleri kapsamında inşa edilen "Şehir Hastaneleri" kapasiteyi artırmıştır. Örneğin, bu projeler kamu hastaneleri bünyesinde destek hizmetleri (radyoloji gibi) sağlamak üzere özel işletmecileri bünyesine katmıştır. Bunlara ek olarak, doktorların maaşları artırılarak hizmet başına ücret teşvikleri getirilmiştir. Hizmet hacmi için prim ödemelerinin getirilmesi, üretkenliği ödüllendirmeyi amaçlamakla birlikte, aynı zamanda hizmet kullanımını artırmak için finansal teşvikler de yaratmıştır. Böylece, performansa dayalı ödeme sistemi bakım hizmetini hızlandırırken niceliğe vurgu yapmanın kalite ve güvenliği azaltabileceği endişesi pek çok çalışma tarafından ortaya konulmuştur (Aran ve Rokx, 2014) Bu sonuçlar, SDP'nın hizmet kullanımına olan erişimi büyük ölçüde genişlettiğini ancak diyagnostik kapasite faktörlerini zorladığını göstermektedir.

Türkiye'de Tanı Kapasitesi Çerçevesi

Bu tez, Türkiye'de yetersiz tanıyı değerlendirmek için bir tanı kapasitesi çerçevesi geliştirmektedir. Çerçeve, tanı kapasitesini üç ayrı kategoride değerlendirilmektedir:

İnsan Kapasitesi: Teşhis ve tedavi sürecinde tanı kararını şekillendiren klinisyenlerin ve tamamlayıcı personelin iş yükünü ve dağılımını ifade eder. Yüksek hasta yoğunluğu ve sınırlı konsültasyon veya raporlama süresine sahip ortamlarda, klinisyenler daha fazla bilişsel yük ile karşı karşıya kalmaktadır. Bu da belirsiz kodlama ve geciken/kaçan tanılar olasılığını yükseltmektedir. Çalışma bağlamımızda bu mekanizma, gün-içi ve hafta-içi zaman baskısı örüntüleri ve günlük başvuru yoğunluğu üzerinden yakalanmakta; beklenti, daha yüksek iş yükü ve geç saat randevularının R-kod atama olasılığını artırması yönündedir. Böylece yetersiz tanı, bireysel yetkinlikten ziyade, kıt insan dikkatinin operasyonel kısıtlar altında nasıl tahsis edildiğinin bir çıktısı olarak ele alınır.

Kurumsal Kapasite: Sağlık hizmetinin örgütsel bağlamını, teşvik yapısını ve hasta akış mimarisini içerir. Hastane tipi (A1–E2), merkezi hastane rolü, sevk merkeziyeti, yatak devir hızı kurumsal kapasiteyi belirleyen faktörlerdir.

Teknolojik Kapasite: Tıbbi görüntüleme ekipmanlarının (MR, BT, MAM) bulunabilirliğini ve kalitesini ifade eder. Türkiye, Sağlıkta Dönüşüm Programı'ndan bu yana yüksek teknoloji ekipman sayısını önemli ölçüde genişletmiştir. 2002 ile 2023 yılları arasında MR cihazı sayısı 58'den 1001'e çıkmıştır. Ancak, bu artış kişi başına cihaz seviyesinde OECD ülkelerinin gerisindeyken kişi başına düşen görüntüleme sayısında ve cihaz başına yapılan görüntüleme sayılardan en yüksek oranlara sahiptir.

Teorik Çerçeve

Bu tez, yetersiz tanıyı açıklamak için Arrow'un (1963) tıbbi bakımdaki belirsizlik teorisini temel almaktadır. Arrow'a göre sağlık hizmetleri, belirsizlik ve bilgi asimetrisi nedeniyle mükemmel piyasa varsayımlarından önemli ölçüde sapar. Hastalar ne zaman hastalanacaklarını, hangi tedavinin etkili olacağını veya bir doktorun hizmet kalitesini kolayca bilemezken; doktorlar hastaların ihtiyaçları ve teknik olanaklar hakkında çok daha fazla bilgiye sahiptir. Bu asimetri, sağlık hizmetlerinin yalnızca piyasa mekanizmalarıyla işlemesini imkânsız kılar; bu nedenle etik kurallar, mesleki lisanslar, sigorta sistemleri ve yasal düzenlemeler ortaya çıkar.

Arrow ayrıca, piyasa dışı müdahalelerin (örneğin, ücret yapıları veya hastane düzenlemeleri) teşhis ve tedavi kararlarını çarpıtabileceğini vurgulamıştır. Türkiye bağlamında bu çerçeve özellikle anlamlıdır: Hastalar çoğu zaman ileri görüntülemeye ihtiyaçları olup olmadığını bilemez ve doktorların yönlendirmelerine güvenmek zorundadır. Doktorlar ise performansa dayalı ödeme sistemleri veya malpraktis riskleri nedeniyle teşhis süreçlerinde piyasa dışı teşvikler ve kısıtlamalarla karşı karşıya kalmaktadır. Dolayısıyla, Arrow'un belirsizlik kavramı Türkiye'de MR kullanım davranışını anlamak için güçlü bir teorik zemin sunmaktadır.

HTP sonrasında doktorlar tarama başına daha fazla gelir elde edebilmekte, ancak aynı zamanda fatura denetimleri ve malpraktis şikayet mekanizmaları gibi daha sıkı düzenlemelerle karşı karşıya kalmaktadır. Burada önemli olan Arrow'un işaret ettiği teşvikler ile kısıtlamaların nasıl etkileştiğidir. Bu çerçeve altında, tez hem teorik (Arrow'un belirsizlik yaklaşımı) hem de ampirik (lojistik regresyon, IV modelleri) yöntemlerle Türkiye'de MR kullanımının yetersiz tanıya nasıl yol açtığını incelemektedir. Bulgular, insan kurumsal ve makine kapasiteleri zorlandığında yetersiz tanı riskinin arttığını göstermektedir.

Literatür Taraması

Bu bölüm, tanısal görüntüleme (MR) kullanımını artıran arz-tarafı teşvikler, defansif tıp ve hasta beklentileri üzerine uluslararası ve Türkiye odaklı bulguları kısaca sentezlemektedir.

Arz kaynaklı talep ve finansal teşvikler: Bilgi asimetrisi, hekimlerin hastalar adına karar verirken finansal teşviklere tepki vermesine yol açarak arz kaynaklı talebi (SID) mümkün kılmaktadır (Evans, 1974; McGuire, 2000). Ücret-başına-hizmet ve ikili ödeme düzenlerinde hizmet miktarının artırılması sık görülür (Clemens & Gottlieb, 2014; Zabrodina et al., 2020). Türkiye'de SUT/SGK kapsamı ve performans ek ödemesi çıktı sayısını ödüllendirir; bu yapı, klinik gereklilik sınırının ötesinde kullanım riskini artırmaktadır (Başar et al., 2019; OECD, 2017).

Defansif Tıp ve Malpraktis Baskısı: Defansif tıp, olası dava riskini azaltma motivasyonuyla gereksiz testlerin istenmesi şeklinde ortaya çıkar ve görüntüleme bu davranışın başlıca araçlarındandır. Mediko-legal risk algısı, özellikle acil servis bağlamında "kaçırmama" saikiyle gereksiz görüntülemeyi teşvik eder; hekimler kimi zaman dava riskini azaltmak ve belirsizliği yönetmek için görüntüleme ister (Brownlee et al., 2017; Tung et al., 2018). Türkiye'de 2004/2012 reformlarıyla malpraktis başvurularının artması, klinik pratikte daha temkinli/defansif kalıpları güçlendirmiştir (Özmen et al., 2015; Arıkan et al., 2017; Başar et al., 2019).

Hasta Beklentileri ve Talep Faktörleri: Türkiye evrensel sağlık sigorta kapsamına ulaştıkça, Arrow'un 'ahlaki tehlike' altında hastaların tedavi kapsamında tanıya olan talebi artmaktadır. Daha iyi bilgilendirilmiş ve teknoloji tabanlı bakım bekleyen hastalar, küçük şikayetler için bile görüntüleme talep edebilir. Bu hasta odaklı talep, daha yüksek görüntüleme oranlarına yol açabilir. Örneğin, hekim ziyaretlerindeki artış, muhtemelen görüntüleme yönlendirmelerini de içeren sağlık hizmeti arayışının arttığını göstermektedir. Hasta talebiyle ilgili literatür, daha iyi eğitimli veya daha varlıklı hastaların genellikle daha fazla görüntüleme hizmeti aldığını göstermektedir.

Literatürün işaret ettiği mekanizmalar doğrultusunda, R-kodlarıyla ölçülen tanısal belirsizliğin iş yükü, kalite ve teşvik kanalları üzerinden nasıl oluştuğunu test etmek üzere aşağıdaki hipotezleri kuruyoruz.

H1: İnsan Kapasitesi ve Zaman Baskısı: Belirsizlik kuramına dayanarak, hekimlerin gün içinde çok sayıda hastaya bakması ve yardımcı personel yetersizliği tanısal süreci kısaltarak yetersiz tanı olasılığını artırır. Yüksek hasta ziyaretleri ve tarama hacimleri (performans-ödeme hedeflerine bağlı olarak) sağlık çalışanlarını yorabilir. Akşam/gece/hafta sonu vardiyalarında uzmanlık ve destek hizmetleri azaldığı için bu etki güçlenir. Çalışma, doktor başına toplam günlük ziyaret sayısı gibi metriklerin eksik tanı olasılığıyla pozitif korelasyon göstermesini beklemektedir.

H2: Teknolojik Baskı: MR Aşırı Kullanımı ve Görüntü Kalitesi: Aşırı MR kullanımı klinik akıl yürütmeden sapmaya ve “cevap vermeyen” incelemelerin artmasına, düşük kalite (acele çekim, yetersiz protokol/yorum) ise belirsiz ya da tekrar çekim gerektiren sonuçlara yol açabilir. Bu nedenle makine başına çekim sayısı çok yüksek olan veya aynı gün tekrar MR oranı fazla olan hastanelerde R-kodu daha sık beklenir. Beklenti: MR yoğunluğu ve tekrar oranları arttıkça R-kod (yetersiz tanı) olasılığı yükselir (Hendee et al., 2010; Brinjikji et al., 2015).

H3: Kurumsal Baskı: Hastane Kademesi ve Kaynaklar: Daha düşük kademeli (örn. Tip C/D/E) hastanelerde uzmanlık ve tanısal altyapı kısıtları, karmaşık vakalarda hekimin R-kodu ile “üst merkeze yönlendirme” stratejisine başvurma olasılığını artırır. Buna karşılık üniversite/eğitim-araştırma ve merkez/referral hastaneler kapasite sayesinde bu baskıyı daha iyi absorbe edebilir. Beklenti: kademeler yükseldikçe R-kodu oranı azalır (OECD, 2017).

H4: Hasta Özellikleri: Son olarak, belirli hasta faktörleri yetersiz tanıyı etkileyebilir. Tez, yaş, cinsiyet ve hastalık şiddet derecesini yardımcı değişkenler olarak ele almaktadır. Çalışma, yetersiz tanının, yaş ve hastalık şiddeti arttıkça artacağını varsaymaktadır.

Veri ve Metodoloji

Analiz, yetersiz tanı ve belirleyicilerini incelemek amacıyla birden fazla Türk sağlık veri setinden yararlanmaktadır: Çalışmanın ana verisi 2015-2016 dönemine ait Hasta Düzeyinde Hastane Verileridir. Bu talep verisi, demografik özellikler, yatış tanıları (ICD kodları) ve görüntüleme talepleri hakkında kapsamlı bilgi sunmaktadır. Bu veri üzerinden, açık bir tanı kaydı bulunmayan vakalar "R kodları" yardımıyla yetersiz tanı göstergesi olarak tanımlanmıştır. Örneklem, farklı hastanelerden on binlerce hasta karşılaşmasını içermektedir. Bir diğer veri seti, görüntüleme teknolojilerinin güncel trendlerini gösteren 2019-2023 dönemine ait hastane düzeyinde görüntüleme kullanım verileridir**.** Sağlık Bakanlığından toplanan bu seri, beş yıl boyunca hastanelerin yıllık MRI kullanımı ve başvuru sayılarının takip etmektedir. Üçüncü veri seti ise 2019-2023 yıllarına ait hasta takip MRI verileridir. Bu kohort, MRI çektiren ve kısa bir süre içerisinde (90 gün) tekrar görüntülemeye alınan hastaları kapsamaktadır. Bu veri setlerine göre regresyon analizlerinde kullanılan başlıca değişkenler şunlardır:

Yetersiz Tanı Göstergesi: Bir vaka R kodu aldıysa = 1, aksi halde = 0.

Aynı Gün Tekrar MR: Aynı gün birden fazla MR çekimini gösteren ikili değişken.

Düşük Kaliteli MR: 2019-2023 MR başvuruları yapan hasta veri setinden tekrar MR çekim durumuna göre türetilmiş düşük kalite göstergesi

Makine Başına MR Kullanım Seviyesi: Her hastanedeki MR ünitelerinin kullanım yoğunluğunu ölçmektedir.

Toplam Doktor Günlük Ziyaretleri: Tedavi eden hekimlerin günlük ortalama ayakta hasta sayısı.

Hastane Türü ve Düzeyi: Merkez hastanesi veya hastane rolüne göre alt kademeli kurum ayrımı.

Hasta Eş değişkenleri: Yaş, cinsiyet ve hastalık şiddeti

Ampirik Metodoloji

Bu çalışma, Türkiye'de tanısal belirsizliğin belirleyicilerini insan, kurumsal ve makine kapasitesi eksenlerinde sınamak için iki tamamlayıcı yaklaşım uygular: i) R-kodu ile sonuçlanma olasılığını açıklayan temel lojistik regresyonlar, ii) hekim günlük ziyaret sayısı ve MR kalite/kullanımı göstergelerindeki endojenliği gidermek üzere kurulan araç değişkenli (IV) modeller. Bağımlı değişken R_Code_Prob olup, başvurunun R kodu (1) ya da spesifik tanı (0) ile sonlanmasını gösterir.

İlk olarak, lojistik çerçevede modelleme aşamalı biçimde ilerler ve her aşama bir kapasite boyutunu tanımlı bir şekilde devreye alarak katsayıların kararlılığını sınar. İlk olarak, İnsan kapasitesi modeli arz tarafı göstergeleri (aynı gün tekrar MR, hekim günlük ziyaret sayısı, MR makine başına kullanım seviyesi) ve hasta/başvuru kontrolleriyle tahmin edilir. İkinci olarak, kurumsal kapasite modeli temel yapıya Sevk Derece Oranı, Yatak Devir Hızı ve Düşük MR Kalitesi göstergelerini ekleyerek kurumların akış-yönetimi ve kalite altyapısını içselleştirir. Üçüncü olarak, makine kapasitesi modeli teknoloji/altyapı boyutunu BT cihaz başına kullanım seviyesi ve MAM cihaz başına kullanım seviyesini ekleyerek ayrıntılandırır. Böylece yalnızca MR yoğunluğunun değil, BT ve mamografinin kurum içi kapasite kullanımının da R-kodu ile ilişkisini bağımsız ve eşzamanlı olarak tahmin edilir. İkili sonuç değişkeni için tüm aşamalarda lojistik tahmin yapılır; katsayılar marjinal etkiler üzerinden yorumlanır ve hastane düzeyi kümelerde küme-robust standart hatalar kullanılır. Bu kurgu, MR-odaklı teknolojik yoğunluk ve kalite göstergelerinin yanı sıra BT ve mamografinin kapasite kullanımını da modelleyerek, teknolojik yoğunluk/kalite ile R-kodu arasındaki ilişkiyi daha temiz biçimde izole eder.

Endojenlik riski iki kanaldan ele alınır. İlk olarak, hekim günlük ziyaret sayısı gözlenemeyen hekim niteliği, vaka karması veya aynı gün talep şoklarıyla ilişkili olabilir; bu, basit lojistik tahminlerde yanlılığa yol açar. Bu nedenle, zaman temelli dışsal dalgalanmadan yararlanan araçlar (gün/saat kategorileri) kullanılarak hekim iş yükünde dışsal varyasyon izole edilir. İkinci olarak, aynı gün tekrar MR ve MR kalite kullanımı göstergeleri de hem klinik karmaşıklık hem de kurum içi uygulama farklılıklarıyla içsel olabilir; burada, cihaz/yorum kalitesini yansıtan kalite ortamı göstergeleri (hastane düzeyinde tekrar oranları) araç olarak kullanılır. IV tahminleri iki aşamalı (2SLS) çerçevede yürütülür; ikincil doğrulama olarak ikili sonuçlar için ivprobit/kontrol-fonksiyonu yaklaşımlarıyla tutarlılık sınanır.

Ekonometrik Modeller ve Sonuçlar

İlk lojistik regresyon, insan kaynağı modelidir. Bu modeldeki temel değişken olan doktorun günlük hasta sayısı, R koduyla tanı konulma olasılığıyla negatif ve anlamlı bir ilişki göstermektedir ($p < 0.05$). Ancak bu sonucun doğrudan doktorun daha fazla çalıştıkça daha iyi tanı koyduğu şeklinde yorumlanması yanıltıcı olabilir. Çünkü sistemde iyi doktorlara yönelim *(self-referral)* ve vaka seçilimi gibi mekanizmalar vardır. Daha fazla hasta gören doktorlar, genellikle daha kolay veya belirgin vakaları görme eğiliminde olabilir; karmaşık vakalar ise daha sakin zamanlara ya da uzman merkezlere yönlendirilebilir. Ayrıca, toplumda *"iyi doktor"* olarak bilinen hekimler hem daha fazla hasta çekebilir hem de daha başarılı tanı koyabilir. Bu nedenle, gözlemlenen bu ilişki nedensel değil, muhtemelen seçilim yanlılığı ve ters nedensellik içeriyor olabilir. Bu endojenlik sorunu, ilerleyen bölümde araç değişkenli modeller ile test edilecektir. Ayrıca günün saati ve tatil günleri, tanı belirsizliğini anlamlı biçimde artırmaktadır. Özellikle akşam/gece saatlerinde ve tatil günlerinde R kodu alma olasılığı ciddi şekilde yükselmektedir. Bu da destek personelinin sınırlı olduğu zamanlarda tanı konulmasının zorlaştığını göstermektedir.

İkinci lojistik regresyon, hastane kapasitesini vurgulamaktadır. Bu modelde hastanenin sevk derecesi, merkezi olup olmaması ve hastane tipi gibi kurumsal faktörler incelenmiştir. Merkezi hastaneler ve yüksek düzeyde genel hastaneler (A2, B), tanı belirsizliğini anlamlı ölçüde azaltmaktadır. Bu, bu hastanelerin daha iyi tanı süreçlerine ve uzman kadroya sahip olduğunu göstermektedir.

Üçüncü lojistik model, teknoloji kapasitesini ölçmektedir. Bu modelde MR cihaz kalitesi, MR çekim yoğunluğu ve diğer görüntüleme (BT, mamografi) cihazlarının kullanımı analiz edilmiştir. Düşük kaliteli MR cihazları, aynı gün içinde tekrar MR çekimi yapılması, ve yoğun şekilde kullanılan MR cihazları, tanı konulamama ihtimalini ciddi ölçüde artırmaktadır ($p < 0.01$ seviyelerinde). Örneğin, yüksek MR kullanım düzeyi olan hastanelerde, R kodu alma ihtimali neredeyse 5 kat artmaktadır. Bu, cihaz başına düşen hasta sayısının çok yüksek olması durumunda tanı kalitesinin düştüğünü göstermektedir. İlginç olarak, yüksek BT ve mamografi kullanım düzeyleri, tanı belirsizliğini azaltmaktadır. Bu durum, bu tür cihazlarda kullanılan daha standart protokoller ve süreçlerin etkili tanı sağlamasıyla ilişkili olabilir.

Araç Değişkenli (IV) Regresyonlar

Bu bölümde, bazı temel değişkenlerin tanı belirsizliği (R kodu ile tanı) üzerindeki etkilerini nedensel olarak ölçmek amacıyla iki farklı IV (araç değişkenli) regresyon modeli uygulanmıştır. Önceki lojistik modellerde, doktorun günlük hasta sayısı ve aynı gün tekrarlanan MR çekimi gibi değişkenler, sonuçlarla istatistiksel olarak ilişkiliydi; ancak bu değişkenlerin hem hasta seçimi, hem doktor niteliği, hem de sistem içi yönlendirmeler (self-referral) gibi nedenlerle endojen olabileceği değerlendirildi. Bu nedenle, söz konusu değişkenlerin etkisini dışsal şoklarla izole edebilmek için zaman kategorileri (sabah/öğle/gece) ve tatil günü gibi zamanlama temelli değişkenler ile MR kalitesi gibi yapısal değişkenler, araç değişken (IV) olarak kullanılmıştır.

Doktor Talep ve İş Yükü IV Modeli: İlk modelde, doktorun bir günde gördüğü hasta sayısının tanı belirsizliği üzerindeki etkisi incelenmektedir. Bu değişken, doğrudan ölçüldüğünde seçilim yanlılığı taşıyabileceğinden, günün saati (sabah/öğle/gece) ve günün türü (tatil mi değil mi) gibi zaman temelli değişkenlerle araçlanmıştır. İlk aşama sonuçları, bu zamanlama değişkenlerinin doktorun hasta sayısını güçlü şekilde etkilediğini göstermektedir $(F\ istatistiği\ > 10)$. İkinci aşama sonuçlarında, doktorun günlük hasta sayısındaki artışın tanı belirsizliği ihtimalini anlamlı şekilde artırdığı görülmüştür ($p < 0.01$). Bu bulgu, önceki modellerdeki ters ilişkiyi düzelterek, yoğun iş yükünün gerçekten tanısal hata ve belirsizliğe yol açtığını göstermektedir. Bu durum, Arrow'un (1963) sağlıkta belirsizlik teorisiyle, ayrıca klinik karar kalitesini etkileyen zaman baskısı ve tükenmişlik literatürüyle örtüşmektedir.

MR Kalitesi ve Tanı Belirsizliği IV Modeli: İkinci IV modeli, aynı gün içinde tekrarlanan MR çekiminin, tanı belirsizliği üzerindeki etkisini analiz etmektedir. Bu değişkenin hasta karmaşıklığı ve doktor davranışıyla ilişkili olabileceği için, araç değişken olarak 2019–2023 dönemine ait hastane düzeyinde tanımlanan "düşük MR kalitesi" kullanılmıştır. İlk aşama sonuçları, düşük kaliteli MR ortamlarının tekrarlayan çekim ihtimalini güçlü şekilde artırdığını ortaya koymaktadır (F istatistiği ≈ 40). İkinci aşamada, aynı gün tekrarlanan MR çekimi ile tanı konulamama olasılığı arasında pozitif ve anlamlı bir ilişki saptanmıştır ($p \approx 0.08$). Bu bulgu, MR kalitesinin sadece teknik bir ayrıntı değil, klinik tanı süreçlerinin doğruluğunda belirleyici bir rol oynadığını göstermektedir. Ayrıca, cihaz başına düşen yüksek çekim sayısının da tanısal doğruluğu düşürdüğü görülmüştür.

Bu iki IV model birlikte değerlendirildiğinde, hem insan kapasitesi (doktor yükü) hem de teknolojik kapasite (MR kalitesi ve cihaz yoğunluğu) sağlık sistemindeki tanısal belirsizlikleri önemli ölçüde şekillendirmektedir. Doğrudan ilişkilere kıyasla IV modeller, daha güvenilir ve nedensel sonuçlar sunmakta; böylece sağlık politikalarında sadece kapasite artırımı değil, aynı zamanda kalite iyileştirme ve iş yükü dengelemesi gerektiğini ortaya koymaktadır.

Temel Bulguların Özeti

Tez, tüm modellerden diyagnostik kapasite faktörlerine göre temel bulguları özetlemektedir:

İnsan Kapasitesi Faktörleri: Hekimin günlük hasta sayısı, lojistik modellerde negatif ve anlamlı görünse de, bu ilişki büyük ölçüde seçilim yanlılığı ve iyi doktor etkisi ile açıklanabilir. Çünkü daha deneyimli ya da güvenilen doktorlara daha çok hasta yönelmekte ve bu doktorlar genellikle daha basit vakaları görebilmektedir. Bu nedenle değişkenin nedensel etkisini anlamak amacıyla IV yöntemi uygulanmıştır. Günün saati ve tatil günü gibi dışsal değişkenlerle araçlanan modelde, doktorun hasta sayısı arttıkça tanısal belirsizlik (R kodu olasılığı) da artmaktadır ($p < 0.01$). Bu sonuç, bilişsel yüklenme, zaman baskısı, ve hızlı karar alma zorunluluğu gibi faktörlerin tanı kalitesini olumsuz etkilediğini ortaya koymaktadır.

Günün saatine ve günün türüne ilişkin değişkenler (ör. sabah-öğle-akşam saatleri ve tatil günleri), yardımcı sağlık personeli kapasitesinin dolaylı göstergeleri olarak değerlendirilmiştir. Tüm modellerde, sabah saatlerine kıyasla öğleden sonra ve özellikle akşam/gece saatlerinde R kodu ile tanı alma olasılığı %30 ila %150 arasında artmaktadır. Aynı şekilde, resmî tatil günlerinde tanı belirsizliği anlamlı şekilde artmaktadır. Bu bulgular, laboratuvar, görüntüleme, danışman uzmanlık hizmetleri gibi destekleyici birimlerin erişilebilirliğinin sınırlı olduğu zamanlarda, doktorların daha fazla belirsizlikle karşılaştığını göstermektedir. Yardımcı personel eksikliği, tanının netleştirilememesine ve geçici R kodlarıyla sistemden çıkış yapılmasına yol açmaktadır.

Kurumsal Kapasite Faktörleri: Tezdeki modeller, hastanenin yapısal özelliklerinin tanı belirsizliği üzerinde belirleyici rol oynadığını ortaya koymaktadır. Özellikle hastane türü değişkeni, tüm lojistik ve IV modellerde güçlü etkiler göstermektedir. Merkezi hastaneler ve yüksek seviyeli genel hastaneler (A2, B tipi), tanı belirsizliğini anlamlı düzeyde azaltmaktadır. Bu hastaneler genellikle daha deneyimli kadrolara, uzmanlaşmış birimlere ve protokol temelli hizmet akışlarına sahiptir. Buna karşın, çevresel hastaneler (D, E1, E2) hem lojistik hem IV modellerde en yüksek R kodu oranlarına sahiptir. Örneğin, E2 tipi küçük entegre hastanelerde tanı konulamama ihtimali A1 tipi üniversite hastanelerine göre 4 kata kadar daha fazladır.

Bununla birlikte, bazı kurumsal göstergeler, sevk derecesi veya yatak devir hızı, bazı modellerde anlamlı çıkmamıştır. Bu durum, tanısal kapasite açısından belirleyici olanın sadece hastanenin fiziksel kapasitesi değil, aynı zamanda örgütsel süreçleri, kalite kontrol mekanizmaları ve multidisipliner çalışma yetkinliği olduğunu göstermektedir (Özen et al., 2022; Chen et al., 2012).

Teknolojik Kapasite: Modellerin önemli bir diğer bulgusu, tanı kalitesinin sadece insan veya kurum faktörlerine değil, aynı zamanda doğrudan kullanılan teknolojik altyapıya da bağlı olduğu yönündedir. Özellikle MR cihazlarının kalitesi, cihaz başına düşen çekim sayısı (yoğunluk) ve aynı gün yapılan tekrar çekimler, tanı belirsizliğini anlamlı şekilde artırmaktadır.

Lojistik modellerde düşük MR kalitesi olan hastanelerde R kodu alma ihtimali yaklaşık 2 kat daha fazladır. Bu bulgunun nedensel etkisini belirlemek amacıyla uygulanan MRI Kalitesi IV Modeli, bu ilişkinin gerçekten MR kalite eksikliklerinden kaynaklandığını göstermektedir. Düşük kaliteli cihazlarda çekilen MR'lar daha sık tekrarlanmakta ve bu tekrarlar tanının netleştirilememesine neden olmaktadır (Başar et al., 2019).

Ayrıca, MR cihazı başına düşen çekim sayısı yüksek olan hastanelerde tanı belirsizliği anlamlı şekilde artmaktadır. Örneğin, yüksek kullanım seviyelerinde R kodu ile tanı alma olasılığı 3 ila 5 kat daha fazladır. Bu durum, yoğunluk baskısının tanısal hassasiyeti düşürdüğünü ve doktorların "hızlı geçiş" yaparak belirsiz kodlara yöneldiğini göstermektedir (Smith-Bindman et al., 2008).

Genel olarak, ekonometrik kanıtlar, Türkiye'deki görüntülemede tanı konulmamasının çok faktörlü olduğu hipotezini desteklemektedir: insan ve makine kapasiteleri aşırı zorlandığında artar ve iyi kaynaklara sahip kurumsal ortamlarda azalır.

Sonuç

Bu tez, Türkiye'nin sağlık sisteminde 2000'li yılların başında başlatılan Sağlıkta Dönüşüm Programı (SDP) sonrasında tanısal kapasitenin gelişimini ve bunun tanısal doğruluk üzerindeki etkilerini incelemiştir. İkinci bölümde gösterildiği gibi, program sayesinde görüntüleme teknolojilerine erişim olağanüstü düzeyde artmış, özellikle MR ve BT cihazlarının sayısı ile hasta başına düşen çekim oranları on kata varan artışlar göstermiştir. Ancak bu niceliksel genişleme, beraberinde iş yükü, performans teşvikleri ve kapasite baskısı kaynaklı yeni sorunları da getirmiştir.

Üçüncü bölümde teorik çerçeve olarak Arrow'un (1963) belirsizlik teorisi kullanılmıştır. Arrow'un öne sürdüğü gibi, sağlık hizmetlerinde bilgi asimetrisi ve belirsizlik ortamı, sistemin teşvik ve kapasite yapıları üzerinden tanısal sonuçları doğrudan şekillendirmektedir. Bu çerçeve, Türkiye örneğinde gözlenen yüksek hacimli hizmet sunumunun, her zaman daha doğru tanılara dönüşmediğini; aksine, insan, kurumsal ve teknolojik kapasite kısıtları altında tanısal belirsizliğin artabileceğini açıklamaktadır.

Dördüncü bölümde tanımlanan hasta ve hastane düzeyindeki kapsamlı veri setleri, beş yıllık dönemi kapsayarak bu kapasite faktörlerini test etmeye imkân vermiştir. Veriler, hem doktor iş yükü hem de MR cihazlarının kalite ve yoğunluk göstergeleri açısından sistematik ölçümler içermektedir.

Beşinci bölümde sunulan ekonometrik analizler, bu faktörlerin etkilerini ortaya koymuştur. Lojistik regresyon sonuçları, üç kapasite boyutunda (insan, kurumsal, teknolojik) belirgin eğilimler göstermiştir: İnsan kapasitesinde, yüksek doktor iş yükü ve özellikle yardımcı sağlık personelinin az olduğu akşam/gece saatleri ve tatil günlerinde tanısal belirsizlik belirgin biçimde artmaktadır. Kurumsal kapasitede, merkezi ve daha üst düzey hastaneler (A2, B) tanı doğruluğunu artırmakta; küçük ve çevresel hastanelerde (D, E1, E2) ise belirsizlik riski ciddi ölçüde yükselmektedir. Teknolojik kapasitede, düşük kaliteli MR cihazları ve cihaz başına yüksek çekim yoğunluğu tanı belirsizliğiyle güçlü şekilde ilişkilidir; aynı gün yapılan tekrar çekimler de bu kalite eksikliklerinin doğrudan yansımasıdır.

Bu bulgular, IV modelleri ile desteklenmiştir. Araç değişkenli tahminler, önceki sonuçlardaki olası endojenlik sorunlarını gidererek iki temel nedensel çıkarımı güçlendirmiştir:

i) Yüksek doktor iş yükü, tanısal belirsizliği artırmaktadır.
ii) Düşük MR kalitesi, tekrar çekimleri tetikleyerek tanı doğruluğunu düşürmektedir.

Sonuç olarak, Türkiye’nin sağlık sisteminde tanısal kapasitenin yalnızca cihaz ve erişim sayısını artırmaktan ibaret olmadığı, aynı zamanda insan gücü, kurum yapısı ve kalite yönetimi ile desteklenmesi gerektiği ortaya çıkmıştır. Bu çalışma, niceliksel genişlemenin yanında kaliteyi ve kapasite dengesini merkeze alan bir yaklaşımın, hastaların tanı konulamadan kalma riskini azaltmada kritik olduğunu göstermektedir.

B. THESIS PERMISSION FORM / TEZ İZİN FORMU

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences	☐
Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences	☒
Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics	☐
Enformatik Enstitüsü / Graduate School of Informatics	☐
Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences	☐

YAZARIN / AUTHOR

Soyadı / Surname : Kaynak
Adı / Name : Mehtap Tuğba
Bölümü / Department : İktisat / Economics

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English): OF MEN AND MACHINES: THE CHALLENGE OF INCREASING DIAGNOSTIC PRECISION IN THE TURKISH HEALTH SYSTEM

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ☒ **Doktora** / PhD ☐

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☒
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years**. * ☐
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. * ☐

***** *Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

Yazarın imzası / Signature .... **Tarih** / Date ....

.