NOVA SCHOOL OF BUSINESS AND ECONOMICS
UNIVERSIDADE NOVA DE LISBOA

Dissertation, presented as part of the requirements for the degree of
Doctor of Philosophy in Economics

Essays on Health Economics

Joana Mª R. N. de P. Pestana

Student nº 32332

A dissertation carried out on the PhD in Economics,

under the supervision of Professor Pedro Pita Barros

October 25, 2023

Abstract

This thesis delves into a series of health economics essays that critically evaluate the broader health care system, particularly focusing on primary care, mental health and on providers behaviour. Chapter 1, investigates the variations among primary care physicians in Portugal, offering a thorough examination across multiple dimensions of care. Chapter 2 investigates the link between the supply of psychologists and the prescribing patterns of antidepressants in primary care. Chapter 3 delves into the key factors influencing Portuguese physicians' job preferences, highlighting the differences between physicians exclusively in the public sector and those concurrently working in both public and private sectors.

Keywords: primary health care, practice variation, physician prescription behaviour, preferences of physicians

To my parents

“On ne fait jamais attention à ce qui a été fait;

on ne voit que ce qui reste à faire.”

- Marie Skłodowska-Curie

Acknowledgements

This thesis represents the culmination of five years of research, and its successful completion owes much to the generous support and assistance of many people and institutions.

I am deeply grateful to the dedicated educators and mentors who have supported and shaped my educational and professional journey. Their insightful lessons not only inspired my lifelong commitment to learning but also drove me to surpass my own expectations. Special recognition is due to those within the European Training Network "Improving the Quality of Care in Europe" consortium, for without them, this PhD journey would have remained a distant dream.

I owe a debt of gratitude to my PhD supervisor, Professor Pedro Pita Barros, whose feedback on my work, patience and continuous support have been indispensable. His enduring curiosity, provocativeness, and commitment to research excellence, coupled with his generous spirit, provided the ideal environment for the successful completion of this work.

I must also express my gratitude to Professors Arthur Attema, Job van Exel, and Werner Brouwer for their support and feedback during my research period as a visiting scholar at Erasmus University Rotterdam.

Throughout my academic journey, I have had the privilege to collaborate with many colleagues, especially those met during seminars, conferences, and events. My appreciation and gratitude extend to my co-authors, fellow ETN Research Fellows, the Nova Health Economics and Management knowledge centre community, and the health economics team at the Nova National School of Public Health and Comprehensive Health Research Centre. Their support and encouragement in pursuing our shared research goals have been vital. Further gratitude goes to the members of the Portuguese Health Economics Association for their persistent cheering from the early stages of my PhD.

I am grateful to the financial support provided by the European Research Council under the European Union's Horizon 2020 research and innovation programme (Marie Sklodowska-Curie Grant Agreement No.721402), as well as the Fundação para a Ciência e a Tecnologia (PhD studentship: 2021.06183.BD) and the Research Office at Nova SBE for their contributions.

I would like to extend my gratitude to ACSS for granting access to the data that was instrumental for this research. A special thanks goes to Doctor Nuno Sousa for his assistance and guidance during my affiliation with the institution.

Lastly, but most importantly, my deepest gratitude goes to my friends and family. My parents, my sister, and extended family have been an endless source of support, affection, and strength throughout my life. They taught me the values of hard work, perseverance, and the importance of learning from both triumphs and setbacks. From zeniths to nadirs, I must acknowledge my partner for being *eine Quelle der Liebe und feste Unterstützung*.

This thesis stands as a testament to the collective support, encouragement, and contributions of these remarkable individuals and institutions.

Joana Pestana

Lisboa, October 2023

Contents

Introduction 1

1 Exploring inter-physician variation across multiple quality dimensions and the role of primary care organization 7

Abstract 7

1.1 Introduction 8

1.2 Related literature 10

1.2.1 Practice variation 10

1.2.2 The meaning of variation 12

1.2.3 Primary health care indicators 13

1.2.4 Hierarchical organization 14

1.3 Institutional Background 16

1.3.1 Primary care reform 17

1.3.2 Primary care governance 20

1.4 Methods 24

1.4.1 Data 24

1.4.2 Outcome variables 25

1.4.3 Explanatory variables 26

1.4.4 Statistical analysis 27

1.5 Empirical results 30

1.5.1 Descriptive statistics 30

1.5.2 Variation across practices and local groups 31

1.5.3 Comparison between types of practice 35

1.5.4 Comparison between indicators linked with incentives 38

1.5.5 Analysis of the determinants of achievement 39

1.6 Discussion 45

1.6.1 Summary 45

1.6.2 Limitations 49

1.6.3 Policy and Practice Implications 51

1.7 References 55

1.8 Appendices 62

1.8.1 Evolution of the number of patients enrolled in primary care in Portugal 62

1.8.2 Diagram of the Portuguese National Health Service 64
1.8.3 Diagram of the study population 66
1.8.4 Description of the indicators 67
1.8.5 Summary statistics from outcome variables 69
1.8.6 Summary statistics from outcome variables per type of indicator and practice 74
1.8.7 Summary statistics from explanatory variables 75
1.8.8 Results per indicator – Variance Partition Coefficient (VPC) at practice- and local- level (stacked) for each Indicator. 76
1.8.9 Results per type of practice – Variance Partition Coefficient (VPC) at practice- and local- level (stacked) for each Indicator, per type of practice. 77
2 Antidepressant therapy prescription, do psychologists help? - Evidence from Portugal .. 79
Abstract 79
2.1 Introduction 80
2.2 Setting 84
2.2.1 Depression and treatment approaches in Portugal 84
2.2.2 Portuguese primary care and mental health support 85
2.3 Methods 87
2.3.1 Data 87
2.3.2 Study Population 88
2.3.3 Variables 88
2.4 Empirical approach 90
2.4.1 Statistical model 92
2.4.2 Heterogeneity and robustness analyses 93
2.5 Results 95
2.5.1 Descriptive statistics 95
2.5.2 Within-between random effects model 99
2.5.3 Heterogeneity analyses 101
2.5.4 Robustness analysis 105
2.6 Discussion 106
2.7 Conclusion 111
2.8 References 112
2.9 Appendices 116
2.9.1 Psychology supply on diagnostic of depression and reverse causality analyses 116

2.9.2 Comparison of within-between random effects model (main) and model without GP utilization rate and patient list size....117
2.9.3 Comparison of within between random effects model with unbalanced (main) and balanced model....118
2.9.4 Diagram of the study population....119
2.9.5 Geographical distribution of the prevalence of depressive disorders diagnosis 120
2.9.6 Relationship between the incidence of depression in the GP patient list and number of psychologists/100,000 patients in the local group....121
3 Heterogeneity in physician's job preferences in a dual practice context - Evidence from a DCE....122
Abstract....122
3.1 Introduction....124
3.2 Setting....127
3.3 Methods....129
3.3.1 Choice experiment design....129
3.3.2 Choice sets....132
3.3.3 Pre-test and pilot study....132
3.3.4 Data....133
3.3.5 Analytical approach....133
3.4 Results....137
3.4.1 Are these preferences different whether professionals work in the public sector only? 142
3.5 Discussion....145
3.6 Conclusions....148
3.7 References....150
3.8 Appendices....155
3.8.1 DCE design....155
3.8.2 Latent Classes....159
3.8.3 Preference heterogeneity according to specialty....162
3.8.4 Semi-structured interviews....163
3.8.5 Pre-test Questions....165

List of Figures

Figure 1.1 Diagram of the Portuguese Primary health care organization and a short description of the influences at each level 21

Figure 1.2: Variance partition coefficient from GP indicators per dimension in 2018. 32

Figure 1.3: Variance partition coefficient from GP indicators per dimension and type of practice in 2018 36

Figure 1.4: Variance partition coefficient from GP indicators per dimension and type of indicator, incentivized (+) and non-incentivized (o) 39

Figure 1.5: Evolution of the number of patients enrolled in Primary care in Portugal: with GP, without GP and opting out of GP. 62

Figure 1.6: Diagram of the health system as of 2018 64

Figure 1.7: Diagram of the study population 66

Figure 1.8: Variance Partition coefficient for each indicator in 2018. 76

Figure 1.9: Variance Partition coefficient for each indicator in 2018 by type of practice 77

Figure 2.1. Geographical characterization of the GP antidepressant prescription share of patients with depression (top panel) and the average number of psychologists FTE in the local group per 100,000 enrolled patients (bottom panel) in 2015 (left) and 2018 (right). 98

Figure 2.2: Relationship between the share of depressed adults with antidepressants prescription and the average number of psychologists available at the local group level. 99

Figure 2.3 Heterogeneity in the association of psychological therapy supply and antidepressants prescription rate by GP and GP practice characteristics 104

Figure 2.4: Diagram of the study population 119

Figure 2.5: Geographical characterization of the prevalence of depressive disorders diagnosis (diagnosed according to P76 of ICPC-2) 120

Figure 2.6: Relationship between the incidence of depression in the GP patient list and number of psychologists/100,000 patients in the local group 121

List of Tables

Table 1.1 Multilevel mixed effects ordered logit regressions of the probability of physicians achieving the expected/acceptable intervals in all indicators grouped by dimension.....40

Table 1.2: Evolution of the number of patients enrolled in Primary care in Portugal and the percentage of patients in a GP list.....62

Table 1.3: List of indicators and corresponding target intervals in 2018. 67

Table 1.4: Descriptive statistics of the outcome variables in 2018.....69

Table 1.5 Summary statistics of achievement for incentivized and non-incentivized indicators per type of practice (PHCU, FHU-A, FHU-B) 74

Table 1.6: Descriptive statistics of the explanatory variables in 2018.....75

Table 2.1: Descriptive statistics of the study population. 97

Table 2.2: Within-between random effect models studying the association of psychotherapy supply and antidepressant prescription share.....100

Table 2.3: Heterogeneity in the association of psychological therapy supply and antidepressants prescription rate by average number of psychologists in the local group.....102

Table 2.4: Heterogeneity in the association of psychological therapy supply and antidepressants prescription rate by average antidepressant prescription rate of GPs. 105

Table 2.5: Psychology supply on diagnostic of depression and reverse causality analyses..116

Table 2.6: Comparison of within-between random effects model (main) and model without GP utilization rate and patient list size.....117

Table 2.7: Comparison of within between random effects model with unbalanced (main) and balanced model 118

Table 3.1: Attributes, levels, description, and expected sign of the regression parameters, E(sign).....130

Table 3.2. Characteristics of the physicians.....137

Table 3.3. Regression results from preference estimation using multinomial, mixed logit and generalized mixed logit models. 139

Table 3.4. Marginal willingness to accept compensation estimates. 141

Table 3.5: Results of the mixed logit model with single interactions presenting preference heterogeneity according to the physicians' contractual circumstances. 143

Table 3.6: Estimation of the latent class logit model – 3 classes.....159

Table 3.7: Characteristics of physicians, by latent classes 160

Table 3.8: Coefficients and significance levels for Mixed Logit Model (MXL1) specification, for all attributes interacted with specialities dummy variable (GP=1 for general practitioners, GP=0 for hospital specialists). 162

List of Acronyms and abbreviations

PHC	Primary Health Care
GP	General Practitioner
ACSS	*Administração Central do Sistema de Saúde, I.P*. Central Administration of the Health System
ARS	*Administração Regional de Saúde*. Regional Health Administration
ACES	*Agrupamento de Centros de Saúde*. Local groups of practices
LHU	Local Health Units. *Unidade Local de Saúde*
PHCU	Primary Health Care Unit. *Unidade de Cuidados de Saúde Personalizados*
FHU-A	Family Health Unit model A. *Unidade de Sáude Familiar tipo A*
FHU-B	Family health units model B. *Unidade de Sáude Familiar tipo B*
QI	Quality indicators
P4P	Pay for Performance
UK	United Kingdom

Introduction

Healthcare systems globally are currently under intensive scrutiny, with an emphasis on the quality, equity, and efficiency of the care they deliver. Primary Health Care (PHC), acknowledged as the cornerstone of numerous health systems, holds the promise of enhancing overall health outcomes, ensuring patient-centric care, and optimizing the utilization of resources (OECD, 2020). However, even within primary care, variations in the consistency and quality of care persist.

Chapter 1 begins my exploration, focusing on the inter-physician variations in primary care indicators across various quality dimensions. It delves into both supply-side and demand-side factors influencing these variations. An integral component of the supply-side equation is the presence and influence of other health professionals. Their integration into primary care settings can substantially shape the clinical decisions of General Practitioners (GPs), as highlighted in Chapter 2. This chapter investigates the relationship between the availability of psychologists in primary care and the antidepressant prescription behaviours of GPs. As global concerns mount over medication over-prescription and there's a push towards more holistic care, understanding the role of mental health professionals in the treatment process becomes pivotal. Nonetheless, the strength of a healthcare system is not solely anchored in its policies, structures, or interdisciplinary collaborations. At its core, a system's efficacy is linked with the adequacy and motivation of its professionals. Despite the leaps in technological advancements, healthcare's quality and accessibility fundamentally still pivot around the competencies and availabilities of its professionals. Chapter 3 shifts the lens towards the physicians. The importance of this topic has been highlighted in several studies and reports, such as the 2022 World Health Organization report, that stresses that many European countries are struggling the with difficulties to attract and retain health professionals. In Portugal, 2023 stands out as a

tumultuous year, marked by significant upheaval between physicians and the government. Numerous strikes, intense union negotiations, and consequent closures of services emphasized the profound tensions and highlighted the pressing challenges within the National Healthcare Service (NHS). Today's physicians, already contending with the aftershocks of the post-financial crisis budgetary constraints and the pressures of the pandemic, face an evolving landscape of challenges. These range from shifting population demographics and the technological resources at their disposal, to evolving health care market, where the private sector is competing for professionals and financing with the NHS, and a decreased purchasing power tied to their employment contracts. While increased remuneration remains a priority, this research exposes the non-monetary factors that play an equally, if not more, significant role in influencing job choices. These insights are valuable for policymakers and healthcare administrators. By acknowledging and acting upon these preferences, they can create environments conducive to both attracting and retaining medical talent.

These three chapters align with the broader framework of health economics research, emphasizing “supply-side economics” and, subsequently, focusing on “planning, budgeting, regulation, and monitoring mechanisms”, as detailed in the schematic by Culyer & Newhouse (2000), specifically boxes D and G. On the supply-side this research explores the behavioural dynamics and institutional responses of health care workers as they face different working environments, organization, and financing structures. This research recognizes that the healthcare system extends beyond prominent institutions like hospitals and primary care practices, encompassing professionals providing care to specific patient groups such as the elderly and those facing mental health problems. On the other hand, by evaluating primary care indicators, we contribute to the assessment of effective monitoring tools and the creation of more effective incentive structures. This analysis aids in understanding the relationship between organizational frameworks, and their influence on practitioner variations. Moreover,

examining monitoring mechanisms and physician preferences regarding job rewards paves the way for innovative approaches in system monitoring, organizational strategy, and compensation.

On a broader scale, Portugal, faces the challenges inherent to its healthcare system, and the *Zeitgeist* affecting healthcare systems globally. This thesis aims to contribute to comprehension of some of these challenges through three distinct yet interconnected studies that study providers behaviours and preferences in Portugal from various perspectives. A shift in the landscape of PHC has taken place in Portugal over recent decades. Initiated in 2006, the reform aimed to bolster the sector by improving accessibility, enhancing patient and professional satisfaction, and elevating the quality and continuity of care. These steps were taken with an overarching objective: to elevate the performance and efficiency of the National Health Service (NHS) and deliver more significant health benefits to the population. Central to these reforms was the evolution of traditional health centres into multidisciplinary teams, like Family Health Units (FHU). A strategic shift was made toward contracting with self-managed public units, operating in a network of units, integrating performance-related pay as an incentive for improved care outcomes. As of 2023, Portugal features 304 traditional health care units, in contrast to the 271 that adopted the team-based organization of the initial type FHU-A, and 348 units that evolved to adopted individual performance-related incentives of the type FHU-B, a testament to the progressive nature of the reforms.

Chapter 1 explores the landscape of primary care after the reform in Portugal, analysing the variation in physician achievement in 50 indicators spanning six dimensions of healthcare – quality, including quality, intermediate health outcomes, access, and efficiency measures. Leveraging nationwide administrative data encompassing a vast cohort of full-time GPs, we employ multilevel linear random intercept models to disentangle the sources of variation in

primary care quality across GPs, practices, and local groups of practices (ACES) in the year 2018. Our findings confirm a substantial degree of patient-adjusted variation in clinical practices, emphasizing the challenges of delivering uniform and consistent care. The highlights the distinct contributions of primary care practices and ACES in shaping these variations. While prior research often focused on variations within specific conditions or procedures, this study provides a holistic assessment of quality variations across diverse healthcare dimensions. The results highlight that the contribution of the practice where the GP provides care varies across dimensions, with access and preventive care indicators reflecting the practice's influence comparably to the individual GP, while efficiency dimensions exhibit a predominant role of the GP. Furthermore, the study examines variations between different types of practices, distinguishing between conventional salaried primary care units (PHCU) and team-based FHU. FHUs exhibit a greater degree of independence from ACES influence for health outcome and quality of process indicators, suggesting distinct strategies employed by FHUs in pursuit of these targets. Moreover, our research demonstrates that the variation in indicators linked with financial incentives are less influenced by the healthcare organization structure. This nuanced perspective on the effectiveness of incentive-based quality improvement strategies contributes to the ongoing debate on physician performance assessment, as well as the importance of considering organizational dynamics alongside individual physician characteristics in healthcare quality assessments.

Chapter 2 investigates the relationship between the number of psychologists working in Portuguese primary care and the prescription of antidepressants by GPs from 2015 to 2018. While the overall effect of increasing psychologist availability on GP prescription was inconclusive, the analysis revealed intriguing heterogeneity in the results. Specifically, the presence of one psychologist per 100,000 patients led to statistically significant reductions in antidepressant prescriptions among GPs in both the lowest and highest quartiles of psychologist

supply. Furthermore, this research sheds light on the interplay of factors influencing prescription behaviour. GPs working in local groups with a higher average number of psychologists per 100,000 patients had a lower share of antidepressant prescriptions. This finding suggests that the impact of psychologist availability on prescription behaviour depends not only on the supply of psychological services but also on physician and organizational characteristics within local healthcare settings. Despite these insights, the study acknowledges limitations in establishing causal relationships and highlights the need for sophisticated research designs in future investigations. Nevertheless, this study offers valuable insights into a field where evidence is sparse about GP treatment and prescription decisions in mental health when pharmacological and psychotherapy treatments are available.

Chapter 3 delves into the job preferences of physicians in Portugal, highlighting the significance of non-monetary attributes in shaping these preferences. Autonomy emerges as the top-valued trait, contrasting with less prioritized attributes like on-site clinical case discussions. While both public and dual private-public sector physicians displayed overlapping preferences, public sector physicians exhibited stronger preferences for frequent training and updated facilities. Interestingly, these preferences diverge from previous findings that primarily emphasized increased earnings, especially in developing countries. The research indicates that factors like autonomy, training frequency, and work-life balance play an important role in influencing physicians' choices in Portugal.

Diverse policy challenges can benefit from health economics insights, such as the decentralized commissioning, mental health policy, and attraction and retention of health professionals. I hope these essays in health economics can contribute to this policy objectives.

References:

OECD. (2020). *Realising the Potential of Primary Health Care* (OECD Health). Paris: OECD Publishing.

World Health Organization. (2022). *Health and care workforce in Europe: time to act*. WHO Regional Office for Europe

Chapter 1

1 Exploring inter-physician variation across multiple quality dimensions and the role of primary care organization [1]

Abstract

Medical practice variation arises from diverse sources, including organizational aspects of healthcare providers. This study explores inter-physician variations in multiple primary care indicators, aiming to assess the extent to which the institutions and hierarchical health care organization contribute to the variation. We further examine whether this contribution differ according to the type of primary care practices, and across distinct categories of indicators, and which factors are associated with physicians' achievement of indicators' targets. Using a comprehensive set of physician-level indicators, we estimated multilevel models to account for the clustering of physicians (n=4,502) within primary care practices (n=851), which, in turn, were nested within local groups of practices (n=55), adjusting for physician, patient, and contextual factors.

Our findings revealed significant variation in clinical practices among physicians across all indicators. Consistent with existing literature, a substantial portion of unexplained variation can be attributed to physician and patient-level factors. However, a non-negligible proportion of this variation, approximately 22%, can be ascribed to practice-level factors, with an additional 6% associated with local practice groups. For dimensions related to access and preventive care, the proportion of variation attributed to the organization is almost as large as that attributed to the physician. Conversely, for indicators concerning intermediate health

[1] This chapter is co-authored with Julio Souza, Andreia Pinto, and João Vasco Santos.

outcomes (e.g., blood pressure), maternal care, and prescription costs, the predominant source of variation lies at the physician level, suggesting a greater degree of physician discretion. Interestingly, the contribution of practice-related factors is less pronounced for indicators associated with financial incentives in child and maternal healthcare. These results highlight that differences in the organization of practices and its local groups may be driving inter-physician variation.

1.1 Introduction

Large practice variations exists at all levels of the healthcare system and are pervasive in all countries (Brownlee et al., 2017; Corallo et al., 2014). Understanding and quantifying practice variations can offer insights into the underuse or overuse of resources and services, impacting health care quality, health outcomes, inequalities and overall costs for both healthcare systems and societies (Lay-Yee, Scott, & Davis, 2013; Skinner, 2011).

In this study, we study the sources of practice variation in primary care to inform the right policy measures and policy targets. We posit that substantial practice variations persist across General Practitioners (GP) even after accounting for patient characteristics and contextual factors, and that those variations differ according to the dimension of the clinical practice being measured. Moreover, we hypothesize that structural elements of the organization of care and financing models contribute to the observed variations across different quality-of-care indicators.

Substantial practice variations across physicians, providers, and regions in various countries have been extensively documented (Busse et al., 2019; Folland & Stano, 1990). Within primary healthcare alone, marked variation has been identified in areas such as diagnosis (Nielen, Schellevis, & Verheij, 2009), consultation time (Irving et al., 2017), treatment (O'Connor et

al., 2008), prescription patterns (Davis & Gribben, 1995), referrals (O'Donnell, 2000), orders for laboratory tests, imaging and other resources utilization (Grytten & Sørensen, 2003), issuing sick leave certificates (Aakvik, Holmås, & Kamrul Islam, 2010), and follow-up requested (Lay-Yee et al., 2013), even after accounting for patient heterogeneity. These variations underscore a quality gap between existing practices and optimal healthcare services (Corallo et al., 2014). This quality gap may originate from different hierarchical levels within the organization of healthcare provision. However, few studies explicitly identify the amount of variation attributable to different hierarchical levels (Krein, Hofer, Kerr, & Hayward, 2002).

Despite differences in physician's beliefs about the efficacy and appropriateness of alternative therapies, commonly referred to as "practice style" (J. E. Wennberg, 1985), there is broad consensus that the organization of healthcare and practice regulations exert influence over practice style, thereby contributing to inter-physician care variations (J. D. De Jong, 2008; Westert & Groenewegen, 1999). This explanatory framework has often been described as the "supply hypothesis" (Davis, Gribben, Scott, & Lay-Yee, 2000).

Our contributions are threefold. First, we investigate inter-physician variations, adding to the limited pool of research that focuses on variations across practitioners (in contrast to geographical variations) (Sutherland & Levesque, 2020). Second, we examine multiple dimensions of quality of care, including the process of care, intermediate health outcomes, access, and efficiency. Lastly, we distinguish between incentivized and non-incentivized indicators, going beyond indicators solely considered in the commissioning process (Sutherland & Levesque, 2020).

1.2 Related literature

1.2.1 Practice variation

Practice variation has long been a central consideration in quality management, since the seminal work of Wennberg and colleagues on small-area variations (Mullan, 2004; J. Wennberg & Gittelsohn, 1973). Variations used to be attributed to physicians' beliefs and experience, especially when medical decisions were mired in uncertainty (J. E. Wennberg, Barnes, & Zubkoff, 1982). Subsequent research has challenged this notion, highlighting the significant influence of social context and organizational aspects on physician behaviour, even when medical uncertainty is low (Cutler, Skinner, Stern, & Wennberg, 2019; Grytten & Sørensen, 2003; Rodriguez, Scoggins, von Glahn, Zaslavsky, & Safran, 2009; Westert & Groenewegen, 1999). Moreover, researchers could not find strong empirical evidence that clinicians' personality traits and interpersonal behaviors had a significant impact on the quality of care (Boerebach et al., 2014).

Two systematic reviews examining medical practice variations have revealed substantial disparities in care provision across regions, hospitals, and physician practices, focusing in particular medical conditions and procedures (Corallo et al., 2014; Fung, Schmittdiel, Fireman, Meer, Thomas, Smider, Hsu, Selby, et al., 2010). Variation can be influenced by various factors at different levels: the organization of the health system (e.g., funding/payment mechanisms, incentives, guidelines), the regional authority (e.g., availability and accessibility of resources, management skills, information dissemination), the practice (e.g., size, organization and management, skill-mix, culture), the physician (e.g., experience, training, beliefs), and the patient (e.g., health literacy, comorbidities, socioeconomic status, preferences) (Fung, Schmittdiel, Fireman, Meer, Thomas, Smider, Hsu, Selby, et al., 2010). Additional conceptual models have expanded this list to encompass social values and regulatory bodies (Cook et al., 2018). Current evidence mostly focuses on inter-region or inter-provider comparisons, with

limited attention to inter-physician variations (Corallo et al., 2014; Fung, Schmittdiel, Fireman, Meer, Thomas, Smider, Hsu, Selby, et al., 2010). Most articles characterize variability as the proportion of total variation attributed to specific levels, and generally find modest proportions; less than 19% of the total variation attributed to physicians, leading researchers to question the justification for targeting quality improvement interventions at physicians (Fung, Schmittdiel, Fireman, Meer, Thomas, Smider, Hsu, & Selby, 2010).

Studies in primary health care (PHC) have highlighted significant variations across different settings. In Israel, inter-physician variation was three times greater than practice-based cvariation (Shashar, Ellen, Codish, Davidson, & Novack, 2021). In northern California, practices and physicians contributed ≤4% to the total variation in almost all quality indicators (except for patient-reported care experience scores, with 8.6%), though for all indicator greater proportion of variance was found for physicians than for facilities (Selby et al., 2010). A study on Veterans Affairs medical centres found that variations between practices and patients were more meaningful than variations between physicians in diabetes care (Krein et al., 2002). In Australia, analyses of PHC performance indicators following the implementation of regional coordination initiatives found that local practice clusters (known as Divisions of General Practice) accounted for 19% to 64% of adjusted total variation (Scott & Coote, 2007).

As in other countries, the study of clinical variations in Portugal has encompassed mostly hospital services (Parchman, 1995), where significant geographic variations were observed in the utilization of different health procedures (OECD, 2014). According to a review (Corallo et al., 2014), only one paper studied inequality in the geographic distribution of physicians in Portugal (I. Correia & Veiga, 2010). In sum, to date, very little attention has been paid to understanding the role of the ACES implemented in Portugal in 2005 on the quality of care and how much of the variation in GP indicator achievement they can explain. Addressing quality variation is particularly pertinent in Portugal, where, in comparison to similar settings such as

the UK and Spain, the continuity and coordination of primary care services have been rated as moderate rather than strong (D. Kringos et al., 2013). While various studies in Portugal have described considerable variations between healthcare providers or geographical areas (I. Correia & Veiga, 2010; da Luz Pereira et al., 2021; Monteiro, 2020), none have investigated intra-practice (i.e., inter-physician) variations.

1.2.2 The meaning of variation

Benchmarking healthcare providers or geographical areas has been a way to reduce variations, envisioning quality of care improvements and cost containment. Comparisons have been enabled by routine use of quality indicators (i.e., administrative data) (Ramalho et al., 2019), often to support the implementation of pay-for-performance programs and public reporting requirements (Emanuel et al., 2016; Rosenthal & Frank, 2006; Werner & Asch, 2005). These programs use monetary and reputational incentives to motivate physicians and practices to achieve specific healthcare quality goals.

Variation is a complex phenomenon that can reflect both appropriate and inappropriate care, depending on the clinical context and evidence base. There is some controversy about the extent to which variation across providers is clinically meaningful (Fung, Schmittdiel, Fireman, Meer, Thomas, Smider, Hsu, Selby, et al., 2010) or unwarranted (Harrison et al., 2019; Mercuri & Gafni, 2011; Selby et al., 2010; Sutherland & Levesque, 2020), and the literature contains few frameworks to determine inappropriate variation (Mercuri & Gafni, 2011). Variation can be benign when it aligns with evidence and patient needs. With personalized care increasing, clinical practice is expected to vary to reflect patients' informed choices. Therefore, variations influenced by factors that are extrinsic to the patient — those that occur across physicians and providers — are potentially more relevant. Identifying the sources of variation can inform the

implementation of interventions to improve quality and reduce inadequate care (Selby et al., 2010), which is what is addressed in this paper.

1.2.3 Primary health care indicators

This analysis focus on primary health care indicators, which serve as robust measures to assess the quality of medical care across various dimensions. These well-defined and measurable indicators are instrumental in evaluating the quality of clinical practice (Lawrence & Olesen, 1997).

In understanding primary care, we adopt the classic Donabedian framework, which categorizes it into three complex levels: structure, process, and outcome, each comprising multiple dimensions (Donabedian, 1988). Given our focus on understanding the organizational factors contributing to practice variation, we use structure indicators primarily as explanatory variables rather than standalone output dimensions, a methodology commonly employed in similar analyses (Ramalho et al., 2019). The core dimensions characterizing primary care, as outlined by Kringos et al. (2010), encompass four elements within primary care processes: access, continuity, coordination, and comprehensiveness of care; and three dimensions evaluating the outcomes of a primary care: quality of care, efficiency of care, and equity in access to health. Remarkably, some authors also include specific clinical measures as part of the outcome dimensions, referred to as intermediate health outcomes, which are influenced by preventive services but are not health outcomes in and of themselves (e.g., measures of HbA1c, blood pressure, cholesterol or physical activity) (Scott et al., 2011).

In our study, we delve into the dimension of access to primary care services, utilizing indicators that assess the accommodation of accessibility, such as appointment systems. Additionally, we analyse the quality of primary care, reflecting the extent to which health services meet patient

needs and adhere to standards of care. Our investigation covers the management of chronic diseases, maternal and child health care, health promotion, preventive care, and specific intermediate health outcomes. Lastly, our analysis extends to the efficiency of the primary care system, with a specific focus on achieving a balance between resource allocation and desired outcomes. This includes an examination of workforce efficiency through metrics like prescription frequency and treatment expenditure.

1.2.4 Hierarchical organization

Since the 1978 Alma-Ata Declaration, the fundamental concept of primary care and its pivotal role in ensuring universal access to healthcare services, has been acknowledged. However, the implementation and delivery of primary care can differ across healthcare systems and regions. In many countries, particularly those with a National Health Service model, GPs stand at the forefront of this crucial healthcare sector, serving as gatekeepers who guide the utilization of elective care (Westert, de Jong, & da Silva, 2015). Furthermore, GPs are increasingly working within group practices alongside other healthcare professionals, which has been shown to influence the way physicians practice medicine. Evidence from several countries suggests that this co-location of physicians within team-based practices can lead to improvements in clinical processes, innovation, and quality assurance, although contrasting evidence exists regarding patient satisfaction (Bonciani, Schäfer, Barsanti, Heinemann, & Groenewegen, 2018; Damiani et al., 2013).

Moreover, primary care practices are frequently organized into local networks or groups, with the overarching goal of enhancing health services planning, facilitating shared services, supporting commissioning, resource allocation, and promoting the diffusion of best practices (J. Smith & Goodwin, 2017). Consequently, it is reasonable to assume that observed practice

variations, whether systematic or random, may emerge at various levels within the healthcare system, including regions, local groups of practices, individual practices, and among practitioners themselves (Westert & Groenewegen, 1999). In recent health system reforms, local primary care networks have been implemented as a hub fostering teamwork and collaboration among primary health care providers and stakeholders in defined geographic areas (Saltman, Bankauskaite, & Vrangbaek, 2006). Notable examples include the local primary care centre groups in Portugal (*Agrupamento de Centros de Saúde*- ACES) (Leone, Dussault, & Lapão, 2014), Primary Care Trusts in England (Campbell, 2002), the Divisions of General Practice in Australia (Scott & Coote, 2007), or the Primary Health Organizations in New Zealand (Hefford, Crampton, & Foley, 2005). Despite the popularity of these organizational changes, their precise impact on reducing variation in healthcare delivery remains uncertain (Scott & Coote, 2009). A systematic review suggested that cohesive and collaborative local health professional networks can improve primary care quality by providing a platform for collaboration and learning, facilitating the sharing of best practices and resources among providers (Cunningham et al., 2012).

By promoting a culture of change management, these networks help standardize care processes, encourage the use of care pathways and clinical guidelines, and promote evidence-based approaches to care, thereby reducing variation in medical service delivery (C. Pearce, Shearer, Gardner, Kelly, & Xu, 2012). Furthermore, they can offer training and support to providers, aiding in the adoption of new techniques and technologies that improve quality and reduce variation. Previous studies in the UK have demonstrated that Managed Clinical Networks for diabetes led to a significant shift in the care of type 2 diabetes, resulting in improved clinical indicators (Greene et al., 2009). Evidence from primary care networks in Germany, which served as platforms for professional peer exchanges, also highlighted their positive impact on reducing variation in antibiotic prescribing (Poss-Doering et al., 2020). The impacts of

implementing such organizational changes depend on the design and implementation of the networks (Phillips et al., 2010), as well as the local context (J. D. de Jong, Groenewegen, & Westert, 2003). Additionally, local networks can improve efficiency by reducing duplicative services, and facilitating care coordination with other local services, such as nutrition or mental health (Sirimsi et al., 2022).

1.3 Institutional Background

The National Health Service (NHS) of Portugal is primarily funded by taxation and follows a universal health coverage model. A visualization of the Portuguese health system can be found in Figure 1.6 in the appendix. Most primary healthcare falls under the NHS, with a limited number of private providers mainly in urban areas (Barros, 2017). Given this, our analysis focuses on the public PHC providers.

In the Portuguese NHS, GP practices have been established since the early 1970s (Branco & Ramos, 2001). Portugal was among the pioneering European countries to integrate primary health care by developing an extensive health center network (Dionne Sofia Kringos, Boerma, Hutchinson, & Saltman, 2015). Presently, Portugal's primary care utilizes a list-based system where inhabitants register with a practice in their residence area. GP act as gatekeepers to specialist care and maintain fixed patient lists, allowing patients to select a family doctor if a slot is available (Simões et al., 2017). Typically, a GP's list encompasses 1500 to 1900 patients, providing services ranging from general adult care to preventive measures. However, there's a noticeable GP shortage in certain regions, leaving not all Portuguese residents with an assigned family doctor (Barros, Machado, & Simões, 2011). From 2016 to 2022, the number of patients enrolled in primary care has steadily increased, reflecting the growing demand for healthcare services. There was a simultaneous rise in the number of patients without a GP during this

period. In 2016, 767,147 patients were without a GP, comprising 7.66% of the total enrolled patients. By December 2022, this figure had increased to 1,494,392 patients, 14.14% of the total (see Appendix Table 1.2). It's crucial to mention that regional disparities exist in primary care access. According to data from the EU-SILC survey in 2019, individuals in the lowest income quintile experienced unmet medical needs due to cost, distance, or waiting time at a rate of 3.5%, whereas individuals in the highest income quintile reported a significantly lower rate of only 0.2% (OECD/EU, 2020). Despite minimal fees for PHC visits[2], several exemptions exist for vulnerable groups.

1.3.1 Primary care reform

In 2005, Portugal initiated a comprehensive healthcare reform to restructure Primary Health Care (PHC). Backed by an advanced information system (OECD, 2015), the reform aimed to strengthen primary care, enhance access, and improve care quality, efficiency, and overall satisfaction. This comprehensive reform introduced several new elements, including the introduction of Family Health Units (FHUs) and reorganization of Primary Health Care Units (PHCUs), and the establishment of local groups of primary care practices *(Agrupamento de Centros de Saúde* - ACES) (Barros & Simões, 2007).

FHUs introduced a novel approach with physician-driven, team-based units. These units comprised multidisciplinary teams and operated with functional and technical autonomy. These multidisciplinary teams typically consist of approximately three to eight General Practitioners (GPs), a corresponding number of nurses, and a variable complement of administrative professionals. They are designed to serve a population ranging from four to fourteen thousand individuals (Ministério da Saúde, 2013). Their funding is partially based on

[2] The GP visits co-payments have been extinct since 2020.

capitation and includes some incentives schemes based on Pay-for-Performance (P4P) schemes (Simões et al., 2017). Healthcare professionals working in PHC are almost exclusively salaried government employees recruited through public tender (Ibid.). The reform saw the emergence of three distinct practice types, with the same mission but distinct organizational features and incentives schemes. (1) the existing GP practices, were designated as Personalized Health Care Units (PHCU) (*Unidades de Cuidados de Saúde Personalizados*), (2) FHU type A (*Unidades de Saúde Familiar*/USF-A) and (3) FHU type B (*Unidades de Saúde Familiar*/USF-B) Financial incentives differ between FHU-A and FHU-B, with the former allocated to the entire team and the latter based on individual P4P incentives, adhering to stringent performance targets (Biscaia & Heleno, 2017; Pisco, 2011). Practices under FHU type A can apply to switch to FHU-B, provided they consistently demonstrate good performance. It is important to clarify that PHC units offer the same fundamental services, ensuring a consistent basis for comparison.

Another key aspect of reform was the established local groups of primary care practices *(Agrupamento de Centros de Saúde* - ACES), hereafter referred to as "ACES". ACES are publicly funded organizations that aggregate several primary care practices, assuming responsibility for delivering primary care within specific geographic areas. The creation of the ACES[3] streamlined resource allocation, and population oversight. Initially there were 74 which were later aggregated to 55 distributed nationwide (Biscaia & Heleno, 2017). They inherently support the decentralized practices by centralizing resources for organizational management, harmonization, the establishing technical-scientific leadership and governance and as channels for community and citizen representation and participation. The administrative independence allows for tailored solutions based on available resources and the specific needs of each community, employing suitable organizational structures and management tools (Dionne S

[3] The ACES were established under Decree-law 28/2008, dated 28th February. [Accessed from https://diariodarepublica.pt/dr/detalhe/decreto-lei/28-2008-247675 in July 2023].

Kringos, Boerma, Hutchinson, Saltman, & Organization, 2015). The geographic demarcation of ACES considers factors such as the number of residents (ranging from 50,000 to 200,000), population structure, aging index, and proximity to the reference hospital (Barros et al., 2011).

In addition to primary care practices, each ACES group governs several units[4]. Each ACES includes a unit of shared resources (*Unidade de recursos assistenciais partilhados* – URAP), comprising professionals like nutritionists, dentists, psychologists, social workers, among others; a management support unit (*Unidade de apoio à gestão* – UAG) providing administrative and general support to the executive director of ACES, the clinical board, and functional units; a Public Health Unit (*Unidade de saúde pública* – USP), responsible for local health profiles, planning, and the management of communicable diseases, as well as the planning and dissemination of public health initiatives, epidemiological surveillance, and the management of population-wide programs in prevention, health promotion, and protection. Furthermore, ACES establish connections with the long-term care network through the Community Care Unit (*Unidade de cuidados na comunidade* - UCC) offering care to groups with special needs and community interventions, focusing on providing health care and support at home and in the community. UCC plays a crucial role in health education, integration into family support networks, and the implementation of mobile intervention units. The UCC team consists of various professionals, including nurses, social workers, doctors, psychologists, nutritionists, physiotherapists, and speech therapists.

Furthermore, eight ACES fall under Local Health Units (LHU). LHUs were introduced in 1999 and further expanded in 2007 by the Ministry of Health to foster vertical integration between

[4] Details of the primary care units can be found on the Central Administration of the Health System website. [Accessed from https://www.acss.min-saude.pt/category/healthcare/primary/?lang=en in July 2023].

PHC and hospital-based care[5]. However, full coordination between care levels has not been achieved successfully (ERS, 2015; Nunes & Ferreira, 2019). Currently, LHUs cater to a limited portion of the mainland population, accountable for delivering care to 12% (Brito Fernandes et al., 2019). Comparative analyses suggest marginal differences between ACES and LHU-led units concerning care access, quality, and efficiency (ERS, 2015). Nevertheless, some challenges persist in LHUs, such as increased avoidable hospitalizations and prolonged inpatient stays (S. Lopes et al., 2017).

1.3.2 Primary care governance

Public primary healthcare services in Portugal, while more centralized compared to countries like Spain and Italy (Costa-Font & Pons-Novell, 2007) operate within a structured hierarchy. National health objectives are set by central administrative bodies, including the Ministry of Health. These objectives are translated into action by the five regional health authorities (RHAs), which coordinate with ACES and LHUs to realize these aims locally. Direct care to patients is facilitated by two types of healthcare units: the PHCUs and FHUs, as shown in Figure 1.1.

[5] Starting in January 2024, the Executive Directorate of the NHS plans to establish an additional 31 LHUs, encompassing all hospitals nationwide and catering to nearly the entire population. Notably, the financing mechanism for this new batch of LHUs will diverge from existing LHU. At the time this article was authored, further details were not available. Source: https://www.publico.pt/2023/08/31/sociedade/noticia/sao-unidades-locais-saude-vai-mudar-2024-2061740.

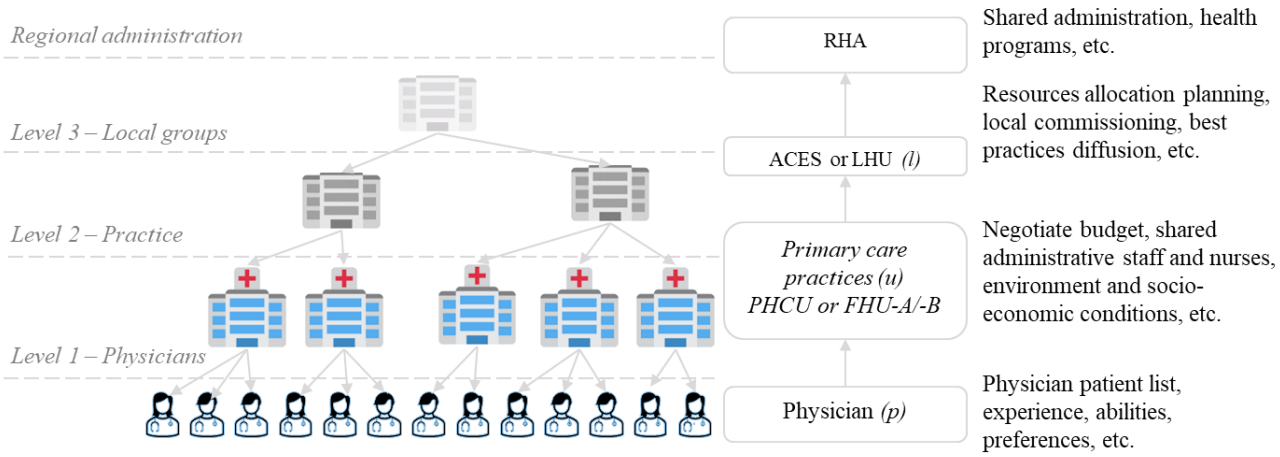


Figure 1.1 Diagram of the Portuguese Primary health care organization and a short description of the influences at each level

Streamlining contractual procedures, the Central Administration publishes yearly the "Terms of Reference for the Contracting of Healthcare in the NHS"[6]. This comprehensive document encapsulates guidelines for both primary care and hospital provider. The contracting process for primary care has two phases. First, an internal commissioning between the ACES and each of its constituent GP practices, resulting in a commitment letter (*Carta de Compromisso*) that includes the plan for the application of the institutional incentive and an action plan. This is followed by an external contract between the regional authorities and the ACES, leading to a program contract (*Contrato-programa*) that details the obligations, counter-obligations, monitoring rules, financial resources underpinning the contract, and methods for tracking and evaluation. The Terms of Reference document from ACSS lays down the guidelines delineated for the indicators that support internal and external contracting, ensuring uniformity in agreements, implementation timelines, and monitoring mechanisms.

[6] Refer to the 2018 document "Termos de Referência para contratualização de cuidados de saúde no SNS" published by the Central Administration of the Health System (ACSS), Available at: https://www.acss.min-saude.pt/wp-content/uploads/2017/11/Termos-Referencia-Contratualizacao-SNS_2018.pdf (Accessed July 2023).

The guidelines for incentive contractualization were based on criteria defined in Order no. 301/2008, dated 18th April[7]. Incentive allocation depended on evaluations from 15 potential indicators out of 49, with 13 being standard across all FHUs and 2 locally defined. These indicators delve into diverse sectors: 4 linked to accessibility, 8 to healthcare delivery efficiency, and 2 to economic viability. They capture aspects such as frequency of home medical visits, proportion of hypertensive patients tracked in recent months, and the average user's medication costs. The internal contract specified the indicator target value for each practice, based on historical data and local goals. The scoring system for each indicator, spanning between 0 and 2, reflected the level of target achievement. Incentive calculations relied on a cumulative score, and the calculation process was automated to alleviate any additional strain on practices (Miguel, 2010).

The incentives scheme is based upon a variety of indicators, mostly related to the care and management of child and maternal health, cancer screening, vaccination, diabetes, and hypertension (Perelman, Lourenço, Russo, & Miraldo, 2017). For an FHU-A to qualify for the team incentive, it must meet specific criteria in areas such as accessibility and healthcare performance. This incentive can be allocated to costs related to information dissemination, conference participation, research facilitation, or enhancing working conditions. The monetary value of the incentive is contingent upon the age-adjusted population size of the FHU-A, fluctuating between €9600 and €15200. In the case of FHU-B, the payment of individual financial incentives (P4P) hinges on the accomplishment of set goals. The maximum annual incentive can peak at €3600 for nurses and €1150 for administrative personnel. Seventeen

[7] Refer to Order no. 301/2008, dated 18th April (Regulates the criteria and conditions for granting institutional and financial incentives to family health units and their professionals, based on improvements in productivity, efficiency, effectiveness, and quality of care provided). Available at: https://diariodarepublica.pt/dr/detalhe/portaria/301-2008-249870 (Accessed July 2023).

performance metrics, covering domains from women's health and prenatal care to early childcare and chronic patient management, dictate these financial rewards.

In 2017, a significant change in the commissioning strategy was performed, transitioning from a static set of indicators to a flexible, multidimensional matrix, allowing individualized negotiations between each GP practice and ACES of the set of indicators, instead of negotiating the indicators targets (Pestana, Dimitrovová, Charondière, Broeiro, & Cardoso, 2019). Under this system, practices can contract varying indicators, providing a diagnostic overview and impacting overall performance (Monteiro, 2020). This approach also expanded the temporal scope of these indicators from annual to triennial contracts. Moreover, the benchmarks for each indicator are now standardized nationwide, facilitating our ability to consistently compare indicators across all GPs in this study.

The institutional developments in Portugal's healthcare landscape offer valuable lessons for health policies globally. At the core, is the distinctive configuration of the Portuguese NHS, which evidences a robust orientation towards PHC, with its structure anchored around GPs. Portugal's strong focus on PHC has led to it having the highest ratios of GPs to other medical professionals among OECD countries (OECD, 2021). Yet, an intriguing paradox emerges from the data. While Portugal boasts a top-tier structural ranking in PHC, it grapples with challenges in the continuity and coordination of primary care, placing it lower among 31 European counterparts (D. Kringos et al., 2013). Highlighting the necessity for enhanced coordination across different care levels, the WHO underscores the imperative of emphasizing the local-level decision-making to adeptly address unique challenges (WHO & World Health Organization, 2018). The evident contrast, especially concerning the dynamics of structure, process, outcomes, and coordination, motivates an in-depth exploration of the contribution of these structure to GP care variation in Portugal.

1.4 Methods

1.4.1 Data

The study was carried out using data for all GPs in mainland Portugal. Three data sources were used. First, we extracted data from primary care indicators database (*Bilhete de Identidade dos Cuidados de Saúde Primários*- BI-CSP) for 2018, that includes clinical indicators about the quality of service provided of PHC in mainland Portugal. It also includes information on the provider characteristics (e.g., type of practice), context indicators (e.g., location, population demographics), and practice human resources (SNS, 2019). Second, the human resources register RHV (*Recursos Humanos e Vencimentos*), which contains information on the primary care workforce, from the Central Administration of the Health System (*Administração Central do Sistema de Saúde; ACSS*), including anonymized information that characterize practitioners (age, gender, patient list characteristics). Third, municipality-level characteristics were obtained from the Portuguese national statistics bureau (*Instituto Nacional de Estatística - INE*) and Eurostat [8]. The first and second sources were linked at the GP level, while the third was linked at the GP practice level, based on the practice location.

Our population of interest consists of the universe of GPs in the Portuguese NHS (n=4,502 in 2018). The study sample includes only full-time GPs throughout the entire year on a given practice. We have excluded practices with only one GP and practices that underwent software migrations during the study period, due to potential disturbances in data collection (exclusions described in Appendix Figure 1.7).

[8] Unemployment data available at https://www.pordata.pt/db/municipios/ambiente+de+consulta/tabela/5833014 [Accessed on 12 Dec 2020]. The classification of the level of urbanization available at https://ec.europa.eu/eurostat/web/degree-of-urbanisation/background [Accessed on 12 Dec 2020]. The distance in kilometres between the practice and the ACES was calculated by the authors using the practice location and the headquarters of the ACES using QGIS software.

1.4.2 Outcome variables

The primary outcome variables in this study are the primary care indicators described previously. These encompass various dimensions including access, quality of care, intermediate outcomes, and efficiency, and are related to multiple clinical areas such as adult health, child and youth health, maternal health, hypertension, diabetes, mental health, cancer screening, and respiratory diseases. These indicators were originally defined by the Central Administration of the Health System (ACSS).[9] From the 143 indicators monitored by the ACSS in 2018, we selected 50 indicators for analysis. We excluded redundancies (e.g., for indicators reporting the same process over different time horizons, the shortest horizon was chosen), indicators monitoring technical aspects not linked with quality or performance (e.g., proportion of consultations performed in the interval 8-11 hours), those that reflected a portion of another indicator (e.g., those aggregating several chronic conditions, single condition indicators were selected), and those with high proportion of missing values (>5%).

Dependent variables are assessed through two lenses: by evaluating variations in the indicators' absolute values and by comparing them to established benchmarks to analyse the factors influencing physicians' relative performance. This dual approach allows us to study both the absolute fluctuations and to delve deeper into the determinants of meaningful variations in each GP's relative performance vis-à-vis the care targets. For gauging relative performance, each indicator is measured against a nationwide benchmark, which is determined by a technical commission[10] and published annually (ACSS, 2017). The benchmark consists of an expected

[9] Details regarding the definitions of the indicators have been published elsewhere (Ministério da Saúde, 2017). The classification in the original language was translated as process indicators (*Processo*), intermediate outcomes (*Resultados em saúde*), efficiency (*Eficiência*) and access *(Acesso*).

[10] The technical commission is composed of trade unions, professional orders, the regional administration, the ACSS, the association of FHU and the National Coordination for NHS Reform in Primary Health Care, as

and an acceptable interval. To calculate the achievement for each indicator, the following scoring system is applied: physicians receive two points if their indicator value falls within the expected interval, one point if they reach the acceptable interval, and zero points otherwise.

For instance, consider a physician who achieved a 55% rate of adult patients with arterial hypertension, with their last recorded blood pressure below 150/90 mmHg. In this case, the physician's achievement would be awarded one point, as their performance falls within the acceptable interval (defined as between 50% and 100% in 2018) but not within the expected interval (which ranged from 67% to 100%). A comprehensive list of these indicators, along with their corresponding target intervals, is provided in Table 1.3 in the appendix. Additionally, descriptive statistics for each indicator can be found in Table 1.4 in the same appendix.

1.4.3 Explanatory variables

The physician-level covariates consisted of GP age and gender, number of patients in the GP list (to account for workload), patient mix (percentage of female, children, and elderly), the proportion of co-payment-exempt and foreign patients (as socioeconomic proxies). Practice-related factors included workforce size (GPs, nurses, administrative staff), nurse-to-physician ratio, proportion of elderly patients (>65), distance to the head office of the ACES (in km), municipality unemployment rate, practice location (city, town or suburb, rural), and type of practice (FHU-A, FHU-B, PHCUs). At the local group (ACES) level, we considered the proportion of FHU-A and FHU-B relative to PHCUs, GPs and nurse numbers within the ACES that are not working in a practice, ratio of unassigned patients (i.e., patients without a denominated family doctor), and a dummy identifying if the ACES is part of a vertically

published by Serviços Partilhados do Ministério de Saúde at Matriz Indicadores Cuidados de Saúde Primários. Available online: https://bicsp.min-saude.pt/pt/investigacao/Paginas/Matrizindicadorescsp_publico.aspx [Accessed on March 2023]

integrated care units (S. Lopes et al., 2017). Description and statistics of the explanatory variables are available at the Table 1.6 of the appendix.

1.4.4 Statistical analysis

The main objective of the analysis was to quantify to which extent the unexplained variation in multiple physician's indicators (dependent variables) was attributed to systematic variation associated with practice and local group, or due to random chance variation. Furthermore, we investigated the associations between organizational factors and achievement of quality standards, while accounting for contextual factors, such as patient and physician characteristics, to remove any effect of different population structures across different regions.

It is expected that two physicians working in the same practice manage patients somewhat similarly and might follow similar lines of decision, contrarily to two physicians working in different practices. The same within-group homogeneity can be expected among physicians in the same ACES. To explore variation in quality adjusting for this nested structure of the data, we employ multilevel models. Multilevel models are commonly employed to explain physician-level heterogeneity when data are hierarchically structured (Duncan, Jones, & Moon, 1998; Gelman & Hill, 2006; Raman & Hedeker, 2005; Rice & Jones, 1997). GPs (n=5,619) were our observation units, and GP practice (n=901), and ACES (n=55) form the macro units, at level 2 and level 3, respectively. The models include random-intercepts to capture the inter-practice and inter-ACES variability while also controlling for physician-level factors as fixed.

First, we estimate three-level random-intercept models for each indicator separately (individual analysis) to study the proportion of variation that can be attributable to each level (Hox, Moerbeek, Moerbeek, & Schoot, 2018). We estimated cut points and variance components using a multi- level/nested linear model as described in Eq. 1:

$$y_{pul} = \alpha + X_p\beta + X_{pu}\gamma + X_{pul}\delta + v_l + u_{ul} + \varepsilon_{pul} \quad (1)$$

The outcome y_{pul} denotes the absolute achievement in each indicator of physician p, working in practice u, that belongs to the local group (ACES) l. This is explained by a constant term (α) and three vectors of explanatory variables (X_p, X_{pu}, X_{pul}). These vectors include the physician-level characteristics (X_p) (including age, gender, patient list size, demographics, and medical conditions), practice-level confounders that are common to all physicians in the practice (X_{pu}) and ACES-level characteristics (X_{pul}). The random term consists of three error terms, one at each level: the random intercept u_{ul} at the practice level (level 2), v_l at the ACES level (level 3), and ε_{pul} which denotes the individual physician error term. The individual error term was considered to be randomly distributed, and the random effects v_l and u_{ul} to follow a normal distribution, in all models (Goldstein, 2011).

In a three-step approach, we checked whether the results were robust to the inclusion of each group of covariates. Likelihood-ratio tests comparing the model with three-level and two-level ordinary linear regression were used and were highly significant for our data, suggesting the importance of using higher-order levels (three-level model results are presented).

For each indicator, we calculated the physician ($\sigma^2_{\varepsilon_{pul}}$), practice ($\sigma^2_{\mu_{ul}}$), and ACES ($\sigma^2_{v_l}$) variance terms, as well as the Variance Partition Coefficient (VPC), to assess the between-group variation. In hierarchical models, this is analogous to the intra-class coefficient in two-level models (Goldstein, Browne, & Rasbash, 2002). The ACES-level VPC_l and the practice-level VPC_u are calculated as the ratio of the ACES or practice variance to the total variance:

$$VPC_l = \sigma^2_{v_l} \times (\sigma^2_{\varepsilon_{pul}} + \sigma^2_{\mu_{ul}} + \sigma^2_{v_l})^{-1}, \text{ where } 0 < VPC_l < 1 \quad (2)$$

$$VPC_u = \sigma^2_{\mu_{ul}} \times (\sigma^2_{\varepsilon_{pul}} + \sigma^2_{\mu_{ul}} + \sigma^2_{v_l})^{-1}, \text{ where } 0 < VPC_u < 1 \quad (3)$$

As an example, a VPC of 10% for practice indicates that 10% of the total variation in the indicator can be explained by differences between practices, after accounting for variation at the ACES level.

The second set of analyses was performed to estimate the determinants of GP achievement of the indicators (pooled analysis). Converting all indicators to the same achievement scale allows us to study the common factors associated with high achievement. We estimated multilevel mixed effects ordered logit regressions of the probability of physicians achieving the expected/acceptable intervals in all indicators grouped by dimension (access, efficiency, intermediate health outcomes, process) to estimate marginal effects of the explanatory variables associated with physician's achievement of high-performance scores. Formally, the model in terms of a latent linear response can be written as follows:

$$y_{ipul} = X_{pul}\beta_1 + \gamma_i + v_l + u_{ul} + \tau_{pul} + \varepsilon_{ipul} \quad (4)$$

The observed ordinal outcome variable y_{ipul} is the achievement of the target associated with indicator i ($i = 1, \ldots, m$) [2 if expectable interval, 1 if acceptable interval, 0 otherwise]. In this case, physicians and indicators are not considered nested rather they are crossed, and therefore we included an indicator-specific term γ_i in the model that takes on the same value for all physicians. Each indicator can have a systematic effect on all physicians, due to the area or process it refers to, or the way it is measured. The error term ε_{ipul} is distributed as logistic with mean 0 and variance $\pi^2/3$ and is independent of u_{ul}. Analyses were performed using STATA 17 software (StataCorp., 2021).

1.5 Empirical results

1.5.1 Descriptive statistics

A total of 4,502 full-time GPs were included in this analysis. Appendix Table 1.6 characterizes physicians and practices. The physicians were on average 53 years old (±12); 63.5% were female, and the average list was 1,692 (± 246) patients. Regarding co-morbidities in the patient list, nearly one in four patients (23%) had arterial hypertension, and one in nine patients (11%) had depressive disorders; the prevalence of asthma, diabetes mellitus, malignant neoplasm, and obstructive chronic pulmonary disease was smaller (3%, 8%, 4%, and 1%, respectively). Most physicians were working in PHCU (41.7%) or FHU-B (30.2%), located in cities (39.6%) or Town or suburb (36.4%). The practices' location was on average 11km (±12) away from the ACES headquarters, and the local unemployment rate was 5.45 on average. Approximately one in four patients in the average practice was above 65 years, and the proportion of women was slightly larger than that of men (52.2%). There were approximately 1.15 nurses per physician, and 23 enrolled patients per professional (including nurses, physicians, and administrative staff) in the practice. In the ACES, on average, 6% of the patients had no family doctor assigned. Shared professionals within the ACES were on average 49 nurses and 12 physicians. Almost one-fifth of the ACES (19%) belong to vertically integrated care units. ACES aggregate different practice types: ranging between no FHU, to 67% FHU-A or 71% FHU-B at maximum.

The achievement of the indicators differed considerably from one GP to another. The descriptive statistics of the 50 indicators can be found in supplementary Table A3.2. In summary, the mean and standard deviation values of these indicators indicate some variation in performance across physicians. For access indicators, the rate of medical consultations by the respective family doctor shows a mean of 85.9% and a standard deviation (SD) of 10.9, reflecting variations in the accessibility of primary care services. Efficiency indicators, such as

the proportion of drugs prescribed that are generic drugs, have a mean of 62.7% and an SD of 6.6, suggesting relatively consistent performance. However, the cost of therapy for patients with Diabetes exhibits a substantial SD of 96.2, indicating significant variation in spending for this condition. Process indicators, like the proportion of pregnant women with timely surveillance consultations, have a mean of 90.7% and an SD of 11.6, while indicators related to child health surveillance present larger variation with SD values up to 27.6. Health outcomes indicators also display strong variation between GPs, with the indicator related to blood pressure control showing SDs of 22.6. Regarding the achievement relative to the benchmark, in 2018 most GPs (>50%) achieved the expected benchmarks in 20 out of 50 indicators, primarily in efficiency and health outcomes indicators. However, in 15 out of 50 indicators, a significant number of physicians fell short of the acceptable benchmark, something particularly noticeable in access indicators. Overall, these findings highlight the diverse quality of care provision across GPs in achievement in absolute terms, as reflected in their mean and SD values, but also in relative terms, reflected in the percentage of physicians that achieved the expected and acceptable intervals.

1.5.2 Variation across practices and local groups

Each indicator absolute achievement per GP was regressed on all the confounders at the physician-, practice-, and ACES-levels, and the variances were estimated. All estimated variances were statistically significant. The VPC ratios for each level were plotted in Fig.1., with the estimated proportion of variation attributed to the practice-level (y-axis) and to the ACES-level (x-axis) for each indicator grouped per dimension (detailed values in the supplement Fig. A.4.).

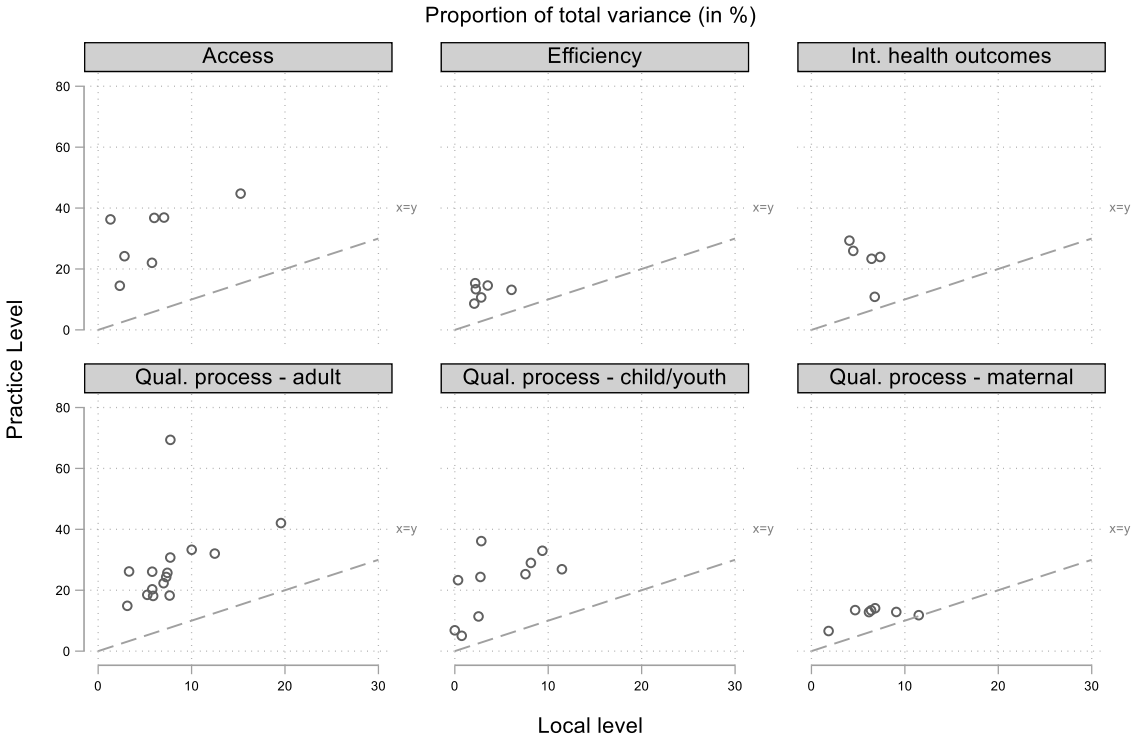


Figure 1.2: Variance partition coefficient from GP indicators per dimension in 2018.

Note: The variation attributed to the practice-level (plotted in the y-axes) and to local-level (at the x-axes). The regressions include all the confounders at the physician-, practice-, and ACES- levels, including physician demographics, patient list size, patient demographics and medical conditions, practice, and ACES characteristics. Each dot represents a different indicator.

Overall, the estimates are notably distant from the origin, indicating a substantial portion of variation attributable to practice and ACES features, rather than the individual GP. Before the adjustment (not plotted), the results revealed that between 8% and 54% (average 29%) of the total variation was attributed to the GP practice level, with 3% to 50% (average 20%) attributed to the ACES level. However, after adjusting for GP and context characteristics, as plotted in Figure 1.2, the proportion of unexplained variation in the adjusted indicator that could be attributed to these two levels was on average lower, i.e., variation was mostly inter-physicians. The contribution of GP practices varied substantially, depending on the indicator, ranging from 5% to 69% (mean 0.227, mode 0.134, SD 0.11). Meanwhile, ACES accounted for 0% to 20%

of the variation (mean 0.59, mode 0.58, SD 0.038). The remaining variation, approximately 71% on average, was attributed to random variation at the physician level, unrelated to observable physician or patient list characteristics. It's worth noting that the range of contribution is quite wide overall but more consistent within groups of indicators. Notably, the estimates from the efficiency indicators were relatively smaller in magnitude, ranging from 9% to 15% for the practice and 2% to 6% for ACES. This suggests that practice and ACES characteristics exert limited influence over the prescription of drugs and laboratory tests by physicians.

All Variance Partition Coefficient (VPC) ratios were situated above the diagonal, indicating that the variability in performance across practices within ACES was stronger than the disparities observed across different ACES. This is particularly noticeable in the context of access indicators, where the VPC ratios are further above the diagonal. A substantial proportion of the unexplained variation, ranging from approximately 16% to 60%, is attributed to practices. Notably, the "proportion of patient visits on the day of the appointment" (344) showed 45% of unexplained adjusted variation attributed to the practice and 15% to the ACES. Other indicators with a considerable portion of the variation attributed to the practice are those measuring the “number of remote visits within 72h after request” (335), and “visits scheduled within 15 days after request” (342), possibly reflecting the GP reliance on practice coordination capacity for this indicator. These results are in line with previous studies in the UK that have found that 20% of the variation in patients’ reported experience of the wait for an appointment occurs at the level of the practice (Salisbury, Wallace, & Montgomery, 2010).

Patient health outcomes indicators also demonstrated dependency on the practice. While most of the unexplained variation is at the GP level, a noteworthy 11% to 29% was attributed to the specific practice where the GP works, and only 4% to 7% to the ACES. This suggests that successful management of patients with chronic conditions can be influenced by practice-level

factors, with a lesser impact from the ACES. Similar findings were observed in an earlier study on diabetic patient management in the United States (Krein et al., 2002), where the greatest amount of variance was attributable to the facility level.

The indicators assessing the quality of the process of care for children, young people, and adults show a stronger influence from the practice level than from the ACES. For the adult population, three indicators stand out: the risk assessment of foot ulceration for diabetic patients (261) (VPC 69% practice and 8% ACES), adult vaccination indicators for tetanus (098) (VPC 42% practice, and 20% ACES), and flu vaccines for elderly patients (030) (VPC 32% practice, and 12% ACES). This is not a surprising result since each ACES has a team monitoring the implementation of the National Vaccination Program, in which tetanus vaccine is included and the team often coordinate the seasonal flu vaccination.

Intriguingly, the indicator for visits related to smoking habits cessation (277) exhibits a considerable contribution from the practice where the patient is registered (VPC 33% practice, and 10% ACES) and registration of smoking habits (47) (VPC 24% practice, and 7% ACES), while visits for alcohol consumption cessation show most of the variation between physicians (VPC 15% practice, and 3% ACES). This suggests that the quality of care for patients seeking assistance in quitting smoking heavily relies on the specific practice they are registered with, while the variation in visits for alcohol consumption cessation is more dependent on individual physicians. In fact, while smoking cessation visits are organised either at a practice- or at a local-level, alcohol cessation visits are often organised outside PHC at specific institutions for addictive behaviours. For maternal care indicators, related to pregnancy surveillance, the contributions from both practice and ACES are reduced to below 20% and 10%, respectively. Most of the variation lies between physicians, indicating that the quality of care for pregnant women primarily depends on the physician rather than the care organization.

Overall, these findings align with previous studies, which observed that GPs working within the same practice were more similar to their colleagues in the same practices than to GPs in other practices (J. de Jong, Groenewegen, & Westert, 2018). The limited contribution from the ACES on the quality of process may be attributed to the findings on multidisciplinary collaboration and strategies for distributing information (evidence-based learning and resources) to change physicians' practices. Research in this area has yielded mixed effects of such interventions (Forsetlund et al., 2021; Hansen & Drivsholm, 2002; Scott & Coote, 2009).

1.5.3 Comparison between types of practice

In this subgroup analyses, we distinguish between two distinct practice types: "PHCU," representing physicians working in conventional salaried primary care units, and "FHU," comprising physicians operating within team-based Family Health Units. This differentiation is important due to notable disparities in achievement rates observed among these practice types.

In general, physicians in team-based practices (both FHU-A and FHU-B) generally outperform those in non-team-based practices (PHCUs) in achieving the defined targets. FHU physicians tended to have lower percentages of physicians who did not achieve the target, and higher percentages of physicians who achieved the accepted and expected intervals for both incentivized and non-incentivized indicators compared to PHCU physicians. For a comprehensive breakdown of these findings, please refer to summary Table 1.5 in the appendix, which provides a detailed overview of the percentage of physicians by type of practice who successfully met the specified targets employed in the commissioning process.

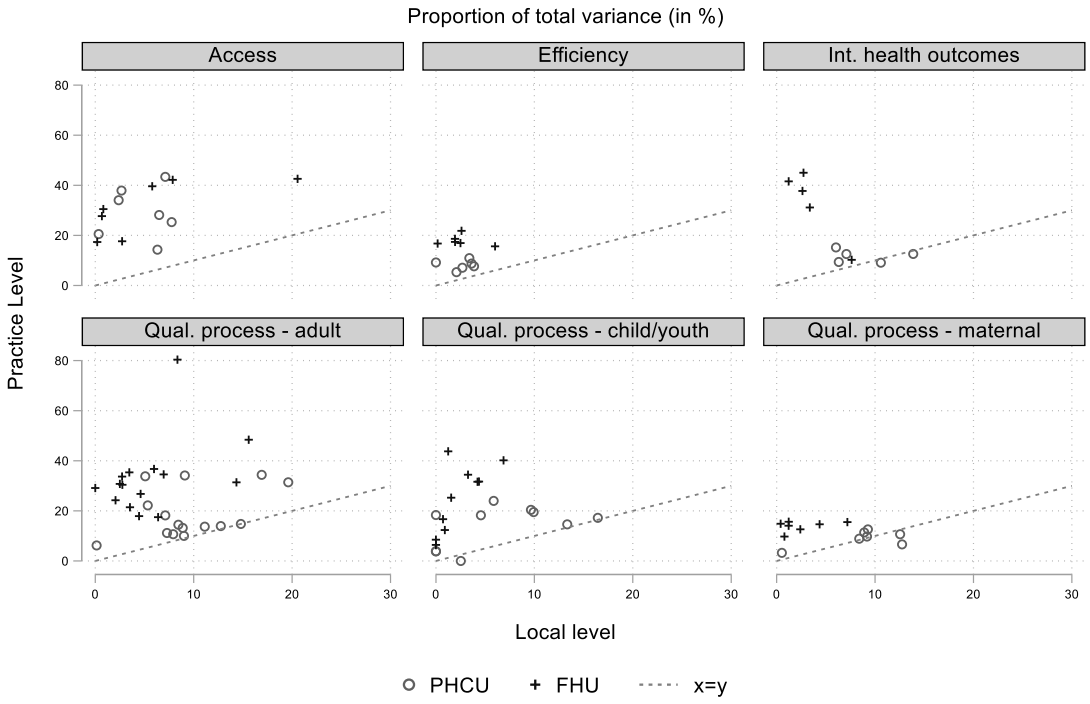


Figure 1.3: Variance partition coefficient from GP indicators per dimension and type of practice in 2018.

Note: The variation attributed to the practice-level is plotted in the y-axes and to local- level at the x-axes. The team-based practices are categorized as FHU (+) and non-team-based practices as PHCU (o). The regressions include all the confounders at the physician-, practice-, and ACES- levels, including physician demographics, patient list size, patient demographics and medical conditions, practice and ACES characteristics. Each dot represents a different indicator.

Across both types of practice, we consistently found that the variation in physician achievement across practices within ACES was larger in magnitude than the variation across different ACES, as depicted in Figure 1.3. Specifically, for PHCU, the average variation across practices was 15.9%, while the variation across ACES was 7.2%. Similarly, for FHU, the average variation across practices was 26.7%, whereas the variation across ACES was 3.9%.

The contribution of each organizational level to the variation in physician performance slightly differed based on the type of indicator. In PHCU, the unexplained variation in intermediate health outcomes was attributed to the practice at a range of 9% to 13%, whereas in FHU,

variation had a much larger contribution from the practice, ranging from 10% to 45%. Conversely, the ACES contribution in PHCU was between 6% and 14%, but in FHU, it was smaller for most indicators, ranging from 1% to 8%. These results indicate that compared to PHCU practices, the variation in intermediate outcomes indicators of physicians in FHU predominantly reflects the influence of the practice where they work, with a lesser impact from the ACES. This suggests that certain practices within FHU may prioritize and allocate specific time to achieve these indicators, possibly organizing regular meetings to discuss and improve their performance.

Similar conclusions were drawn from the analysis of process indicators, with one exception for the indicator related to flu vaccinations for at-risk patients (30). In this case, we observed a relatively larger contribution of the ACEs to physician performance variation in FHU compared to PHCU.

In terms of access, we noticed some heterogeneity between indicators referring to access to appointments and those related to utilization. Notably, the variation in the rate of same-day visits (344) was large and identical in both PHCU and FHU. However, the contribution of ACES to the unexplained variation was strikingly larger in FHU (21%). This suggests that unjustified variation in providers' capacity to provide care on the day of booking in FHU is partly due to unmeasured local factors, such as geographical isolation of the population and transportation infrastructures.

For efficiency indicators related to physician-initiated costs for the treatment of chronic conditions patients, we found in both PHCU and FHU that the variation was substantially at the GP level, as in the aggregate analysis. The ACES contribution to these indicators was marginal (under 6% in both types of practices). However, the unexplained variation in physician-initiated spending could be attributed to practice-level factors to some extent in FHU, ranging from 16% to 22%, but to a lesser extent in PHCU, ranging from 5% to 11%. These

results suggest that most of the unexplained variation across physicians in PHCU can be attributed to physician characteristics and their patient lists, while practice-level factors play a more significant role in the spending heterogeneity across physicians in FHU. This indicates that certain practices within FHU may have distinct approaches to managing chronic condition patients' costs and utilization of resources.

1.5.4 Comparison between indicators linked with incentives

In our analysis, we also distinguish between the indicators that could be used in 2018 in the commissioning process with associated financial incentives (referred to as incentivized) and those that were not (non-incentivized), as illustrated in Figure 1.4. As indicated by the descriptive statistics presented in Table 1.5 in the appendix, it is evident that physicians achieve incentivized indicators at higher rates when compared to non-incentivized indicators, irrespective of their practice type.

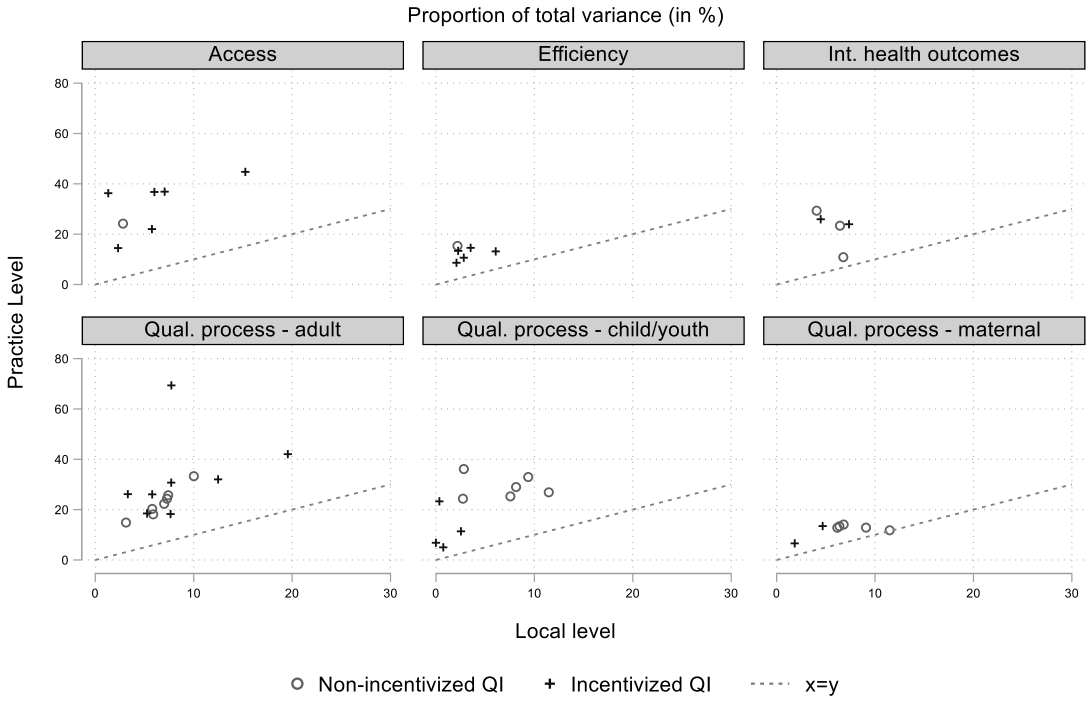


Figure 1.4: Variance partition coefficient from GP indicators per dimension and type of indicator, incentivized (+) and non-incentivized (o).

Note: The variation attributed to the practice-level is plotted in the y-axes and to local-level at the x-axes. The regressions include all the confounders at the physician-, practice-, and ACES- levels, including physician demographics, patient list size, patient demographics and medical conditions, practice, and ACES characteristics. Each dot represents a different indicator.

The graphs show that the contribution from the practice and ACES to the variation in the dimensions of quality of the process for child/youth and maternal health is stronger for the non-incentivized indicators compared to the incentivized ones. In these dimensions, the proportion of unexplained variation at the GP level for incentivized indicators exceeds that of non-incentivized indicators by a factor of 1.1 or more.

However, for the remaining dimensions, no discernible difference emerges between incentivized and non-incentivized indicators. This observation highlights the significant impact of performance-based rewards in positively shaping physician behavior and motivation to attain specific targets and uphold quality standards, irrespective of the practice type or the level of support they may receive from their peers and management. Additionally, this finding may suggest a potential positive spill-over effect of incentives, aligning with previous research (da Luz Pereira et al., 2021).

1.5.5 Analysis of the determinants of achievement

To analyse the factors associated with physician achievement across diverse quality indicators, we performed a pooled analysis of all indicators per dimensions. In order to perform this analysis we do not analyse the absolute achievement but the achievement relative to each indicator target. The dependent variable is converted into an achievement score, classified as 0, 1, or 2, for each indicator. To study the likelihood of physicians reaching the highest score, we employed a multilevel mixed-effects ordered logit estimation. The three-level model

accounts for unobserved heterogeneity between physicians using two random components: practice and ACES. Additionally, fixed effects were incorporated to adjust for variations across different indicators and unobserved differences across the five regional authorities in Portugal (ARS).

Table 1.1 Multilevel mixed effects ordered logit regressions of the probability of physicians achieving the expected/acceptable intervals in all indicators grouped by dimension.

	Outcomes (1)		Process (2)		Access (3)		Efficiency (4)	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Physician-level covariates								
GP Age	-0.0014	(0.00296)	-0.0011	(0.00106)	-0.0014	(0.00132)	-0.0025	(0.00246)
GP gender=female	0.1	(0.0694)	0.02	(0.0246)	0.038	(0.0313)	0.005	(0.0579)
1st GP list size quintile	[ref]							
2nd GP list size quintile	0.1	(0.12)	0.055	(0.0436)	-0.065	(0.0538)	-0.05	(0.0976)
3rd GP list size quintile	0.13	(0.134)	0.099**	(0.0486)	-0.083	(0.0593)	-0.18	(0.108)
4th GP list size quintile	0.23	(0.143)	0.12**	(0.0514)	-0.06	(0.0625)	-0.16	(0.114)
5th GP list size quintile	0.27*	(0.152)	0.14**	(0.0546)	-0.065	(0.066)	-0.074	(0.12)
GP list % Female	30.1***	(0.809)	0.68**	(0.286)	0.26	(0.361)	-0.88	(0.657)
GP list % 0 – 6 years	10.6	(20.87)	-0.012	(10.03)	2.7**	(10.28)	-2.3	(20.32)
GP list % >65 years	-20.8***	(0.942)	-0.56*	(0.334)	-20.2***	(0.429)	0.48	(0.784)
Practice-level covariates								
Size in FTE	-0.0012	(0.00545)	-0.00021	(0.00229)	0.0019	(0.00215)	-0.0039	(0.00368)
Ratio Nurses to GPs	0.053	(0.199)	0.052	(0.0829)	-0.03	(0.0817)	0.13	(0.142)
% Elderly in practice	0.022	(0.0169)	0.024***	(0.00699)	0.0072	(0.0069)	0.024*	(0.0123)
Unemployment rate	-0.14***	(0.0432)	-0.092***	(0.019)	-0.06***	(0.0157)	0.072**	(0.0296)
City	[ref]							
Towns and suburbs	0.1	(0.182)	0.14*	(0.0785)	0.086	(0.0698)	-0.11	(0.121)
Rural areas	0.46*	(0.246)	0.41***	(0.105)	0.23**	(0.0967)	-0.014	(0.167)
PHCU	[ref]							
FHU-A	2.9***	(0.178)	1.4***	(0.0732)	0.87***	(0.0726)	0.74***	(0.126)
FHU-B	4.6***	(0.166)	2.0***	(0.0658)	1.4***	(0.0659)	1.4***	(0.114)
Age of the practice	0.0076	(0.0208)	-0.0012	(0.00884)	-0.0018	(0.00844)	0.0031	(0.0148)
Distance from practice to ACES	0.0028	(0.0063)	0.0012	(0.00269)	0.0022	(0.00255)	-0.0093**	(0.0045)

Local group level covariates (ACES)								
Shared Physicians	-0.57*	(0.3)	-0.14	(0.145)	-0.16*	(0.0955)	-0.014	(0.198)
Shared Nurses	0.12	(0.105)	-0.00085	(0.0508)	-0.055	(0.0338)	0.015	(0.0697)
Integrated Care (LHU=1)	-0.38	(0.369)	-0.07	(0.177)	0.061	(0.12)	0.097	(0.244)
% Population not listed with GP	-0.034	(0.0221)	-0.02*	(0.0107)	-0.0068	(0.00699)	-0.011	(0.0145)
FHU-A in ACES (%)	-0.014*	(0.00795)	-0.0075**	(0.00381)	-0.0033	(0.00257)	0.00032	(0.00525)
FHU-B in ACES (%)	-0.015**	(0.00705)	-0.0059*	(0.00339)	-0.0044*	(0.00227)	0.009*	(0.00468)
Indicator FE	Yes		Yes		Yes		Yes	
Region FE (ARS)	Yes		Yes		Yes		Yes	

(continues)

(continuation)	Outcomes (1)		Process (2)		Access (3)		Efficiency (4)	
ƙ1 cutpoint	-0.99	(0.918)	-3.7***	(0.408)	-0.71**	(0.339)	-0.085	(0.643)
Ƙ2 cutpoint	0.69	(0.918)	-3***	(0.408)	0.51	(0.339)	0.76	(0.643)
σ2 Local group (Aces)	0.29***	(0.0909)	0.078***	(0.0213)	0.018**	(0.00885)	0.11***	(0.0372)
σ2 Unit	1.2***	(0.113)	0.25***	(0.0192)	0.16***	(0.0178)	0.4***	(0.0539)
σ2 Physician	2.3***	(0.115)	0.32***	(0.0121)	3.9e-34	(70.20e-19)	1.9***	(0.085)
N	22380		143232		31332		26856	
n Physicians	4,502		4,502		4,502		4,502	
n Practices	831		831		831		831	
n local Custer	55		55		55		55	
Likelihood-Ratio Test	-16985.0		-107964.9		-20409.1		-21827.3	
P-value (Assumption: 2-level nested within 3-level)	0.00		0.00		0.00		0.00	

Note0. The dependent variable is the physician level of the achievement (0,1 or 2) of the indicator from each group of indicators0. Coefficients of three-level multilevel mixed-effects ordered logit estimation that uses two random components (practice and local group) to account for part of the unobserved heterogeneity between physicians0. Fixed effects were included to account for differences across indicators, and for unobserved differences between the five regional authorities in Portugal (ARS)0. Robust SEs clustered on practices are in brackets, *** p<0.01, ** p<0.05, * p<0.10.

To analyse the factors associated with physician achievement across diverse quality indicators, we performed a pooled analysis of all indicators per dimensions. In order to perform this analysis we do not analyse the absolute achievement but the achievement relative to each indicator target. The dependent variable is converted into an achievement score, classified as 0, 1, or 2, for each indicator. To study the likelihood of physicians reaching the highest score, we employed a multilevel mixed-effects ordered logit estimation. The three-level model accounts for unobserved heterogeneity between physicians using two random components:

practice and ACES. Additionally, fixed effects were incorporated to adjust for variations across different indicators and unobserved differences across the five regional authorities in Portugal (ARS).

Table 1.1 presents coefficients, standard errors, and model fit statistics. Likelihood-ratio tests show sufficient variability between ACES and practices to justify using a mixed-effects model over standard ordered logistic regression.

The results for the physician level factors are mixed. On the one hand physician attributes, such as age and gender, are not significantly associated with their achievements after considering patient list specifics and the structural characteristics of healthcare provision. The patient list's size and composition emerge as significant predictors of GP performance concerning outcome and process indicators. Specifically, GPs in the 5th quintile (more than 1865 patients per GP) exhibit a higher likelihood of reaching the highest score compared to those in the 1st quintile.

The correlation between list sizes and achievement levels suggests that GPs managing larger patient lists are more likely to meet the targets. This aligns with previous research conducted in Portugal, which indicates that physicians with larger patient lists tend to allocate more of their workday to direct patient care and less to coordinating, teaching, and management tasks (Granja, Ponte, & Cavadas, 2014). These non-clinical activities often compete for time and are not factored into the physician indicators. Furthermore, it's worth noting that FHUs typically have larger lists of registered patients compared to PHCUs, so the practice type incentives could potentially confound the results related to list size. In terms of patient list composition, we observed that the likelihood of a physician meeting quality targets increases with a higher

proportion of female patients on their list. Specifically, for each unit increase in the proportion of women on a GP's list, the ordered log-odds of achieving a higher score increase by 0.27 for outcomes and 0.14 for process indicators, given variables remain constant. This trend aligns with recognized gender differences in health care seeking behaviour and utilization (Bertakis, Azari, Helms, Callahan, & Robbins, 2000). Moreover, GPs with a higher ratio of children on their lists are more likely to reach access targets, while GPs with older patient lists tend to underachieve. The nature of medical encounters and their associated challenges differ markedly among age groups, as the disease burden and the risk of multimorbidity rise sharply with age, introducing complexity in diagnosis and treatment (Prince et al., 2015).

In terms of practice-level determinants, neither workforce size nor skill composition significantly influenced achievement levels. Prior research using routinely collected data in the UK emphasized that the strongest predictors of clinical care quality weren't necessarily nurse staffing levels but their interplay with organizational elements (Griffiths, Maben, & Murrells, 2011). On the other hand, external factors, such as the practice's location and the surrounding population's socio-economic conditions, are important in defining GP performance metrics. Physicians in urban areas, have a higher likelihood of underperforming compared to those in rural areas, towns, and suburban areas. Physicians serving an urban population face distinctive challenges: they serve both a transient population and those with easier access to private healthcare. This can disrupt the continuity of care, a key factor in achieving quality targets. Interestingly, practices further from the ACES central office tend to exhibit poorer performance in cost containment metrics, suggesting possible proximity-driven influences from ACES, which does not reveal to be significant in the quality of care and access to care indicators. Regarding the socio-economic conditions, the likelihood of meeting access and quality targets diminishes as the local unemployment rate rises. A recent study had also found a significant relationship between unemployment and the commissioning indicators (Pereira et al., 2022).

Conversely, in our study, areas with more adverse socio-economic circumstances showed a higher likelihood of achieving cost containment (efficiency) targets. Such observations hint at a potential bias in prescription decisions based on the socio-economic profiles of the population being served. These findings imply that prescription cost decisions might be biased by the socio-economic profiles of the served community. This further underscores the notion that the performance of General Practitioners is intrinsically linked to the socio-economic environment of their patient base. Finally, GPs in team-based FHU types of practice were more likely to achieve the targets than the non-team-based practices. The magnitude of the positive effect is higher for those practices with P4P (FHU-B). These findings are not surprising as they are partially related to incentive structures and the commissioning process of FHU.

Lastly, At the local (ACES) level, we found that larger groups, measured by the number of physicians in shared services, often had fewer GPs meeting targets in access and preliminary health outcomes. This implies that even if larger groups have more resources, these might be spread too thin, potentially hindering individual performance. This contrasts with the performance in smaller ACES. However, since we couldn't determine the specific roles of these physicians within shared services, it's possible that a higher number of doctors could be taking on more administrative duties. Moreover, the fact that larger ACES exist predominantly in rural and inland areas of the country, this coefficient might suffer from the confounding effect of the distance to primary care, and to specialist care. The practice's affiliation with a vertically integrated healthcare system (LHU) didn't significantly influence achievement across any metric. While earlier research emphasized certain challenges associated with the vertical integration in (C. Costa, Santana, & Lopes, 2013; Santana, R. & Costa, 2008), they also did not find evidence to suggest a beneficial impact on access (Amado, Ferreira, & Nunes, 2022). However, a negative effect on cost metrics was observed (Ibid.). Additionally, as the proportion of patients without a designated physician in an area increases, the chances of meeting the

process quality targets drop. This trend is not surprising. Typically, these patients can only access care during restricted open visit times in the GP's schedule and often arrive with more acute conditions, adversely affecting preventive care measures. Furthermore, treating these patients doesn't contribute to the physician's indicators, hence diverting valuable time from the care of list patients.

1.6 Discussion

1.6.1 Summary

Our analysis unveiled substantial variation in physicians' practices across multiple quality dimensions, highlighting the inherent complexity of delivering consistent care across regions.

A central element of our research was discerning the primary care hierarchical organization's relative contributions to variations across different quality dimensions. While prior research has explored variation in specific conditions or procedures (Corallo et al., 2014), our study delineates the variation in GPs' achievement across 50 measures of primary care practice spanning six dimensions (process quality, intermediate health outcomes, access, efficiency measures). Leveraging comprehensive nationwide administrative data, which encompasses all GPs, and multilevel linear random intercept models, we distinguished between GP, practice, and ACES sources of variation in primary care quality for the Portuguese population in 2018.

We find that the local groups and practices where GPs provide care significantly influence the inter-physicians adjusted variation observed. Our findings indicated that an average of 22% of the unexplained variation in PHC indicators is influenced by the practice itself, and an additional 6% on average by ACES. Interestingly, our analysis revealed nuances within different quality dimensions. In terms of access and preventive care, the contribution of the practice was comparable to that of the physician. The fact that more than 16% of the variation

in indicators of access to GP visits occurs at the level of the practice provides support for the use of these measures to assess practice performance but may poorly reflect GPs' efforts.

In contrast, efficiency-related dimensions, especially prescription costs, are predominantly influenced at the physician level, emphasizing the broader physician discretion in achieving these indicators. This can be interpreted in light of previous evidence suggesting that physicians are keenly aware of costs when making prescribing decisions (Sá et al., 2015). Similarly, for measures assessing intermediate health outcomes and maternal care quality, the practice's influence is limited as these outcomes are contingent on factors like patient adherence to treatment and subsequent specialized care. Variations in care processes for patients with chronic conditions and children also reflect the contribution practices where GPs work and, to a smaller degree, the ACES. Quality indicators for adult care, especially preventive care services, indicate a substantial influence from local-level services (ACES).

Moreover, our evaluation of different practice types, specifically PHCU and FHU, provided valuable insights into the role of practice type in shaping variations. Our findings indicated that both types of practices exhibited greater intra-ACES variation compared to inter-ACES differences. FHUs, in comparison to PHCUs, displayed a more pronounced independence from ACES when achieving health outcome and process quality indicators, hinting at the distinct strategies FHUs might employ to attain these objectives.

An interesting aspect of our study was the analysis of indicators potentially linked to financial incentives on performance through the commissioning process. Incentivized quality indicators for children and maternal health showed a diminished practice contribution, underscoring the potential of financial incentives to offset the organization's influence. In other care dimensions, contributions from practices and ACES were comparable for all indicators irrespective of being linked to incentives. This observation underscores the substantial influence of performance-based incentives in fostering changes in physician motivation and behaviour, as they may drive

physicians to achieve specific targets and maintain high-quality standards, regardless of their practice type or the level of support they may receive from peers and management.

In our analysis of the determinants shaping GPs' relative achievements vis-à-vis the defined targets in section 1.5.5, we find some interesting results that support and provide more detail our prior conclusions. Regarding the variation that arises at the physician level that was deemed predominant in most of the indicators studied, these were more likely attributed to patient factors than physician attributes. Specifically, attributes like physician age and gender had negligible impacts on performance achievements. However, the demographic composition of patients significantly predicted GP performance across most dimensions, except efficiency. We found that this must be mostly due to the patients' factors than physician factors. Particularly, that physician attributes such as age and gender do not significantly affect their performance achievements. However, the size and age and gender composition of a patient list significantly predict GP performance in almost all dimensions (except efficiency). These results suggest that patient demand-side factors might explain variations in access and quality at physician level to a greater extent than physician supply-side factors —a conclusion consonant with prior studies examining variations in sickness absenteeism duration in Norway (Aakvik et al., 2010) and other contexts where patient factors were more influential than GP or municipality attributes. The exception was made to the efficiency indicators linked with cost containment for which no significant association was found neither with physician age and gender nor patient characteristics. This aligns with findings from the US, suggesting that only physician beliefs influenced Medicare expenditure variations, although such beliefs weren't measured in our study (Cutler et al., 2019).

The understanding of the influence of practice and ACES in inter-physician variation is further expanded in the determinants' analysis. This analysis revealed that neither workforce size nor skill-mix significantly influenced achievement levels. However, external factors, particularly

the socio-economic conditions and practice location, play an important role. Socio-economic conditions, especially the local unemployment rate, influences physician achievement of the targets. Physicians in areas with poorer socio-economic statuses are more likely to underperform in almost all indicators, except for efficiency. Similarly, urban-based physicians also underperform, likely due to unique continuity-of-care challenges. Meanwhile, practices located further from the ACES central office tend to exhibit poorer performance in cost containment metrics, suggesting proximity-driven influences from ACES in efficiency measures. However, this influence did not extend to the other quality or access metrics.

In terms of physician greater discretion on efficiency measures found in the variation analysis is supported by the analysis of the determinants, as areas with more adverse socio-economic circumstances showed a higher likelihood of achieving cost containment (efficiency) targets, implying potential bias in prescription decisions based on the socio-economic profiles of the population being served. The determinants analysis also illuminated how practice types, specifically PHCU and FHU, could influence variation, by pointing out that GPs in team-based FHU types of practice were more likely to achieve targets in all dimensions than non-team-based practices.

Finally, at the ACES level, there appeared to be an inverse relationship between the size of shared local group resources and GPs achieving targets. This suggests that larger groups, despite abundant resources, might be forced to dilute resources and support, thereby affecting the degree of support or resources available to physicians, compared to smaller ACES. Such a finding might elucidate why ACES's contribution to adjusted variation remained marginal for most indicators. Nevertheless, it is important to note that the relationship between ACES size and GP achievement, may be confounded by other factors determining the size of the ACES, such as the scattered populations, and the population distance to hospitals, which in turn also affect the GP indicators.

In summary, the factors associated with physician achieved align with existing literature examining variation in ambulatory care, which has explored potential reasons for variation. Previous studies have identified factors such as access or resource constraints and socioeconomic determinants as significant contributors to variability (Corallo et al., 2014). Contrary to some studies that pinpoint the absence or inadequate adherence to guidelines by doctors as a cause for variation in practice, our findings do not indicate that physician attributes play a significant role in determining GP outcomes. It's important to note, however, that many of these studies focus on use of imaging, a specific aspect, in contrast to the broader dimensions we considered in our analysis.

1.6.2 Limitations

Some methodological shortcomings in the current study must be considered when assessing these findings. In the first place, our interpretation of the contribution of practice-level and local -level are based on administrative quantitative data. Qualitative studies would complement the GP perceived contribution stemming from different levels that are affecting each dimension of quality of care. Additionally, it would be important to quantify practice-level and local-level clinical support services and potential hindrances to GPs work, that are difficult to obtain in observational studies using routinely collected data. Some authors were able to analyse the effect of local-level networks support for practice nurses and IT support (Scott & Coote, 2009), and measures of clinical support arrangements and autonomy over the structure of care delivery (Yano, Soban, Parkerton, & Etzioni, 2007) that are important determinants of physician's performance.

Secondly, estimating practice variation with physician-level aggregated data creates some difficulties to disentangle the agency role of the physician. While the variation analysis

primarily looked at broader factors like practice and ACES, the determinants' analysis emphasized the importance of patient list size and composition as significant predictors of GP performance. The challenge of distinguishing whether the variation is shaped by the patient demand or by the supplier in aggregated data has received much attention within the literature on supplier-induced demand (McGuire, 2000a). To address this challenge, our strategy involved adjusting the estimation for patient list demographic and morbidity data, a method commonly employed in similar studies (Grytten & Sørensen, 2003). Further analysis would benefit from studying this issue using patient-level information.

Thirdly, it is not implied in this study that the variation is always problematic. In this study it was not possible to distinguish between warranted and unwarranted variation, as neither does most of the literature (Harrison et al., 2019). These aspects will be key considerations for future studies on the organizational sources of primary care quality variation. Upcoming research could also be valuable in extending this analysis to further dimensions, such as ambulatory care utilization, not yet studied in Portugal (van Loenen et al., 2014).

Another limitation pertains to our study's exclusive reliance on data from 2018, without the capacity to observe variation within individual physicians over time. However, a previous study in Israel identified greater variation in practice patterns between physicians than within individual physicians over time (Shashar et al., 2021).

Lastly, our study primarily evaluates the extent of practice variation and the contributing factors but does not ascertain the implications or meaning of this variation. It remains uncertain whether this variation signifies wasteful practices, poses obstacles to patients' interests, or is associated with health care inequities. Policymakers should express concerns not only about variations in practice due to potential wastefulness but also because they may not consistently prioritize the best interests of patients. If practice variation is primarily driven by patient dissimilarities rather than factors such as payment methods, geographic location, or systemic

biases, the healthcare system should accommodate it. Future investigations should contrast the absolute results with desirable clinical and efficiency targets and conduct subgroup analyses focusing on specific population demographics.

Unlike previous studies that reported poor reliability at the physician or facility level due to limited proportional variability and small sample sizes per unit (Fung, Schmittdiel, Fireman, Meer, Thomas, Smider, Hsu, & Selby, 2010), this study relies on strong variability between practitioners and a large sample even at the local-level.

1.6.3 Policy and Practice Implications

The implications of our findings extend to healthcare policy and practice. Our findings underscore the limitations of assessing physician performance solely based on individual indicators, emphasizing the necessity for a holistic perspective that recognizes the roles of practices and ACES in achieving quality targets.

This study contributes to the ongoing debate on the suitability of primary care indicators for measuring physician performance. The substantial influence from the institution where the GP works observed in the variation levels of certain indicators raises concerns about their adequacy in fairly assessing physician performance. This issue is particularly pertinent given the increasing reliance on indicator-based accountability, especially when tied to financial incentives (Campbell, Braspenning, Hutchinson, & Marshall, 2002; Marshall & Roland, 2017). A poorly chosen set of measures carries the risk of unintended consequences, such as misinterpretation or "measure fixation" (Jones, Shepherd, Wells, Le Fevre, & Ameratunga, 2014; Lester, Hannon, & Campbell, 2011; P. Smith, 1995). Moreover, the proliferation of indicators can strain resources allocated to data collection and analysis, as exemplified by the numbers of indicators included in Pay-for-Performance schemes across different countries ranging from 11 in the Primary Care Renewal Models, to 134 in the UK QOF (Eijkenaar, 2012).

Furthermore, our results highlight the need to examine best practices at all levels, including practice and local management. A common objective in this literature is aimed at identifying potential targets for behavioural modification interventions to reduce variation and improve healthcare quality and outcomes (Shashar et al., 2021). When formulating strategies to enhance healthcare quality, it is crucial to acknowledge the substantial roles played by practices and ACES in specific dimensions. Previous research has identified levers for driving change in primary care, including financial incentives, performance feedback, peer learning, and leadership support (Gosling, Mays, Erens, & Reid, 2019). Effective improvement strategies must align priorities across various structures and resonate with clinicians' personal motivations, drawing from a well-established body of literature on incentive structures within organizations (e.g., (Holmstrom & Milgrom, 1994).

Another policy implication pertains to local grouping initiatives aimed at harmonizing clinical and management practices. While these policies aim to create networks among primary care providers, enhance information sharing, and support commissioning and resource allocation, our findings suggest that their impact on physician achievement varies across dimensions of care. These structures significantly influence certain preventive campaign indicators (such as tetanus and flu vaccination, TSHPKU tests in babies, and third-trimester pregnancy tests) but appear to play a negligible role in health outcomes and cost targets.

Recognizing the ways in which hierarchical structures impact physician practice can guide quality improvement strategies that target these local structures. Although not the central focus of our analysis, understanding mechanisms like social capital and “social influence mechanism” (i.e., the performance of peers to whom the GP was directly connected) within collaborative networks can shed light on the impact of these networks on practice variation. Prior research has shown that the performance of peers connected to a GP can yield a stronger influence than the GP's centrality in the network. This insight suggests that collaborative

arrangements can promote more homogeneous behaviour among GPs but may not necessarily enhance their ability to meet specific healthcare objectives set by local authorities (Fattore, Frosini, Salvatore, & Tozzi, 2009)

In conclusion, our study enriches the expanding body of literature examining medical practice variation. We emphasise the significance of primary healthcare organizational structures in influencing physician performance, offering a nuanced view of healthcare quality determinants. Our research identifies specific dimensions of quality of care and performance where disparities emerge and distinguishes those predominantly influenced by physician behaviour from those shaped by organizational factors.

Further analyses are necessary to detail the institutional mechanisms through which organizations shape clinical activities and achievements, encompassing network effects, provided services, managerial styles, and cultural influences. Additionally, identifying best practices at each level will be instrumental in advancing healthcare quality and performance optimization.

As healthcare systems continue to evolve, our findings emphasize the need to consider organizational dynamics in tandem with individual physician characteristics when assessing performance and devising quality enhancement strategies. Our study also underscores the value of select indicators from a comprehensive set, as they demonstrate the capacity to differentiate between the performance aspects of local-level, practice-level, and individual GPs within those practices. Furthermore, our results inform the design of future research studies in this domain and highlight the utility of multilevel modelling in investigating sources of variation.

By accounting for the hierarchical structure of these data, reflecting correlations within facilities and local structures, our estimates offer greater realism compared to prior single-level studies. Replicating similar information and methods in analogous contexts can better equip

researchers and policymakers to strategically plan and target interventions for cost control, quality improvement, and variation reduction, as well as evaluate their effectiveness post-implementation.

1.7 References

Aakvik, A., Holmås, T. H., & Kamrul Islam, M. (2010). Does variation in general practitioner (GP) practice matter for the length of sick leave? A multilevel analysis based on Norwegian GP-patient data. *Social Science and Medicine*, 70(10), 1590–1598. Retrieved from https://doi.org/10.1016/j.socscimed.2010.01.031

ACSS. (2017). *Operacionalização da Contratualização de Cuidados de Saúde Primários 2018*.

Amado, G. C., Ferreira, D. C., & Nunes, A. M. (2022). Vertical integration in healthcare: What does literature say about improvements on quality, access, efficiency, and costs containment? *The International Journal of Health Planning and Management*, 37(3), 1252–1298. Retrieved from https://doi.org/10.1002/hpm.3407

Barros, P. P. (2017). Competition policy for health care provision in Portugal. *Health Policy*, 121(2), 141–148. Retrieved from https://doi.org/10.1016/j.healthpol.2016.12.005

Barros, P. P., Machado, S. R., & Simões, J. de A. (2011). Portugal. Health system review. *Health Systems in Transition*, 13(4), 1–156. Retrieved from https://doi.org/10.1136/bmj.4.5787.575-a

Barros, P. P., & Simões, J. D. A. (2007). Portugal: health system review. *Health Systems in Transition*.

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).

Biscaia, A. R., & Heleno, L. C. V. (2017). Primary Health Care Reform in Portugal: Portuguese, modern and innovative. *Ciência & Saúde Coletiva*, 22(3), 701–712. Retrieved from https://doi.org/10.1590/1413-81232017223.33152016

Boerebach, B. C. M., Scheepers, R. A., van der Leeuw, R. M., Heineman, M. J., Arah, O. A., & Lombarts, K. M. J. M. H. (2014). The impact of clinicians' personality and their interpersonal behaviors on the quality of patient care: a systematic review. *International Journal for Quality in Health Care*, 26(4), 426–481. Retrieved from https://doi.org/10.1093/intqhc/mzu055

Bonciani, M., Schäfer, W., Barsanti, S., Heinemann, S., & Groenewegen, P. P. (2018). The benefits of co-location in primary care practices: The perspectives of general practitioners and patients in 34 countries. *BMC Health Services Research*, 18(1). Retrieved from https://doi.org/10.1186/s12913-018-2913-4

Branco, A. G., & Ramos, V. (2001). Cuidados de saúde primários em Portugal. *Revista Portuguesa de Saúde Pública*, 5–12.

Brito Fernandes, Ó., Lopes, S., Marques, A. P., Moita, B., Sarmento, J., & Santana, R. (2019). Local Health Units in Portugal: The Influence of Chronic Conditions on Inpatient Readmissions. *Portuguese Journal of Public Health*, 37(2–3), 91–99. Retrieved from https://doi.org/10.1159/000506015

Brownlee, S., Chalkidou, K., Doust, J., Elshaug, A. G., Glasziou, P., Heath, I., … Korenstein, D. (2017). Evidence for overuse of medical services around the world. *The Lancet*, 390(10090), 156–168. Retrieved from https://doi.org/10.1016/S0140-6736(16)32585-5

Busse, R., Klazinga, N., Panteli, D., Quentin, W., WHO, Busse, R., … Organization, W. H. (2019). *Improving healthcare quality in Europe: characteristics, effectiveness and implementation of different strategies*. (R. Busse, N. Klazinga, D. Panteli, & W. Quentin,Eds.) (Health Pol). Copenhagen PP - Copenhagen: World Health Organization. Regional Office for Europe,Organisation for Economic Co-operation and Development (OECD).

Campbell, S. M. (2002). Implementing clinical governance in English primary care groups/trusts: reconciling quality improvement and quality assurance. *Quality and Safety in Health Care*, 11(1), 9–14. Retrieved from https://doi.org/10.1136/qhc.11.1.9

Campbell, S. M., Braspenning, J., Hutchinson, A., & Marshall, M. (2002). Research methods used in developing and applying quality indicators in primary care. *Quality and Safety in Health Care*, 11, 358–365. Retrieved from https://doi.org/10.1136/qhc.11.4.358

Cook, D. A., Pencille, L. J., Dupras, D. M., Linderbaum, J. A., Pankratz, V. S., & Wilkinson, J. M. (2018). Practice variation and practice guidelines: Attitudes of generalist and specialist physicians, nurse practitioners, and physician assistants. *PLOS ONE*, 13(1), e0191943. Retrieved from https://doi.org/10.1371/journal.pone.0191943

Corallo, A. N., Croxford, R., Goodman, D. C., Bryan, E. L., Srivastava, D., & Stukel, T. A. (2014). A systematic

review of medical practice variation in OECD countries. *Health Policy*, 114(1), 5–14. Retrieved from https://doi.org/10.1016/j.healthpol.2013.08.002

Correia, I., & Veiga, P. (2010). Geographic distribution of physicians in Portugal. *The European Journal of Health Economics*, 11(4), 383–393. Retrieved from https://doi.org/10.1007/s10198-009-0208-8

Costa-Font, J., & Pons-Novell, J. (2007). Public health expenditure and spatial interactions in a decentralized national health system. *Health Economics*, 16(3), 291–306.

Costa, C., Santana, R., & Lopes, S. (2013). *Custos e Preços na Saúde Passado, presente e futuro*. *Fundação Francisco Manuel dos Santos*.

Cunningham, F. C., Ranmuthugala, G., Plumb, J., Georgiou, A., Westbrook, J. I., & Braithwaite, J. (2012). Health professional networks as a vector for improving healthcare quality and safety: a systematic review. *BMJ Quality & Safety*, 21(3), 239–249. Retrieved from https://doi.org/10.1136/bmjqs-2011-000187

Cutler, D., Skinner, J. S., Stern, A. D., & Wennberg, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal: Economic Policy*, 11(1), 192–221. Retrieved from https://doi.org/10.1257/pol.20150421

da Luz Pereira, A., Ramalho, A., Viana, J., Pinto Hespanhol, A., Freitas, A., & Biscaia, A. (2021). The effect of commissioning on Portuguese Primary Health Care units' performance: A four-year national analysis. *Health Policy*, 125(6), 709–716. Retrieved from https://doi.org/10.1016/j.healthpol.2021.02.008

Damiani, G., Silvestrini, G., Federico, B., Cosentino, M., Marvulli, M., Tirabassi, F., & Ricciardi, W. (2013). A systematic review on the effectiveness of group versus single-handed practice. *Health Policy*, 113(1–2), 180–187. Retrieved from https://doi.org/10.1016/j.healthpol.2013.07.008

Davis, P., & Gribben, B. (1995). Rational prescribing and interpractitioner variation: a multilevel approach. *International Journal of Technology Assessment in Health Care*, 11(3), 428–442.

Davis, P., Gribben, B., Scott, A., & Lay-Yee, R. (2000). The "supply hypothesis" and medical practice variation in primary care: testing economic and clinical models of inter-practitioner variation. *Social Science & Medicine*, 50(3), 407–418. Retrieved from https://doi.org/10.1016/S0277-9536(99)00299-3

De Jong, J. D. (2008). Explaining medical practice variation. *Social Organization and Institutional Mechanisms. NIVEL*.

de Jong, J. D., Groenewegen, P. P., & Westert, G. P. (2003). Mutual influences of general practitioners in partnerships. *Social Science & Medicine*, 57(8), 1515–1524. Retrieved from https://doi.org/10.1016/S0277-9536(02)00548-8

de Jong, J., Groenewegen, P., & Westert, G. (2018). Medical practice variation: does it cluster within general practitioners' practices? In *Morbidity, Performance and Quality in Primary Care* (pp. 125–134). CRC Press.

Donabedian, A. (1988). The quality of care. How can it be assessed? *JAMA: The Journal of the American Medical Association*, 260(12), 1743–1748. Retrieved from https://doi.org/10.1001/jama.260.12.1743

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science & Medicine*, 46(1), 97–117. Retrieved from https://doi.org/10.1016/S0277-9536(97)00148-2

Eijkenaar, F. (2012). Pay for performance in health care: an international overview of initiatives. *Medical Care Research and Review : MCRR*. Retrieved from https://doi.org/10.1177/1077558711432891

Emanuel, E. J., Ubel, P. A., Kessler, J. B., Meyer, G., Muller, R. W., Navathe, A. S., … Volpp, K. G. (2016). Using behavioral economics to design physician incentives that deliver high-value care. *Annals of Internal Medicine*, 164(2), 114–119. Retrieved from https://doi.org/10.7326/M15-1330

ERS. (2015). *Estudo sobre o Desempenho das Unidades Locais de Saúde*.

Fattore, G., Frosini, F., Salvatore, D., & Tozzi, V. (2009). Social network analysis in primary care: The impact of interactions on prescribing behaviour. *Health Policy*, 92(2–3), 141–148. Retrieved from https://doi.org/10.1016/j.healthpol.2009.03.005

Folland, S., & Stano, M. (1990). Small area variations: a critical review of propositions, methods, and evidence. *Medical Care Review*, 47(4), 419–465.

Forsetlund, L., O'Brien, M. A., Forsen, L., Mwai, L., Reinar, L. M., Okwen, M. P., … Rose, C. J. (2021). Continuing education meetings and workshops: effects on professional practice and healthcare outcomes.

Cochrane Database of Systematic Reviews, (9).

Fung, V., Schmittdiel, J. A., Fireman, B., Meer, A., Thomas, S., Smider, N., … Selby, J. V. (2010). Meaningful Variation in Performance. *Medical Care*, 48(2), 140–148. Retrieved from https://doi.org/10.1097/MLR.0b013e3181bd4dc3

Fung, V., Schmittdiel, J. A., Fireman, B., Meer, A., Thomas, S., Smider, N., … Selby, J. V. (2010). Meaningful Variation in Performance A Systematic Literature Review, 48(2), 140–148.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. *Cambridge university press*. Retrieved from https://doi.org/10.2277/0521867061

Goldstein, H. (2011). *Multilevel statistical models*. John Wiley & Sons.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, 1(4), 223–231.

Gosling, J., Mays, N., Erens, B., & Reid, D. (2019). Quality improvement in general practice. *The Health Foundation*.

Granja, M., Ponte, C., & Cavadas, L. F. (2014). Whatc keeps family physicians busy in Portugal? A multicentre observational study of work other than direct patient contacts. *BMJ*, 4(6). Retrieved from https://doi.org/10.1136/bmjopen-2014-005026

Greene, A., Pagliari, C., Cunningham, S., Donnan, P., Evans, J., Emslie-Smith, A., … Guthrie, B. (2009). Do managed clinical networks improve quality of diabetes care? Evidence from a retrospective mixed methods evaluation. *BMJ Quality & Safety*, 18(6), 456–461.

Griffiths, P., Maben, J., & Murrells, T. (2011). Organisational quality, nurse staffing and the quality of chronic disease management in primary care: Observational study using routinely collected data. *International Journal of Nursing Studies*, 48(10), 1199–1210. Retrieved from https://doi.org/10.1016/j.ijnurstu.2011.03.011

Grytten, J., & Sørensen, R. (2003). Practice variation and physician-specific effects. *Journal of Health Economics*, 22(3), 403–418. Retrieved from https://doi.org/https://doi.org/10.1016/S0167-6296(02)00105-4

Hansen, L. J., & Drivsholm, T. B. (2002). Interventions to improve the management of diabetes mellitus in primary health care and outpatient community settings. *Ugeskrift for Laeger*, 164(5), 607–609.

Harrison, R., Manias, E., Mears, S., Heslop, D., Hinchcliff, R., & Hay, L. (2019). Addressing unwarranted clinical variation: A rapid review of current evidence. *Journal of Evaluation in Clinical Practice*, 25(1), 53–65. Retrieved from https://doi.org/10.1111/jep.12930

Hefford, M., Crampton, P., & Foley, J. (2005). Reducing health disparities through primary care reform: the New Zealand experiment. *Health Policy*, 72(1), 9–23.

Holmstrom, B., & Milgrom, P. (1994). The firm as an incentive system. *The American Economic*, 84(4), 972–991. Retrieved from https://doi.org/10.2307/2118041

Hox, J. J., Moerbeek, M., Moerbeek, M., & Schoot, R. van de. (2018). Multilevel analysis: Quantitative methodology series.

Irving, G., Neves, A. L., Dambha-Miller, H., Oishi, A., Tagashira, H., Verho, A., & Holden, J. (2017). International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open*, 7(10), e017902.

Jones, P., Shepherd, M., Wells, S., Le Fevre, J., & Ameratunga, S. (2014). Review article: What makes a good healthcare quality indicator? A systematic review and validation study. *EMA - Emergency Medicine Australasia*, 26(2), 113–124. Retrieved from https://doi.org/10.1111/1742-6723.12195

Krein, S. L., Hofer, T. P., Kerr, E. A., & Hayward, R. A. (2002). Whom Should We Profile? Examining Diabetes Care Practice Variation among Primary Care Providers, Provider Groups, and Health Care Facilities. *Health Services Research*, 37(5), 1159–1180. Retrieved from https://doi.org/10.1111/1475-6773.01102

Kringos, Boerma, W. G., Hutchinson, A., van der Zee, J., & Groenewegen, P. P. (2010). The breadth of primary care: a systematic literature review of its core dimensions. *BMC Health Services Research*, 10(1), 65. Retrieved from https://doi.org/10.1186/1472-6963-10-65

Kringos, D., Boerma, W., Bourgueil, Y., Cartier, T., Dedeu, T., Hasvold, T., … Groenewegen, P. (2013). The strength of primary care in Europe: An international comparative study. *British Journal of General Practice*,

63(616), 742–750. Retrieved from https://doi.org/10.3399/bjgp13X674422

Kringos, Dionne S, Boerma, W. G. W., Hutchinson, A., Saltman, R. B., & Organization, W. H. (2015). *Building primary care in a changing Europe: case studies*. World Health Organization. Regional Office for Europe.

Kringos, Dionne Sofia, Boerma, W. G., Hutchinson, A., & Saltman, R. B. (2015). *Building primary care in a changing Europe*. (Dionne S Kringos, W.G.W. Boerma, A. Hutchinson, & R.B. Saltman,Eds.), *European Observatory on Health Systems and Policies*.

Lawrence, M., & Olesen, F. (1997). Indicators of quality in health care. *The European Journal of General Practice*, 3(3), 103–108.

Lay-Yee, R., Scott, A., & Davis, P. (2013). Patterns of family doctor decision making in practice context. What are the implications for medical practice variation and social disparities? *Social Science & Medicine*, 76, 47–56. Retrieved from https://doi.org/10.1016/j.socscimed.2012.10.003

Leone, C., Dussault, G., & Lapão, L. V. (2014). Primary health care reform and implications for the organizational culture of Health Center Groups in Portugal. *Cadernos de Saúde Pública*, 30(1), 149–160. Retrieved from https://doi.org/10.1590/0102-311X00135112

Lester, H. E., Hannon, K. L., & Campbell, S. M. (2011). Identifying unintended consequences of quality indicators: A qualitative study. *BMJ Quality and Safety*, 20(12), 1057–1061. Retrieved from https://doi.org/10.1136/bmjqs.2010.048371

Lopes, S., Fernandes, Ó. B., Marques, A. P., Moita, B., Sarmento, J., & Santana, R. (2017). Can Vertical Integration Reduce Hospital Readmissions? A Difference-in-Differences Approach. *Medical Care*, 55(5), 506–513. Retrieved from https://doi.org/10.1097/MLR.0000000000000704

Marshall, M., & Roland, M. (2017). The future of the Quality and Outcomes Framework in England. *BMJ*, 359, j4681. Retrieved from https://doi.org/10.1136/bmj.j4681

McGuire, T. G. (2000). Chapter 9 Physician agency (pp. 461–536). Retrieved from https://doi.org/10.1016/S1574-0064(00)80168-7

Mercuri, M., & Gafni, A. (2011). Medical practice variations: what the literature tells us (or does not) about what are warranted and unwarranted variations. *Journal of Evaluation in Clinical Practice*, 17(4), 671–677. Retrieved from https://doi.org/10.1111/j.1365-2753.2011.01689.x

Miguel, L. S. (2010). Modelos e gestão nos cuidados de saúde primários. In *30 Anos do Serviço Nacional de Saúde - Um Percurso Comentado* (Almedina). Coimbra.

Ministério da Saúde. (2013). *Relatório Anual sobre o Acesso a Cuidados de Saúde no SNS*.

Ministério da Saúde. (2016). *Organograma SNS*.

Ministério da Saúde. (2017). Bilhete de Identidade dos Indicadores dos Cuidados de Saúde Primários para o ano de 2017. Administração Central do Sistema de Saúde (ACSS).

Monteiro, B. R. (2020). Monitoring and performance indicators in family health units and the objectives of sustainable development goals (SDG 3) in health: A comparative analysis in Portugal in the 2013-2018 period. *Ciencia e Saude Coletiva*, 25(4), 1221–1232. Retrieved from https://doi.org/10.1590/1413-81232020254.31422019

Mullan, F. (2004). Wrestling With Variation: An Interview With Jack Wennberg. *Health Affairs*, 23(Suppl2), VAR-73-VAR-80. Retrieved from https://doi.org/10.1377/hlthaff.var.73

Nielen, M. M. J., Schellevis, F. G., & Verheij, R. A. (2009). Inter-practice variation in diagnosing hypertension and diabetes mellitus: a cross-sectional study in general practice. *BMC Family Practice*, 10(1), 6. Retrieved from https://doi.org/10.1186/1471-2296-10-6

Nunes, A. M., & Ferreira, D. C. (2019). Reforms in the Portuguese health care sector: Challenges and proposals. *The International Journal of Health Planning and Management*, 34(1). Retrieved from https://doi.org/10.1002/hpm.2695

O'Connor, P. J., Rush, W. A., Davidson, G., Louis, T. A., Solberg, L. I., Crain, A. L., … Whitebird, R. R. (2008). Variation in Quality of Diabetes Care at the Levels of Patient, Physician, and Clinic. *Prev Chronic Dis*.

O'Donnell, C. A. (2000). Variation in GP referral rates: what can we learn from the literature? *Family Practice*, 17(6), 462–471.

OECD/EU. (2020). *Health at a Glance: Europe 2020*. OECD. Retrieved from https://doi.org/10.1787/82129230-en

OECD. (2014). *Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance?* (OECD Healt). Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/9789264216594-en

OECD. (2015). *OECD Reviews of Health Care Quality: Portugal 2015: Raising Standards*. *OECD Publishing*. OECD. Retrieved from https://doi.org/10.1787/9789264225985-en

OECD. (2021). *Health at a Glance 2021* (Health at a Glance). Retrieved from OECD: 10.1787/ae3016b9-en

Parchman, M. L. (1995). Small area variation analysis: a tool for primary care research. *Family Medicine*, 27(4), 272–6.

Pearce, C., Shearer, M., Gardner, K., Kelly, J., & Xu, T. B. (2012). GP Networks as enablers of quality of care: implementing a practice engagement framework in a General Practice Network. *Australian Journal of Primary Health*, 18(2), 101. Retrieved from https://doi.org/10.1071/PY11121

Pereira, A., Biscaia, A., Calado, I., Freitas, A., Costa, A., & Coelho, A. (2022). Healthcare Equity and Commissioning: A Four-Year National Analysis of Portuguese Primary Healthcare Units. *International Journal of Environmental Research and Public Health*, 19(22), 14819. Retrieved from https://doi.org/10.3390/ijerph192214819

Perelman, J., Lourenço, A., Russo, G., & Miraldo, M. (2017). *Pagamento pelo Desempenho nos Cuidados de Saúde Primários*. *Edições Almedina* (Olhares pa).

Pestana, J., Dimitrovová, K., Charondière, P., Broeiro, P., & Cardoso, S. (2019). Necessidade de um novo impulso na reforma dos Cuidados de Saúde Primários. In F. Mendes, F. Duarte-Ramos, H. Barros, P. L. Ferreira, & R. S. Gaspar (Eds.), *Saúde um Direito Humano. Relatório de Primavera 2019* (OPSS, p. 190). Lisbon: OPSS.

Phillips, C. B., Pearce, C. M., Hall, S., Travaglia, J., Lusignan, S., Love, T., & Kljakovic, M. (2010). Can clinical governance deliver quality improvement in Australian general practice and primary care? A systematic review of the evidence. *Medical Journal of Australia*, 193(10), 602–607. Retrieved from https://doi.org/10.5694/j.1326-5377.2010.tb04071.x

Pisco, L. (2011). Primary healthcare reform in Portugal on two fronts: autonomous family healthcare units and management of groupings of health centers. *Ciência & Saúde Coletiva*, 16(6), 2841–2852. Retrieved from https://doi.org/10.1590/S1413-81232011000600022

Poss-Doering, R., Kamradt, M., Glassen, K., Andres, E., Kaufmann-Kolle, P., & Wensing, M. (2020). Promoting rational antibiotic prescribing for non-complicated infections: understanding social influence in primary care networks in Germany. *BMC Family Practice*, 21(1), 51. Retrieved from https://doi.org/10.1186/s12875-020-01119-8

Prince, M. J., Wu, F., Guo, Y., Gutierrez Robledo, L. M., O'Donnell, M., Sullivan, R., & Yusuf, S. (2015). The burden of disease in older people and implications for health policy and practice. *The Lancet*, 385(9967), 549–562. Retrieved from https://doi.org/10.1016/S0140-6736(14)61347-7

Ramalho, A., Castro, P., Gonçalves-Pinho, M., Teixeira, J., Santos, J. V., Viana, J., … Freitas, A. (2019). Primary health care quality indicators: An umbrella review. *PLoS ONE*, 14(8), 1–20. Retrieved from https://doi.org/10.1371/journal.pone.0220888

Raman, R., & Hedeker, D. (2005). A mixed-effects regression model for three-level ordinal response data. *Statistics in Medicine*, 24(21), 3331–3345. Retrieved from https://doi.org/10.1002/sim.2186

Rice, N., & Jones, A. (1997). Multilevel models and health economics. *Health Economics*, 6(6), 561–575. Retrieved from https://doi.org/10.1002/(sici)1099-1050(199711)6:6<561::aid-hec288>3.3.co;2-o

Rodriguez, H. P., Scoggins, J. F., von Glahn, T., Zaslavsky, A. M., & Safran, D. G. (2009). Attributing Sources of Variation in Patients' Experiences of Ambulatory Care. *Medical Care*, 47(8), 835–841. Retrieved from https://doi.org/10.1097/MLR.0b013e318197b1e1

Rosenthal, M. B., & Frank, R. G. (2006). What is the empirical basis for paying for quality in health care? *Medical Care Research and Review*. Sage PublicationsSage CA: Thousand Oaks, CA. Retrieved from https://doi.org/10.1177/1077558705285291

Sá, L., Costa-Santos, C., Teixeira, A., Couto, L., Pereira, A. C., Hespanhol, A., … Martins, C. (2015). Portuguese

family physicians' awareness of diagnostic and laboratory test costs: A cross-sectional study. *PLoS ONE*, 10(9), 1–14. Retrieved from https://doi.org/10.1371/journal.pone.0137025

Salisbury, C., Wallace, M., & Montgomery, A. A. (2010). Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling. *BMJ*, 341(oct12 1), c5004–c5004. Retrieved from https://doi.org/10.1136/bmj.c5004

Saltman, R., Bankauskaite, V., & Vrangbaek, K. (2006). *Primary care in the driver's seat? Organizational reform in European primary care*. *European Observatory on Health Systems and Policies* (Vol. 51). McGraw-Hill Education (UK).

Santana, R. & Costa, C. (2008). A integração vertical de cuidados de saúde: aspectos conceptuais e organizacionais. *Revista Portuguesa de Saúde Pública*, Vol. Temát, 29–56.

Scott, A., & Coote, W. (2007). Whither Divisions of General Practice? An empirical and policy analysis of the impact of Divisions within the Australian health care system. *Medical Journal of Australia*, 187(2), 95–99. Retrieved from https://doi.org/10.5694/j.1326-5377.2007.tb01149.x

Scott, A., & Coote, W. (2009). Do regional primary-care organisations influence primary-care performance? A dynamic panel estimation. *Health Economics*, 362(October 2008), n/a-n/a. Retrieved from https://doi.org/10.1002/hec.1509

Scott, A., Sivey, P., Ait Ouakrim, D., Willenberg, L., Naccarella, L., Furler, J., … Young, D. (2011). The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database of Systematic Reviews*, 9(9), CD008451. Retrieved from https://doi.org/10.1002/14651858.CD008451.pub2

Selby, J. V., Schmittdiel, J. A., Lee, J., Fung, V., Thomas, S., Smider, N., … Fireman, B. (2010). Meaningful variation in performance: What does variation in quality tell us about improving quality? *Medical Care*, 48(2), 133–139. Retrieved from https://doi.org/10.1097/MLR.0b013e3181c15a6e

Shashar, S., Ellen, M., Codish, S., Davidson, E., & Novack, V. (2021). Medical Practice Variation Among Primary Care Physicians: 1 Decade, 14 Health Services, and 3,238,498 Patient-Years. *The Annals of Family Medicine*, 19(1), 30 LP – 37. Retrieved from https://doi.org/10.1370/afm.2627

Simões, J. de A., Augusto, G. F. G. F., Fronteira, I. I., Hernandez-Quevedo, C., Simões, J. de A., Augusto, G. F. G. F., … Hernandez-Quevedo, C. (2017). Portugal: Health System Review. *Health Systems in Transition*, 19(2), 1–184.

Sirimsi, M. M., De Loof, H., Van den Broeck, K., De Vliegher, K., Pype, P., Remmen, R., & Van Bogaert, P. (2022). Scoping review to identify strategies and interventions improving interprofessional collaboration and integration in primary care. *BMJ Open*, 12(10), e062111. Retrieved from https://doi.org/10.1136/bmjopen-2022-062111

Skinner, J. (2011). Causes and Consequences of Regional Variations in Health Care. In *Handbook of Health Economics (Vol. 2)*. (pp. 45–93). Retrieved from https://doi.org/10.1016/B978-0-444-53592-4.00002-5

Smith, J., & Goodwin, N. (2017). *Towards managed primary care: the role and experience of primary care organizations*. Routledge.

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18(2–3), 277–310. Retrieved from https://doi.org/10.1080/01900699508525011

SNS. (2019). BI-CSP - Portuguese Health System Control Panel. Retrieved from https://bicsp.min-saude.pt/pt/investigacao/Paginas/Matrizindicadorescsp publico.aspx?isdlg=1

StataCorp. (2021). Stata Statistical Software: Release 17. *College Station: StataCorp LLC*.

Sutherland, K., & Levesque, J. (2020). Unwarranted clinical variation in health care: Definitions and proposal of an analytic framework. *Journal of Evaluation in Clinical Practice*, 26(3), 687–696. Retrieved from https://doi.org/10.1111/jep.13181

van Loenen, T., van den Berg, M. J. J., Westert, G. P. P., Faber, M. J. J., Loenen, T. Van, Berg, M. J. Van Den, … Faber, M. J. J. (2014). Organizational aspects of primary care related to avoidable hospitalization: A systematic review. *Family Practice*, 31(5), 502–516. Retrieved from https://doi.org/10.1093/fampra/cmu053

Wennberg, J. E. (1985). On patient need, equity, supplier-induced demand, and the need to assess the outcome of

common medical practices. *Medical Care*, 23(5), 512–520.

Wennberg, J. E., Barnes, B. A., & Zubkoff, M. (1982). Professional uncertainty and the problem of supplier-induced demand. *Social Science & Medicine*, 16(7), 811–824.

Wennberg, J., & Gittelsohn, A. (1973). Small Area Variations in Health Care Delivery. *Science*, 182(4117), 1102–1108. Retrieved from https://doi.org/10.1126/science.182.4117.1102

Werner, R. M., & Asch, D. A. (2005). The Unintended Consequences of Publicly Reporting Quality Information. *JAMA*, 293(10), 1239. Retrieved from https://doi.org/10.1001/jama.293.10.1239

Westert, G. P., de Jong, J. D., & da Silva, P. (2015). Medical Practice Variations in Primary Care. In A. Johnson & T. Stukel (Eds.), *Medical Practice Variations* (pp. 1–11). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-1-4899-7573-7_74-1

Westert, G. P., & Groenewegen, P. P. (1999). Medical practice variations: changing the theoretical approach. *Scandinavian Journal of Public Health*, 27(3), 173–180. Retrieved from https://doi.org/10.1177/14034948990270030801

WHO, & World Health Organization. (2018). Health System Review Portugal. *European Observatory on Health Systems and Policies*, (April).

Yano, E. M., Soban, L. M., Parkerton, P. H., & Etzioni, D. A. (2007). Primary Care Practice Organization Influences Colorectal Cancer Screening Performance. *Health Services Research*, 42(3p1), 1130–1149. Retrieved from https://doi.org/10.1111/j.1475-6773.2006.00643.x

1.8 Appendices

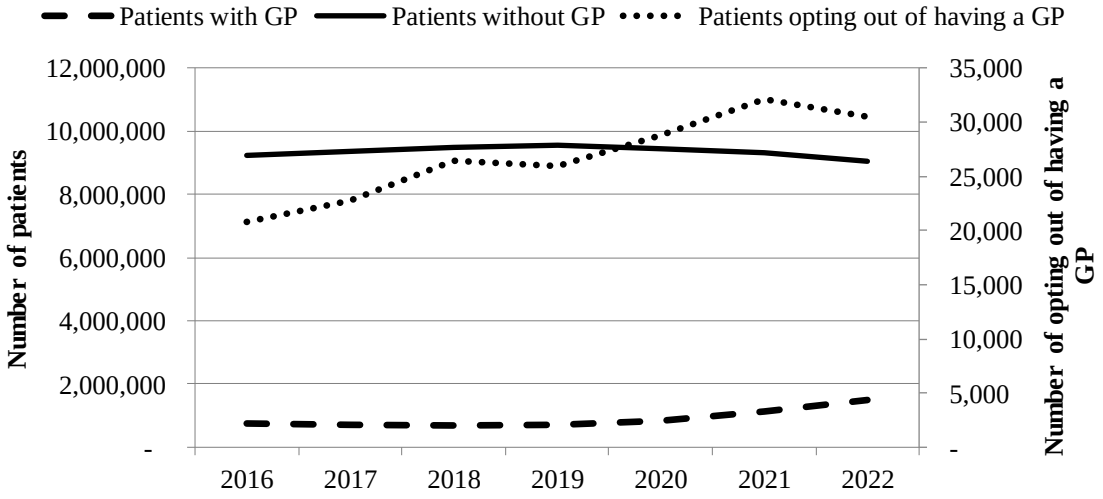
1.8.1 Evolution of the number of patients enrolled in primary care in Portugal

Note: Data from "*Transparencia SNS*" provided by Serviços Partilhados do Ministério de Saúde. Available online: https://transparencia.sns.gov.pt/explore/dataset/utentes-inscritos-em-cuidados-de-saude-primarios (accessed on 1 Mar 2023)

Year	Patients enrolled in Primary care	Patients without GP	Percentage of the patients without GP	Min per ACES	Max per ACES
2016	10,010,709	767,147	7.66%	0.00%	31.53%
2017	10,090,681	710,329	7.04%	0.01%	28.25%
2018	10,195,079	690,232	6.77%	0.00%	27.42%
2019	10,307,383	730,232	7.08%	0.00%	28.62%
2020	10,313,708	835,658	8.10%	0.10%	27.45%
2021	10,477,387	1,139,340	10.87%	0.02%	30.51%
2022	10,568,683	1,494,392	14.14%	0.00%	39.15%

Table 1.2: Evolution of the number of patients enrolled in Primary care in Portugal and the percentage of patients in a GP list.

Figure 1.5: Evolution of the number of patients enrolled in Primary care in Portugal: with GP, without GP and opting out of GP.



Describes the number of patients with GP, without a GP and without a GP per option that represents those who have voluntarily opted not to have a GP.

Data from Table 1.2 and Figure 1.5, that describes the number of patients with GP, without a GP and without a GP per option that represents those who have voluntarily opted not to have a GP, reveals that the number of

patients with a GP has remained relatively stable or even increased slightly, suggesting a consistent availability of GP services. Secondly, the number of patients choosing to opt out of having a GP has steadily increased over the years, growing from 20,784 in 2016 to 30,452 in 2022. This can indicate a consistent preference among a portion of the population to forego primary care that might have several causes, such as the increasing presence private supply primary care services, and the increasing mobility of the population.

Lastly, and most alarmingly, the number of patients without a GP has risen substantially from 767,147 in 2016 to 1,494,392 in 2022. This points to a potential mal distribution issue and may give rise to access issue, indicating that while GP services are available, an increasing portion of the population in some areas may be struggling to secure these essential healthcare providers.

1.8.2 Diagram of the Portuguese National Health Service

An overview of the hierarchical organization of the health system as of 2018 is presented, drawing insights from the organogram by the *Ministério da Saúde* (2016) and the report by Barros, Machado, and Simões (2011).

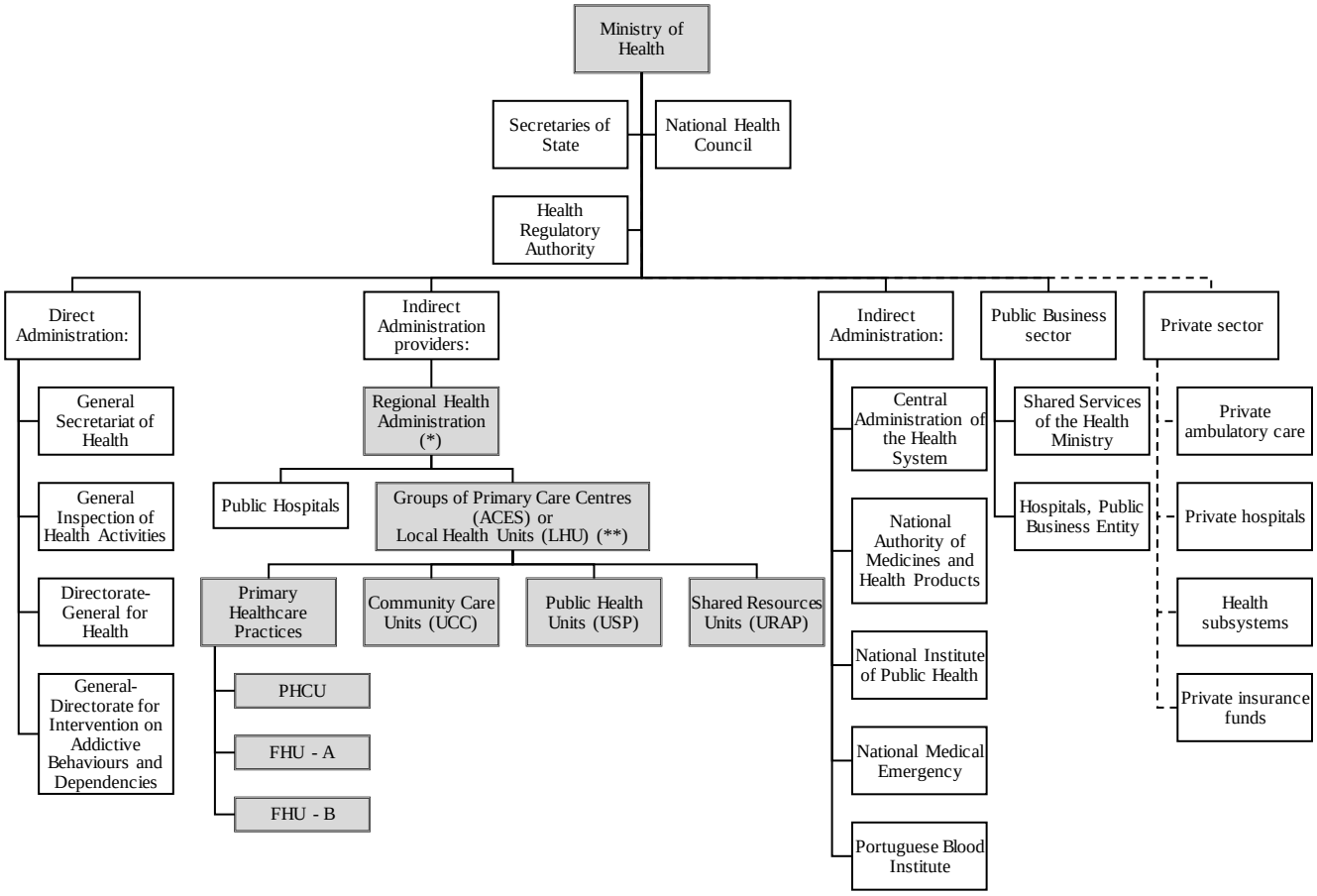


Figure 1.6: Diagram of the health system as of 2018

Portugal's health care delivery, a blend of public and private providers. The Ministry of health oversees healthcare coordination of all providers and the financing of public health care delivery. The administrative structure of Portuguese healthcare comprises both direct and indirect administrations of the public health care sector, the public business sector, and the private sector, illustrated in Figure 1.6. Most citizens can choose between the National Health Service (NHS) or private health insurance/subsystems, or a combination of both.

Elements highlighted in grey within Figure 1.6. represent the primary care structures under study in this article. There are different financing agreements and methods across entities. Specifically, the entities emphasized in the figure derive their funding through (*) historical and activity-based budgets, and (**) prospective payments determined via annually negotiated contracts (Barros et al., 2011).

In 2022, a significant reorganization of the health system was undertaken with the inception of the Executive Directorate of the National Health Service. The Executive Directorate of the National Health Service was established per Decree-Law in 2022 and the statutes approved in 2023[11]. This Directorate's mandate is to oversee and coordinate the care provided by health units within the SNS and manage national networks related to continuous and palliative care. Notwithstanding this institutional alteration, the operational configuration of the primary care providers examined in this study remains unchanged.

[11] Respectively: Decree-Law No. 61/2022, of September 23 – Organization of the Executive Directorate of the National Health Service; and Ordinance No. 306-A/2023, of October 12 Approves the Statutes of the Executive Directorate of the National Health Service.

1.8.3 Diagram of the study population

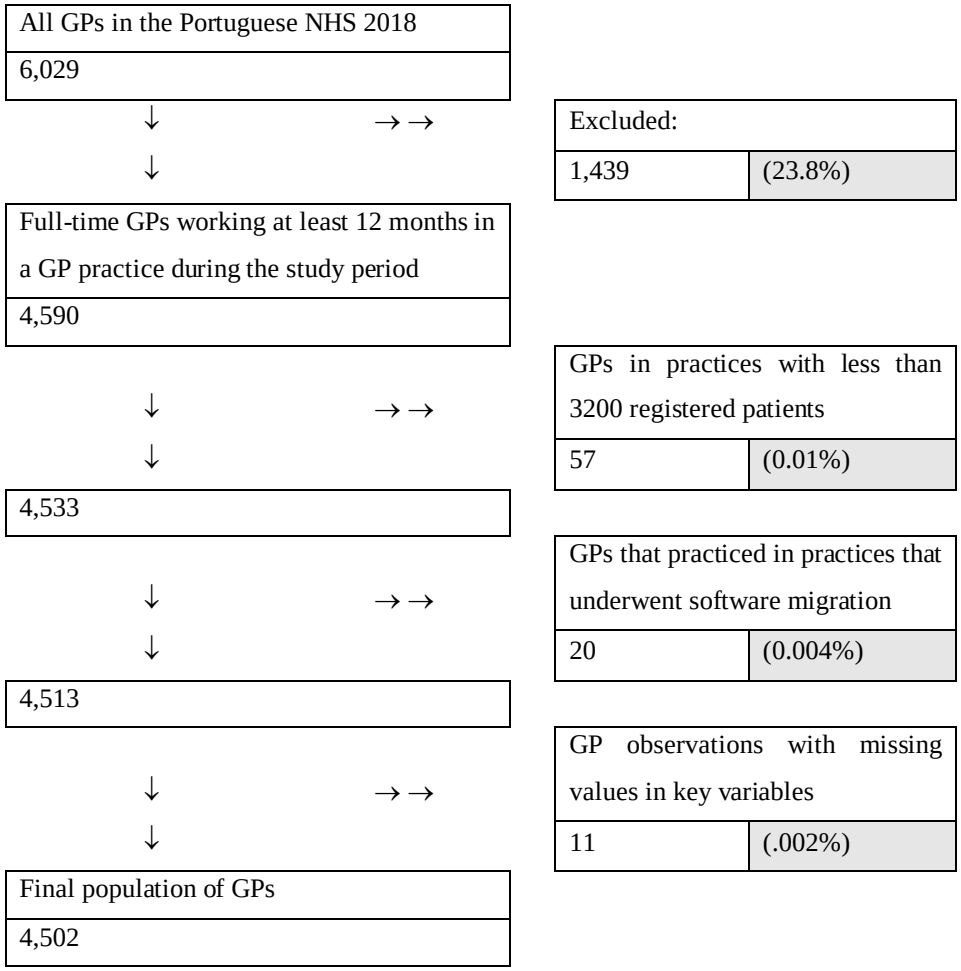


Figure 1.7: Diagram of the study population

1.8.4 Description of the indicators

Table 1.3 presents a comprehensive list of all the indicators categorized into four key domains: Access, Efficiency, Process, and Health Outcomes. It also includes information on whether these indicators could be utilized in the contractual process between the practice and the ACES in 2018, along with the respective target intervals (acceptable and expected).

Table 1.3: List of indicators and corresponding target intervals in 2018.

Type	Indicator	Used for incentives	Min acceptable	Min expected	Max expected	Max acceptable
Access	001 - Prop. of medical consultations by the respective family doctor	Y	75	78	88	90
	003 - Rate of medical home visits per 1000 enrolled	Y	12	18	35	40
	010 - Utilization rate of family planning medical consultations	N	35	45	55	65
	330 - Usage index scaled to the estimated annual need for medical appointments	Y	0.8	0.85	2	2
	335 - Prop. of non-face-to-face consultations with prescription in the first 72 hours after the respective request.	Y	80	85	100	100
	342 - Prop. of medical appointments required by the patients scheduled in less than 15 working days	Y	60	65	100	100
	344 - Prop. of medical consultations on the day of the appointment.	Y	15	20	35	45
Efficiency	278 - Prop. of packages of prescription medications of therapeutic classes with generic	N	50	60	80	90
	350 - Cost of therapy for patients with Diabetes Mellitus	Y	120	120	300	320
	351 - Cost of therapy for patients with controlled Diabetes Mellitus	Y	120	120	300	320
	352 - Cost of therapy for patients with Arterial Hypertension	Y	50	50	90	95
	353 - Cost of therapy for patients with controlled Arterial Hypertension	Y	50	50	100	105
Process	011 - Prop. of pregnant women with 1st pregnancy surveillance medical consultation, held in the 1st trimester	Y	70	75	100	100
	014 - Prop. of newborn with at least 1 medical consultation for child health surveillance within the first 28 days of life	Y	85	95	100	100
	016 - Prop. of children with at least 6 medical consultations for child health surveillance in the first year of life	N	65	70	100	100
	017 - Prop. of children with at least 3 medical appointments for child health surveillance in the 2nd year of life	N	65	70	100	100
	023 - Prop. of patients with arterial hypertension, with determination of cardiovascular risk in the last 3 years	N	70	80	100	100
	030 - Prop. of patients with diabetes or chronic respiratory disease or with chronic heart disease or older than 65 years, with the flu vaccine prescribed or administered in the last 12 months.	Y	50	55	100	100
	031 - Prop. of children aged 7, with weight and height registered in the interval between 5 and 7 years	N	85	90	100	100
	032 - Prop. of young people aged 14, with weight and height registered in the interval between 11 and 14 years	N	85	90	100	100
	038 - Prop. of DM patients with 1 HgbA1c per semester	N	70	75	100	100
	045 - Prop. of women among 25 and 60 years old, with cervical cancer screening	Y	47	52	100	100
	046 - Prop. of patients aged between 50 and 75 years, with colon and rectum cancer screening	Y	47	52	100	100
	047 - Prop. of patients with aged 14y or older, with smoking habits (in the last 3 years)	N	60	66	100	100
	049 - Prop. of COPD patients with at least one FeV1 assessment record in the past 3 years	Y	40	60	100	100

	053 - Prop. of patients aged 14 years old or older, with alcohol consumption, recorded in the last 3 years (53)	N	60	66	100	100
	054 - Prop. of patients aged 14 years old or older and with the problem of "excessive alcohol consumption", who received at least one related consultation in the last 3 years	N	55	70	100	100
	057 - Prop. of newborns with early diagnosis (TSH PKU) performed up to sixth day of life	N	90	95	100	100
	059 - Prop. of children aged 2, with weight and height registered in the last year	N	90	95	100	100
	093 - Prop. of children aged 2 years, with National Vaccination Plan completed or in progress at the reference date of the indicator	Y	95	98	100	100
	094 - Prop. of children aged 7 years, with National Vaccination Plan completed or in progress at the reference date of the indicator	Y	95	98	100	100
	095 - Prop. of children aged 14 years, with National Vaccination Plan completed or in progress at the reference date of the indicator	Y	95	98	100	100
	097 - Prop. of DM patients, with microalbuminuria in the last year	N	75	80	100	100
	098 - Prop. of patients aged 25 or over who have an updated tetanus vaccine.	Y	85	92	100	100
	261 - Prop. of DM patients, with risk of foot ulceration in the last year	Y	75	80	100	100
	262 - Prop. of patients with risk evaluation for type 2 DM in the last 3 years	Y	10	22	100	100
	276 - Ratio between the sum of the daily defined dose prescribed in DPP-4 inhibitors and the sum of the daily prescribed dose prescribed in oral antidiabetics in patients with type 2 Diabetes Mellitus	Y	0	0	36	40
	277 - Prop. of patients aged 14 years or older with smoking habits, whose an appointment on smoking was conducted in the last year	N	18	25	100	100
	297 - Prop. of patients aged 65y or older, without prolonged prescription of anxiolytics, sedatives, or hypnotics	Y	77	80	100	100
	307 - Prop. of pregnant women who had at least one ultrasound examination during the first trimester of pregnancy	Y	70	75	100	100
	308 - Prop. of pregnant women who had at least one ultrasound examination during the second trimester of pregnancy	N	70	75	100	100
	309 - Prop. of pregnant women who had at least one ultrasound examination during the third trimester of pregnancy	N	40	45	100	100
	310 - Laboratory test index in the first trimester of pregnancy	N	0.7	0.75	1	1
	311 - Laboratory test index in the second trimester of pregnancy	N	0.7	0.75	1	1
	312 - Laboratory test index in the third trimester of pregnancy	N	0.7	0.75	1	1
Health Outcomes	018 - Prop. of patients with arterial hypertension, with at least one BMI record in the last 12 months	N	70	80	100	100
	019 - Prop. of patients with arterial hypertension, with blood pressure record in each semester	N	65	75	100	100
	020 - Prop. of patients with arterial hypertension, <65y, with blood pressure below 150/90 mmHg	Y	50	67	100	100
	039 - Prop. of DM patients with the last HgbA1c <= 8.0%	Y	50	60	100	100
	091 - Prop. of DM patients <65y, with HgbA1c<= 6.5%	N	25	30	55	70

Legend: GP (General Practitioner), DM (Diabetes Mellitus), HT (Hypertension), COPD (Chronic obstructive pulmonary disease), W (Woman), FeV (Forced Expiratory Volume), TSHPKU (Thyroid Stimulating Hormone and Phenylketonuria or Heel Prick or Guthrie test), sem (semester).

1.8.5 Summary statistics from outcome variables

Table 1.4 presents summary statistics of the dependent variables in absolute terms (including the mean, standard deviation, median, maximum, and minimum values) and in relative terms, i.e., the achievement based on benchmark intervals (including the percentages of physicians who achieved the acceptable interval and the expected interval, while the remainder achieved neither the acceptable nor expected intervals).

Table 1.4: Descriptive statistics of the outcome variables in 2018.

Type	Indicator	Mean (SD)	Median [Min,Max]	Acceptable	Expected
Access	001 - Prop. of medical consultations by the respective family doctor	85.9 (10.9)	88.3 [1.4, 100]	15%	34%
	003 - Rate of medical home visits per 1000 enrolled	0.017 (0.016)	0.015 [0, 0.4]	0%	0%
	010 - Utilization rate of family planning medical consultations	36.4 (17.1)	36.7 [0, 97.5]	31%	19%
	330 - Usage index scaled to the estimated annual need for medical appointments	0.827 (0.126)	0.833 [0.3, 3.9]	19%	44%
	335 - Prop. of non-face-to-face consultations with prescription in the first 72 hours after the respective request.	95.1 (8.8)	98.5 [4.8, 100]	3%	91%
	342 - Prop. of medical appointments required by the patients scheduled in less than 15 working days	88.1 (12.1)	91.8 [14, 100]	2%	95%
	344 - Prop. of medical consultations on the day of the appointment.	39.4 (11.1)	38.5 [5.8, 100]	39%	35%
Efficiency	278 - Prop. of packages of prescription medications of therapeutic classes with generic	62.7 (6.6)	63.5 [21.7, 81.3]	21%	74%
	350 - Cost of therapy for patients with Diabetes Mellitus	341.5 (96.2)	333.5[60, 1192]	9%	34%
	351 - Cost of therapy for patients with controlled Diabetes Mellitus	345.2 (118.7)	329.4 [18.2,1847.9]	8%	37%
	352 - Cost of therapy for patients with Arterial Hypertension	79.7(18.4)	77.581 [24.3,202]	8%	71%
	353 - Cost of therapy for patients with controlled Arterial Hypertension	91.8 (25.1)	88 [15.0,757]	7%	71%
Process	011 - Prop. of pregnant women with 1st pregnancy surveillance medical consultation, held in the 1st trimester	90.7 (11.6)	93.3 [0, 100]	2%	92%
	014 - Prop. of newborn with at least 1 medical consultation for child health surveillance within the first 28 days of life	60.3 (24.7)	63.5 [0, 100]	26%	46%
	016 - Prop. of children with at least 6 medical consultations for child health surveillance in the first year of life	71.3 (27.6)	80 [0, 100]	6%	61%
	017 - Prop. of children with at least 3 medical appointments for child health surveillance in the 2nd year of life	67.8 (28.0)	75 [0, 100]	5%	56%
	023 - Prop. of patients with arterial hypertension, with determination of cardiovascular risk in the last 3 years	60.9 (0.32.2)	71.6 [0, 100]	13%	40%
	030 - Prop. of patients with diabetes or chronic respiratory disease or with chronic heart disease or older than 65 years, with the flu vaccine prescribed or administered in the last 12 months.	45.7 (10.7)	45.2 [0, 100]	14%	21%
	031 - Prop. of children aged 7, with weight and height registered in the interval between 5 and 7 years	77.1 (23.0)	84.0 [0, 100]	10%	39%
	032 - Prop. of young people aged 14, with weight and height registered in the interval between 11 and 14 years	63.6 (25.5)	66.7 [0, 100]	8%	17%
	038 - Prop. of DM patients with 1 HgbA1c per semester	63.4 (24.4)	69.8 [0, 100]	9%	42%
	045 - Prop. of women among 25 and 60 years old, with cervical cancer screening	49.5 (19.6)	52.5 [0, 100]	9%	52%

	046 - Prop. of patients aged between 50 and 75 years, with colon and rectum cancer screening	53.6 (20.8)	57.6 [0, 100]	7%	61%
	047 - Prop. of patients with aged 14y or older, with smoking habits (in the last 3 years)	61.2 (22.7)	68.2 [0, 100]	9%	55%
	049 - Prop. of COPD patients with at least one FeV1 assessment record in the past 3 years	44.9 (30.9)	44.0 [0, 100]	18%	37%
	053 - Prop. of patients aged 14 years old or older, with alcohol consumption, recorded in the last 3 years (53)	57.9 (24.1)	65.2 [0, 100]	10%	49%
	054 - Prop. of patients aged 14 years old or older and with the problem of "excessive alcohol consumption", who received at least one related consultation in the last 3 years	51.2 (23.2)	50.0 [0, 100]	22%	22%
	057 - Prop. of newborns with early diagnosis (TSH PKU) performed up to sixth day of life	85.6 (18.4)	90.9 [0, 100]	16%	37%
	059 - Prop. of children aged 2, with weight and height registered in the last year	83.6 (20.9)	91.7 [0, 100]	13%	40%
	093 - Prop. of children aged 2 years, with National Vaccination Plan completed or in progress at the reference date of the indicator	96.9 (6.5)	100 [0, 100]	3%	71%
	094 - Prop. of children aged 7 years, with National Vaccination Plan completed or in progress at the reference date of the indicator	96.1 (6.7)	100 [0, 100]	5%	64%
	095 - Prop. of children aged 14 years, with National Vaccination Plan completed or in progress at the reference date of the indicator	96.9 (6.0)	100 [0, 100]	6%	67%
	097 - Prop. of DM patients, with microalbuminuria in the last year	66.1 (24.9)	73.0 [0, 100]	10%	37%
	098 - Prop. of patients aged 25 or over who have an updated tetanus vaccine.	86.9 (11.4)	90.7 [0, 100]	24%	45%
	261 - Prop. of DM patients, with risk of foot ulceration in the last year	68.2 (31.5)	81.8 [0, 100]	6%	54%
	262 - Prop. of patients with risk evaluation for type 2 DM in the last 3 years	34.9 (27.1)	34.2 [0, 100]	10%	62%
	276 - Ratio between the sum of the daily defined dose prescribed in DPP-4 inhibitors and the sum of the daily prescribed dose prescribed in oral antidiabetics in patients with type 2 Diabetes Mellitus	34.8 (10.1)	34.8 [0, 85.9]	15%	55%
	277 - Prop. of patients aged 14 years or older with smoking habits, whose an appointment on smoking was conducted in the last year	22.8 (15.6)	20.2 [0, 100]	14%	40%
	297 - Prop. of patients aged 65y or older, without prolonged prescription of anxiolytics, sedatives, or hypnotics	77.8 (6.9)	77.9 [0, 100]	18%	37%
	307 - Prop. of pregnant women who had at least one ultrasound examination during the first trimester of pregnancy	61.3 (28.2)	68.4 [0, 100]	5%	42%
	308 - Prop. of pregnant women who had at least one ultrasound examination during the second trimester of pregnancy	61.6 (30.0)	69.2 [0, 100]	5%	44%
	309 - Prop. of pregnant women who had at least one ultrasound examination during the third trimester of pregnancy	50.5 (28.1)	53.8 [0, 100]	7%	61%
	310 - Laboratory test index in the first trimester of pregnancy	0.635 (0.224)	0.689 [0, 1]	10%	38%
	311 - Laboratory test index in the second trimester of pregnancy	0.421 (0.200)	0.424 [0, 1]	4%	5%
	312 - Laboratory test index in the third trimester of pregnancy	0.338 (0.176)	0.333 [0, 1]	1%	1%
Health Outcomes	018 - Prop. of patients with arterial hypertension, with at least one BMI record in the last 12 months	75.2 (22.9)	82.8 [0, 100]	15%	56%
	019 - Prop. of patients with arterial hypertension, with blood pressure record in each semester	60.3 (24.7)	63.5 [0, 100]	14%	35%
	020 - Prop. of patients with arterial hypertension, <65y, with blood pressure below 150/90 mmHg	53.5 (22.6)	55.0 [0, 100]	25%	34%
	039 - Prop. of DM patients with the last HgbA1c <= 8.0%	62.6 (19.8)	67.4 [0, 100]	13%	65%
	091 - Prop. of DM patients <65y, with HgbA1c<= 6.5%	28.1 (12.9)	27.8 [0, 76.6]	17%	42%

Legend: GP (General Practitioner), DM (Diabetes Mellitus), HT (Hypertension), COPD (Chronic obstructive pulmonary disease), W (Woman), FeV (Forced Expiratory Volume), TSHPKU (Thyroid Stimulating Hormone and Phenylketonuria or Heel Prick or Guthrie test), sem (semester).

These statistics provide a comprehensive understanding of the variation and performance of the measured indicators among physicians, grouped into four key categories: Access, Efficiency, Process, and Health Outcomes.

In terms of access indicators, we can see that the mean percentage of medical consultations by the respective family doctor (001) is 85.9%, with a standard deviation of 10.9. The utilization rate of family planning medical consultations (010) has a mean of 36.4% and an SD of 17.1. The indicator "Proportion of non-face-to-face consultations with prescription in the first 72 hours after the respective request" (335) exhibits the highest mean of 95.1% and an impressive percentage of GPs achieving the expected interval (91%), implying that most physicians consistently meet or exceed the expected benchmark. On the other hand, the "Proportion of medical consultations by the respective family doctor" (001) has a lower achievement of 15% for the acceptable interval and 34% for the expected interval. Even though, on average, 85.9% of all medical consultations are pursued with the patients' family physicians, in some practices, a substantial number of visits are made with other physicians. Curiously, the benchmark sets the acceptable interval above 75% but below 90%, indicating a certain level of substitution is envisioned. However, aiming for the percentage above 75% in this indicator may lead to an improvement in patients' continuity of care and strengthen the role of family physicians in primary healthcare services.

Moving to efficiency indicators, the "Proportion of prescription medications of therapeutic classes with generic" (278) stands out with a mean of 62.7%, a SD of 6.6, and the achievement of the expected interval is notably high at 74%, highlighting an encouraging trend towards generic drug usage. Conversely, the indicator "Cost of therapy for patients with Diabetes

Mellitus" (350) has a mean of 341.5 and an SD of 96.2, showing substantial variation in spending for this condition. Moreover, it has a low achievement of 34% for the expected interval, indicating that only a small proportion of physicians meet the acceptable benchmark for the cost of therapy in diabetic patients. Similar variation is observed in cost of therapy for patients with Arterial Hypertension (352), with a mean of 79.7 and an SD of 18.4.

Process indicators display varying levels of achievement. For instance, the proportion of pregnant women with 1st pregnancy surveillance medical consultation held in the 1st trimester (011) has a remarkable mean of 90.7% and a SD of 11.6. 92% of GPs achieved the expected interval which indicates that most physicians meet the acceptable benchmark in providing early pregnancy surveillance medical consultations during the crucial first trimester. In contrast, the proportion of children with at least 6 medical consultations for child health surveillance in the first year of life (016) has a lower mean of 71.3% and a higher SD of 27.6.

Lastly, the health outcomes indicators also exhibit diversity in their performance. For example, the proportion of patients with arterial hypertension with at least one BMI record in the last 12 months (018) has a mean of 75.2% and an SD of 22.9, while the proportion of DM patients with the last HgbA1c <= 8.0% (039) has a mean of 62.6% and an SD of 19.8. The indicator "Prop. of patients with arterial hypertension, <65y, with blood pressure below 150/90 mmHg" (020) exhibits the highest mean of 53.5. However, only 34% of physicians meet the expected benchmark for managing blood pressure levels. Additionally, the indicator "Prop. of DM patients <65y, with HgbA1c<= 6.5%" (091) has the lowest mean of 28.1, indicating that, on average, only a small proportion of diabetic patients below 65 years of age maintain HgbA1c levels below 6.5%. In this case, only 41% (=100-17-42) of physicians did not meet the acceptable benchmark for managing HgbA1c levels in these patients.

1.8.6 Summary statistics from outcome variables per type of indicator and practice

Table 1.5 Summary statistics of achievement for incentivized and non-incentivized indicators per type of practice (PHCU, FHU-A, FHU-B)

Type of indicators	Number of observations	Achievement [% of GPs]	PHCU	FHU-A	FHU-B
Incentivized (n=27)	N=121554	Did not achieve the target (0)	53%	34%	22%
		Achieved accepted interval (1)	10%	13%	12%
		Achieved expected interval (2)	37%	54%	66%
Non-incentivized (n=23)	N=103546	Did not achieve the target (0)	71%	47%	30%
		Achieved accepted interval (1)	9%	13%	13%
		Achieved expected interval (2)	20%	40%	57%
Nr of physicians			1,594	1,177	1,731

Among the physicians that did not adopt the team-based practice (PHCU), 47% of GPs achieved the targets for the incentivized indicators, and for the non-incentivized indicators, only 29% did. The discrepancy between incentivized and non-incentivized indicators was smaller for team-based practices (FHUs). While 67% of physicians achieved the incentivized indicators in FHU-A, 53% did so for the non-incentivized, and 78% of the physicians in FHU-B achieved the incentivized indicator, while 70% did so for the non-incentivized indicators.

1.8.7 Summary statistics from explanatory variables

Table 1.6: Descriptive statistics of the explanatory variables in 2018.

	Mean	Std. dev.	Min	Max
Physician-level covariates				
GP Age	52.894	11.926	30.000	76.000
GP gender = female	0.640	0.480		
1st GP list size quintile [1203 to 1573[	0.200	0.400		
2nd GP list size quintile [1573 to 1717[	0.200	0.400		
3rd GP list size quintile [1717 to 1790[	0.201	0.401		
4th GP list size quintile [1790 to 1864[	0.199	0.399		
5th GP list size quintile [1864 to 2221[	0.200	0.400		
GP list % Female	0.527	0.070	0.000	1.134
GP list % 0 – 6 years	0.063	0.019	0.000	0.228
GP list % >65 years	0.226	0.066	0.000	0.586
Practice-level covariates				
Size in professionals FTE	27.238	13.101	2.000	106.528
Ratio Nurses to GPs	1.096	0.313	0.417	5.071
% Elderly in practice	22.467	4.920	8.564	46.781
Unemployment rate	5.459	1.897	2.400	12.000
City	0.450	0.497		
Towns and suburbs	0.383	0.486		
Rural areas	0.167	0.373		
PHCU	0.354	0.478		
FHU-A	0.261	0.439		
FHU-B	0.384	0.486		
Age of the practice	8.464	3.147	1.000	12.000
Distance from practice to ACES	9.819	11.536	0.001	79.627
Local group level covariates (ACES)				
Shared Physicians	0.703	0.473	0.228	3.428
Shared Nurses	2.677	1.383	0.651	7.865
Integrated Care (LHU=1)	0.124	0.330		
% Population not listed with GP	6.982	7.358	0.037	29.495
FHU-A in ACES (%)	30.616	14.093	0.000	66.667
FHU-B in ACES (%)	29.467	19.394	0.000	71.429

1.8.8 Results per indicator – Variance Partition Coefficient (VPC) at practice- and local- level (stacked) for each Indicator.

Figure 1.8: Variance Partition coefficient for each indicator in 2018.

Note: The regressions include all the confounders at the physician-, practice-, and ACES-levels, including physician demographics, patient list size, patient demographics and medical conditions, practice, and ACES characteristics.

Legend: GP (General Practitioner), DM (Diabetes Mellitus), HT (Hypertension), COPD (Chronic obstructive pulmonary disease), W (Woman), FeV (Forced Expiratory Volume), TSHPKU (Thyroid Stimulating Hormone and Phenylketonuria or Heel Prick or Guthrie test), sem (semester).

1.8.9 Results per type of practice – Variance Partition Coefficient (VPC) at practice- and local- level (stacked) for each Indicator, per type of practice.

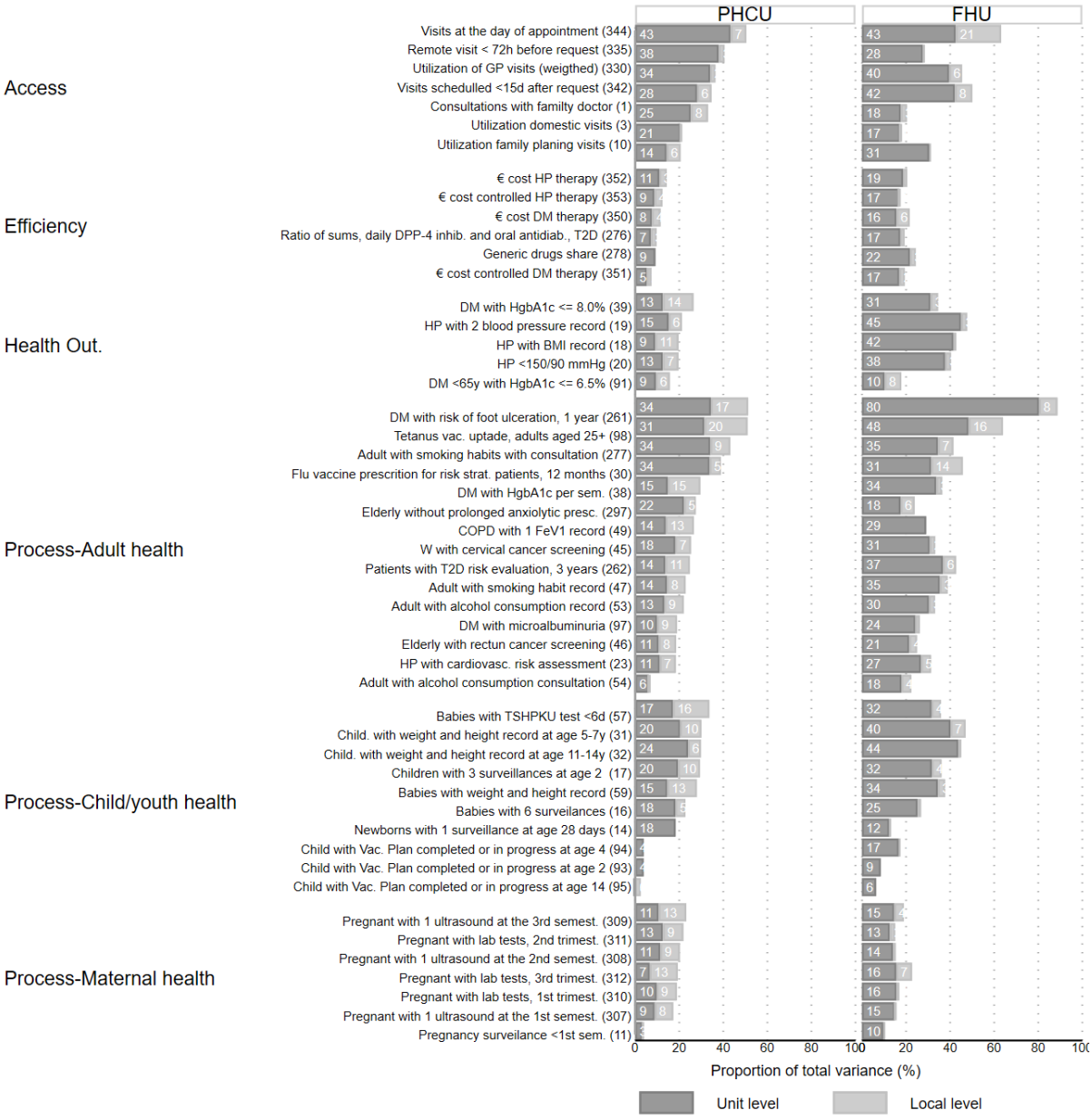


Figure 1.9: Variance Partition coefficient for each indicator in 2018 by type of practice.

Note: The regressions include all the confounders at the physician-, practice-, and ACES-levels, including physician demographics, patient list size, patient demographics and medical conditions, practice, and ACES characteristics.

Legend: GP (General Practitioner), DM (Diabetes Mellitus), HT (Hypertension), COPD (Chronic obstructive pulmonary disease), W (Woman), FeV (Forced Expiratory Volume), TSHPKU (Thyroid Stimulating Hormone and Phenylketonuria or Heel Prick or Guthrie test), sem (semester).

Chapter 2

2 Antidepressant therapy prescription, do psychologists help? - Evidence from Portugal[12]

Abstract

Most patients with depressive disorders are treated in primary care, where General Practitioners (GPs) may prescribe them pharmacological treatment and/or psychotherapy, both endorsed by guidelines. This choice is influenced by both GP and patient preferences. While lack of access to psychological therapy is often pointed out as a determinant of antidepressant prescription, there is limited empirical evidence to substantiate this claim. We investigated the relationship between psychotherapy supply and GP antidepressant prescription in Portuguese primary care, a country with one of the highest consumptions of these drugs and a very low number of psychologists working in the national health service.

We used panel data on all GPs in Portugal between 2015 and 2018 and a within-between random effects model to investigate the effects of changing numbers of psychologists in GPs´ prescription over time (within variation). Additionally, we explored the association of prescription patterns with between GPs variation in the number of psychologists. We found heterogeneous effects of increasing one psychologist per 100,000 patients on GPs´ prescription share overtime. While there was no average effect, the subgroup analysis of GPs exposed to the lowest and highest quartiles of average number of psychologists showed that increasing one psychologist/100,000 patients overtime led to a reduction of 2.9% and 4.4% in the

[12] This chapter is co-authored with Francisca Vargas Lopes.

prescription share, respectively. Findings for the between term showed that GPs facing a one-unit higher average number of psychologists/100,000 patients have a lower prescription share, by 3.6%. Our results suggest that in the context of scarce resources, small increases in the number of psychologists might not lead to a reduction of antidepressant prescription across all levels of supply. Furthermore, policy action should consider the range of unmeasured characteristics and processes behind the fact that GPs with lower antidepressant prescription are working in settings with higher average supply of psychological therapy.

2.1 Introduction

Depression is a common mental health disorder and a major cause of disability and economic burden worldwide (Vigo, Thornicroft, & Atun, 2016; Vos et al., 2020). Most patients with depressive disorders have mild and moderate cases and are treated in primary health care (Kovess-Masfety et al., 2007; Olfson, Kroenke, Wang, & Blanco, 2014; Young, Klap, Sherbourne, & Wells, 2001). When deciding how to treat these patients, general practitioners (GPs) can choose between prescribing pharmacotherapy, psychological interventions – usually provided by mental health workers – or a combination of both (Park & Zarate, 2019). Psychotherapy is the first-line recommendation in most treatment guidelines for patients with mild cases, whereas antidepressants, psychotherapy or a combination of both are all first-line recommendations for moderate depression (American Psychiatric Association, 2009; Anderson et al., 2008; Haute Autorité de santé, 2017; Kennedy et al., 2016; National Institute for Health Care Excellence, 2022; Schneider, Härter, & Schorr, 2017).

In most health care systems, the integration of mental health workers into primary care has been happening for the past decades. This integration can lead to organizational changes and modification of GPs´ clinical routines (Harkness & Bower, 2009) and it increases the availability of psychological interventions such as psychotherapy, which GPs would usually not have sufficient training or time to provide (Verdoux, Cortaredona, Dumesnil, Sebbah, &

Verger, 2014). At the same time, increased antidepressant utilization has also been registered in several countries (Kjosavik, Hunskaar, Aarsland, & Ruths, 2011; Mojtabai & Olfson, 2011; Munoz-Arroyo, Sutton, & Morrison, 2006; Van Marwijk, Bijl, Adèr, & De Haan, 2001). Despite representing an expansion in patient treatment, these trends have raised concerns about the benefit-risk ratio of antidepressants for the treatment of mild and moderate cases, and about increasing drug costs (Druss, 2006; Fournier et al., 2010; Kirsch et al., 2008). As primary prescribers of antidepressants (Druss, 2006; Mojtabai & Olfson, 2011), GPs have been at the core of this debate, and a literature stream studying their opinions and practices emerged, mostly using surveys and qualitative studies (Dumesnil et al., 2012; Mercier, Auger-Aubin, Lebeau, Royen, & Peremans, 2011; Verdoux et al., 2014). Some of these studies suggest that limited access to non-pharmacological treatment resources is a determinant of choice favouring medicines prescription (Morrison et al., 2009; Verdoux et al., 2014); others failed to identify such association (Dumesnil et al., 2012; Sibbald, Addington-Hall, Brenneman, & Freeling, 1996). In fact, there is not enough empirical evidence to back up the commonly stated hypothesis that increasing psychological therapy resources available to GPs would decrease antidepressant prescriptions (Estrela, Herdeiro, Ferreira, & Roque, 2020; Verdoux et al., 2014).

We used nationwide data from Portugal to investigate the relationship between the number of psychologists available in primary care and GP prescription of antidepressants to patients with depression, between 2015 and 2018. In Portugal, psychologists are hired as part of local groups, which are primary care governance structures that bring together several GP practices and manage their shared resources. The number of psychologists in primary care is in all regions considered too low in comparison to other countries, and to the standards set by the Portuguese Assembly recommendation that defines a ratio of one psychologist per 5000 inhabitants (ERS, 2023). International comparisons do also suggest that Portugal has a high level of antidepressant use, in fact the third highest sale volume of antidepressants amongst OECD-26

countries in 2015 (Estrela et al., 2020). As in other countries with barriers to access to psychological services (Verdoux et al., 2014), the limited availability of psychologists in the Portuguese national health system (NHS) is believed to contribute to the excessive use of psychotropic medications (Estrela et al., 2020).

Several mechanisms can be outlined both in favour and against this theoretical hypothesis that an increased supply of psychological therapy leads to the decreased prescription of antidepressants. As in other countries, Portuguese GPs work with large patient lists and limited appointment time. This means that shifting care to other practitioners may have considerable efficiency gains, especially when it does not have direct budget implications for GPs and when there is no competition for patients on the quality of care, as in the Portuguese case. With psychotherapy being the first-line recommendation for mild depression and the accumulation of evidence supporting its efficacy versus care-as-usual but also pharmacotherapy (Cuijpers et al., 2021; Cuijpers, Miguel, Harrer, Plessen, Ciharova, Ebert, et al., 2023; Cuijpers, Miguel, Harrer, Plessen, Ciharova, Papola, et al., 2023; Plessen, Karyotaki, Miguel, Ciharova, & Cuijpers, 2023), it would be reasonable for GPs to substitute antidepressants prescription by a referral to mental health workers working in primary care. This substitution would also free GP´s time for additional activities, as the (initial) prescription of antidepressants requires monitoring and a more intensive follow-up than a referral. On the other hand, GP surveys and qualitative studies suggest that treatment decisions are significantly influenced by GPs' beliefs, attitudes, and personal experiences (Andersson, Troein, & Lindberg, 2001; Dumesnil et al., 2012; Mercier et al., 2011). Practical aspects such as medication fitting more easily into the routine of clinical practice might also influence the choice (Van Schaik et al., 2004). Literature also shows that while GP perception of psychotherapy effectiveness is high and has increased over the years, GPs still prefer to use it in combination with antidepressants, rather than psychotherapy alone (Dumesnil et al., 2012; Haddad et al., 2012; Verdoux et al., 2014). Faced

with the lack of guidelines consensus on first-line treatment for moderate depression, GPs may prefer to use the increasingly available psychotherapy to combine it with antidepressants for higher severity patients, instead of using it as a substitute for medication in milder cases.

We exploited the nested and longitudinal nature of our data to study the variation in the number of psychologists (in full time equivalents) - a proxy of psychological therapy supply – by decomposing it in within GP and the between GP variation. The within variation shows how changes in the number of psychologists in each local group led to changes in prescription by GPs exposed to these changes overtime, between 2015 and 2018. As in a fixed effects frameworks, the within estimate obtained from our model is not affected by omitted variable bias due to unobserved time-invariant confounders, thereby providing stronger evidence about the causal relationship of interest (Bell, Fairbrother, & Jones, 2019). Such biases are absorbed by the between term, which captures the variation in the number of psychologists between GPs working at different local groups (Bell & Jones, 2015; Fairbrother, 2016). Importantly, the between estimate is influenced by unmeasured factors related to both the number of psychologists and GPs' prescription behaviour. The potential bias of the between GP coefficient makes its comparison with the within GP coefficient informative. A large difference between these terms suggests an important role of unmeasured characteristics and processes in the relationship observed between the supply of psychological therapy and antidepressant prescription.

We conducted heterogeneity analyses to further explore the mechanisms at play in the relationship of interest. Firstly, we explored heterogeneity in the within term by the average level of psychologists available to each GP during the study period, which ranged from 0 to 8 psychologists/100,000 patients in the local groups. Given this wide range, the addition of one psychologist/100,000 patients might represent a relative change of 100% to 12% in supply, respectively, and is expected to have a non-linear association with prescription. Secondly, we

examined how the relationship of interest was influenced by certain characteristics of GPs, such as sex, age, tenure, patient list size, and degree of urbanization. These characteristics have been presented in the literature as determinants of antidepressant prescription, and heterogeneous results would provide insight into the role of GPs' beliefs, attitudes, and personal experiences in the treatment decision . Thirdly, the magnitude of the relationship was explored based on the average antidepressant prescription share of each GP. Because our measure of prescription is a share, being standardized for the prevalence of depression in the patient list, prescription patterns should reflect mostly GP treatment preferences towards psychotherapy or antidepressants, rather than patient case mix. In grouping GPs by how often they prescribe we explore how the unmeasured characteristics shaping GP preferences influence our results. The paper is structured as follows: section 2 presents background information describing the Portuguese context and the study setting, and section 3 introduces the data, our empirical approach and statistical analysis. Section 4 starts with some descriptive results and moves on to the findings of our models, and section 5 discusses these results and concludes.

2.2 Setting

2.2.1 Depression and treatment approaches in Portugal

Portugal has a high prevalence of psychiatric disorders compared to other European countries, with common mental disorders making the biggest difference (Wang et al., 2011). According to the National Epidemiological Study of Mental Health, the 12-month prevalence of depression was 6.8% in 2013, and 19,3% of the Portuguese population would experience depression during their lifetime (Caldas de Almeida et al., 2013). As in other countries, the prevalence of psychiatric disorders in Portugal tends to differ from diagnosis rates in primary care (Ayuso-Mateos et al., 2001). Compared to the 12-month prevalence of psychiatric

disorders of 22,9% (Caldas de Almeida et al., 2013), the proportion of patients identified as having a mental health condition in Portuguese GP´s patient list ranged from 29% to 59% (Comissão Técnica de Acompanhamento da Reforma da Saúde Mental, 2017).

Portuguese consumption of antidepressants has also been consistently high for the past years. The country had the highest percentage of the population reporting taking antidepressants in Europe in 2010 (15,7%) (Lewer, O'reilly, Mojtabai, & Evans-Lacko, 2022) and antidepressants accounted for 3.7% of the volume of drugs sold in Portugal in 2020, ranking second among OECD countries (Estrela et al., 2020). On the other hand, there is limited availability of psychotherapy services in the Portuguese NHS. In 2019, there were 9.6 NHS psychologists for every 100,000 inhabitants in Portugal (Caldas et al., 2020), a figure that is much lower than most European countries and the (minimum) recommended 20 psychologists / 100,000 inhabitants. Most of these psychologists work in hospitals, with only approximately 300 being part of primary care (ERS, 2023). These reduced numbers persist despite the high number of qualified clinical psychologists in the country, around 5,000, that work mostly in the private sector (Ordem dos Psicólogos Portugueses, 2015).

2.2.2 Portuguese primary care and mental health support

The Portuguese NHS is funded through taxes and follows the Beveridge model, in which GPs act as a gatekeeper to other primary and secondary care treatment options. Patients can only register with a GP practice in their residence area, and choice is limited due to the GP shortages in some areas. In 2018, 7.5% of patients enrolled in GP practices were not yet registered with a GP. Primary care practices are typically organized into teams consisting of GPs, nurses, and administrative staff, and are owned by the state. The staff, which almost exclusively includes salaried civil servants, is recruited through public hiring processes.

Since 2008, GPs practices were administratively grouped into local groups, known as ACeS (*Agrupamentos de Centros de Caúde*). Local groups' aim was to manage primary care resources, services planning and articulating the various GP practices within the coverage area and to develop health promotion and disease prevention actions. Local groups are typically responsible for overseeing the population of one municipality (except in more sparsely populated areas, where local groups cover more than one municipality, and densely populated areas, such as metropolitan areas, where there may be multiple local groups). Each local group encompasses a unit of shared resources (URAP, *Unidade de Recursos Assistenciais Partilhados*) which provides access to various professions such as nutritionists, physiotherapists, occupational therapists, psychologists, and social workers, pending GP referral. Psychologists integrated into primary care units offer various mental health services such as psychological assessment, psychotherapy, and behavioural counselling. While they cannot prescribe medicines, they should work with GPs to ensure that patients receive appropriate medication as needed, and collaborate with other healthcare professionals to develop comprehensive care plans that address patients' physical and psychological needs (Ordem dos Psicologos, 2018). Besides the national guideline for major depression from the Directorate-General for Health (DGS, 2012), most local groups have their recommendations indicating which patient groups should be referred to psychologists by GPs[13]. These local protocols were often developed in a context of short supply of psychologists, to aid prioritization of certain groups. GPs' global prescription pattern is monitored periodically in a national reporting system published online. Despite that, there are no incentives (monetary or

[13] Informal queries to GPs in various locations suggest that referral protocols used by local groups can vary according to the population needs and resources available. Referral recommendations range from referring all patients with mild depressive symptoms, to prioritizing long-term care, patients in smoking cessation, pregnant women, etc. or excluding some patients, such as excluding some psychopathologies in children and adults that should resort to psychiatric and pedo-psychiatric services.

reputational) that shape GPs' prescription of antidepressants and psychologist referral decisions.

During our study period, patients paid a small fee for primary care appointments, which ranged between 4.5 and 5 euros per GP consultation and 3.5 euros per psychologist appointment[14]. Several groups were exempted from this co-payment, including poor patients and pregnant women (Barros, 2012). Most patients would also need to co-pay for their antidepressants. While the exact amount paid out of pocket varied over time and with the drug used, brand/generic choice, dosage and pack size, the monthly co-payment of antidepressant medicines would usually not exceed a few euros.[15]

2.3 Methods

2.3.1 Data

We used data from three nationwide sources. First, the administrative database SDM@SNS (*Sistema de Dados Mestre do Serviço Nacional de Saúde*), from the Central Administration of the Health System (*Administração Central do Sistema de Saúde; ACSS*), which contains data on a vast set of indicators monitoring all primary care providers. Second, the human resources register RHV (*Recursos Humanos e Vencimentos*), which contains information on the primary care workforce, including GPs and psychologists. Third, municipality-level characteristics

[14] Price of mental health services available at the Central Administration of the Health System (ACSS) website in https://www.acss.min-saude.pt/2016/10/03/tabelas-meios-complementares-de-diagnostico-e-terapeutica/ (visited in January 2023). Primary care visits copayments were extinguished in 2020.

[15] Taking sertraline 50 mg and the year of 2023 as an example we observe the out-of-pocket cost for packages with 60 pills ranges between 1,99 and 7,43 euros for the generic drugs, and 12,05 euros for the brand drug. Given that sertraline has a daily defined dosage (DDD) of 50 mg the monthly cost of treatment could then go from 1 to 6 euros. Information about prices is made available by the national drug authority (Infarmed): https://app10.infarmed.pt/genericos/genericos_II/lista_genericos.php?tabela=spr&fonte=dci&escolha_dci=U2VydHJhbGluYQ==(visited in April 2023).

were obtained from the Portuguese national statistics bureau (*Instituto Nacional de Estatística - INE*) and Eurostat. The first and second sources were linked at the GP level, while the third source was linked at the GP practice level, based on the practice location.

2.3.2 Study Population

Our study population consists of an unbalanced panel of the universe of GPs in the Portuguese NHS over 4 years (n= 5,359, 2015 to 2018). The study population includes full-time GPs working at least 12 months in a GP practice during the study period. The small number of GPs that switched practices between different local groups were excluded (3.8%), as well as observations from GP practices that underwent migration of the information software during the study period (0.4%); and observations of GPs with patient lists smaller than the first percentile of the list size distribution (1.9%; likely data coding issues or exceptional situations occurring in practice transition periods). Last, we excluded observations with missing values in the variables of interest for our main models (2.9%). A flow diagram describing the selection of the study population is presented in the supplement (Figure 2.4).

2.3.3 Variables

The outcome variable is GP antidepressant prescription share, among patients identified with depression in GP´s patient list. This share was calculated for each year by dividing the total number of adults (18 years and older) identified by the GP with a diagnosis of depression and who were prescribed antidepressant therapy, by the total number of adults identified by the GP with a diagnosis of depression. All major classes of antidepressants alone or in combination

were considered (group 2.9.3 of the national pharmacotherapeutic classification of drugs[16], equivalent to group N06A from the Anatomical Therapeutic Chemical Classification System). GPs diagnosis of depression followed the International Classification of Primary Care 2 (ICPC-2) and included affective psychosis, depressive neurosis, depressive psychosis, mixed anxiety and depression, reactive depression, and postnatal or puerperal depression (P76). Data on our outcome variable - from now onwards often referred to as antidepressants prescription - is usually collected in primary care to monitor the prescription behaviour. Importantly, this indicator is not used to measure GPs performance or to apply incentives, because it does not allow conclusions to be drawn about the appropriateness of antidepressant use.

Our exposure variable of interest is the supply of psychological therapy services by psychologists. It was computed for each year as the number of psychologists in full-time equivalents (FTE) per 100,000 patients enrolled in the local group. Primary care psychology teams are multidisciplinary and can include specialized nurses, occupational therapists, and social workers. We excluded other professionals rather than psychologists from our explanatory variable because, in the Portuguese primary care, they do not provide psychological therapy interventions.

Furthermore, a comprehensive set of GP and GP practice covariates were included in our models, informed by literature and the Portuguese context. Physician characteristics included are age in years, sex (female or male), tenure in the GP practice (1 to 2, 3 to 5, or more than 5 years), patient list size and the utilization share (share of patients with at least one yearly visit). Other covariates accounting for the patients' list characteristics measured its demography (proportion of female patients and patients in each age group) and morbidity (prevalence of dementia, chronic obstructive pulmonary disease, cancer, asthma, hypertension, obesity, and

[16] Ministry of health order nr 2977/2014. Available online: https://dre.pt/dre/detalhe/despacho/4742-2014-25681511 (accessed on January 2023).

diabetes). Practice-level covariates controlled for the practice organizational model (personalized health care units – PHCU; family health units type A and B - FHU-A and FHU-B), the number of years the practice had been operating in the current arrangement, the share of patients enrolled in the practice but not listed with a GP, the ratio of nurses to GP in the practice and the practice incidence rate of depression (i.e. the share of yearly diagnosed cases among the patients enrolled in the practice). Additionally, we controlled for the local context in which the practice was located by considering whether it is part of an integrated care scheme (*Unidades Locais de Saúde* - ULS) but also by accounting for municipality-level characteristics such as the population density, average earnings of employees, unemployment rate and the degree of urbanization (cities, towns or suburbs and rural areas [17]).

2.4 Empirical approach

To examine how the GP antidepressant prescription changes with different levels of supply of psychological therapy we start from a (pooled) baseline model:

$$Y_{ijl} = \alpha_0 + \alpha_1 P_l + X'_{ijl}\alpha_2 + Z'_{jl}\alpha_3 + \varepsilon_{ijl} \quad (5)$$

where Y_{ijl} represents the antidepressant prescription rate of GP i, working in practice j, that belongs to local group l. P_l is the measure of psychotherapy supply facing all GPs at local group l; X_{ijl} is a vector of GP characteristics and patients list characteristics, that accounts for observed differences in average medical need, in the absence of individual patient data; Z_{jl} is a vector of GP practice characteristics, and ε_{ijl} is the error term. In the baseline model the coefficient of interest, α_1, measures the association between antidepressant prescription and the availability of psychological therapy. It is important to note that the measure of

[17] Municipality urbanization classification according to Eurostat 2011. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Degree_of_urbanisation_classification_-_2011_revision (visited in Jan 2023)

psychological therapy supply deals with the availability of therapy services rather than their actual utilization, rendering it exogenous in our analysis. In contrast, the utilization of psychotherapy services is considered endogenous, as it is decided alongside prescription. The estimation of this association (α_1) might be nevertheless biased by unobserved provider characteristics such as GP characteristics that systematically influence their decision-making in treating depression (e.g., knowledge, beliefs, etc.), and unmeasured factors driving psychologists hiring by local groups. While mental health need should explain a great part of psychologist's supply, other factors such as clinical leadership and the maturity of mental health services integration in each local group might also play a strong role. These could, for example influence GP and psychologist sorting into the local groups. One other potential threat to the robustness of α_1 might arise from reverse causality, namely when no psychologists would be hired because GPs prefer to prescribe antidepressants than refer patients to psychotherapy.

Therefore, to better investigate what is driving the association between antidepressant prescription and psychological therapy supply captured by α_1, we decomposed this relationship by examining two types of variation. On the one hand, the variation in the number of psychologists faced by each GP in a local group overtime *(within term)*. On the other hand, the variation in the average number of psychologists available to GPs across different local groups (*between term*). We estimated these terms using a within-between random effects model (REWB), sometimes designated as an "hybrid" model (Bell et al., 2019; Bell & Jones, 2015).

The within term uses the yearly variation in the number of psychologists to which the GP is exposed to through her local group to measure the effects on antidepressant prescription for this GP overtime. The within term is free from omitted time-invariant bias, as in the statistically equivalent fixed-effects model (Bell et al., 2019; Bosque-Mercader & Siciliani, 2023). Because the identification is based on observing the same GP under different levels of psychotherapy supply, the within term is purged of bias by GP characteristics that are likely constant during

the 4-year study period (e.g., education and preferences). It is also free from the influence of time invariant local group characteristics that are eliminated as addictive effects captured by the nested structure of GPs into local groups (Wooldridge, 2010).

In addition to the within term, the REWB model offers the flexibility to include random effects and other time-invariant predictors, as noted by Mundlak (1978). In a fixed effects model, time-invariant GP variables would be absorbed by the GP fixed-effect. This would prevent investigating how the variation in the average level of psychologists between GPs in local groups is related to their antidepressant prescription, which we can do with the REWB. Unlikely the within term, the between GPs term is not free from time-invariant bias, and it is likely confounded by GP/local groups unobserved variables linked to both psychologist supply and antidepressants prescription. The economic interpretation of the between term is interesting precisely because it proxies a range of unmeasured characteristics and processes shaping the association observed. By comparing the between and within terms we can conclude about the relevance of these unmeasured characteristics and processes. If the terms are similar in magnitude, then the variation within GP is the most relevant. On the other hand, a larger between term should flag the need to understand and target the unobserved GP and local groups' determinants of depression treatment which are driving the association. Examples of factors explaining why certain local groups have lower (higher) psychologists supply, and higher (lower) antidepressant therapy include the level of mental health services integration, local referral and treatment protocols, and GPs awareness and literacy on the topic.

2.4.1 Statistical model

The key aspect in estimating the REWB model lies in the inclusion of the mean of the variable of interest as an additional explanatory variable ($\overline{P_l}$). This is often described as the Mundlak approach (Mundlak, 1978) and is closely related to the correlated random effects model (Wooldridge, 2010). Following Bell and Jones (2019) (Bell et al., 2019) we further included in

the model the demeaned explanatory variable ($P_{lt} - \bar{P}_l$) instead of its raw time-varying form (P_{lt}). While the models themselves would be equivalent, the use of the demeaned explanatory variable allows for the interpretation of the between term as not being controlled for the raw value of P_{lt}. The REWB model is specified as:

$$Y_{it} = \beta_0 + \beta_1(P_{it} - \bar{P}_l) + \beta_2\bar{P}_l + (W'_{it} - \bar{W}'_i)\beta_3 + \bar{W}'_i\,\beta_4 + \varepsilon_{ijt} \quad (6)$$

Where W'_{it} includes X'_{it}, Z'_{ijt} and θ_t and $\bar{W}'_i$ includes $\bar{X}'_i$, $\bar{Z}'_{ij}$ and $\bar{\theta}$; all explanatory variables – both the number of psychologists but also covariates - are replaced by their GP-specific means over time and deviation from this mean. For better estimation of the partial association captured by the between term β_2 we adjusted for the patient case mix, and for GP, GP practice and local context characteristics described in section 2.2.1. Finally, we controlled for year (θ_t) to account for time trends and (policy) shocks that are common to all GPs, practices, and local groups. Importantly, we have adjusted standard errors by clustering at the local group level, which is the level at which the psychologist's number is defined. All analyses were performed using STATA 17 software (StataCorp., 2021).

2.4.2 Heterogeneity and robustness analyses

We conducted several heterogeneity analyses to explore the potential mechanisms driving the relationship of interest, particularly looking at heterogeneity in the within estimates. First, we explored whether any changes in prescription due to increased number of psychologists overtime would have a different magnitude for GPs in local groups with higher vs. lower average number of psychologists. We hypothesize that such non-linearities might exist because some local groups have a very low (or zero) number of psychologists, where adding 1 psychologist/100,000 patients leads to a much larger relative increase in psychotherapy supply than adding the same 1 psychologist in an established team of n psychologists. Furthermore, pharmacotherapy and psychotherapy might be used as substitutes (mild patients) or in

combination (moderate patients). When used in combination, no (immediate) reduction in prescription would be observed due to additional psychotherapy. Depending on their preferences and beliefs about the effectiveness of psychotherapy as a single or combined therapeutic approach GPs might choose to prioritize additional psychotherapy supply differently depending on the previous (un)availability of such resources in their local groups. We classified GPs in quartiles of the average number of psychologists they are exposed to in the local group during the study period and estimated our main model stratified by quartile.

Second, we conducted additional stratified analysis by determinants of antidepressant prescription described in the literature: GP sex, GP age, GP length of tenure in NHS, patient list size and urbanization degree of the practice location (Morrison et al., 2009; Verdoux et al., 2014). With these analyses, we aimed at understanding whether GPs with certain (practice) characteristics were more prone to respond to the changes in the psychotherapy supply over time. Following a similar rationale, we have conducted stratified analysis by dividing GPs into quartiles of antidepressant prescription (average over the study period). This analysis aimed at exploring heterogeneity by unmeasured GP characteristics that influence prescription behaviour, such as their personal history of depression or personal experience with pharmacotherapy and psychotherapy (Dumesnil et al., 2012; Verdoux et al., 2014).

We also conducted four additional analyses to support the robustness of our findings. We started by examining whether psychologists' supply was associated with changes in prescription but not with the depression diagnosis. Because the capacity and ability of GPs diagnosing depression may be affected by the presence of more psychologists, this could distort the findings of our outcome, which measures an antidepressant prescription share standardized to depression prevalence. For examining this possibility, we replaced the dependent variable in Model 2 with the number of newly diagnosed cases of depression at the GP level. Second, we examined whether the variation in the number of psychologists could be partially a response

to GP patterns in antidepressants prescription. Based on the concept of reverse causality it could be that no (more) psychologists would be hired because GPs would not refer patients to psychotherapy anyway. To exclude this hypothesis, we estimated whether changes in the number of psychologists were associated with former antidepressant prescription shares (Y_{it-1} and Y_{it-2}). Third, we estimated our baseline model excluding the variables of physician list size and the utilization share of the GP visits, to avoid potential endogeneity if diagnosis capacity would be improved due more interaction and closer relationship between GP and patients. Furthermore, excluding the utilization share helps us to addressing the concern that amount of GP visits may be affected by the depression treatment decision and thus bias the association with psychotherapy. Last, we investigated the robustness of our results using a balanced panel of GPs in practice during the 4 years of the study period.

2.5 Results

2.5.1 Descriptive statistics

From 2015 to 2018 our population follows the prescription decisions of 5,359 unique GP (n=17,210 observations) in 888 unique GP practices grouped in the 55 local groups that exist in Portugal. The smallest local group provided care to 30,493 patients and the largest to 386,027 patients in 2018.

Error! Reference source not found. summarizes the characteristics of the study population. The share of antidepressants prescribed to adults identified as having depression was on average 27%. The standard deviation (SD) figures suggest that there is limited variation over time (SD 3.97) but a considerable variation between GP (SD 14.6, ranging from 0% to 96%). Most GPs were female (73%) and worked in the same GP practice for more than 5 years (60%), having an average of 1,710 patients on their patient list, from which 69.9% effectively used

services. The most prevalent conditions were hypertension (22.7 % of the total population in the GP list), obesity (9.2%) and diabetes (8.2%). About 47.6% of the patients on the list were exempted from co-payments. GP practices would more frequently follow the model of PHCU (38.7%) and FHU-B (34.7%) and had operated on average for 7.3 years[18]. The proportion of patients in the practice without GP varied substantially from 0 to 77.8%, and the ratio of nurses to GPs was on average 1.12. The incidence rate of depression at the practice level was 9.2%. Last, there were on average 1.98 FTE psychologists for every 100,000 patients in the local group, varying from local groups that had 0 psychologists at some point during our study to a maximum of 7.59/100,000.

Figure 1. provides some insight into the time and regional variation of our key variables. The antidepressant prescription share fell from 27.6% to 25.7% from 2015 to 2018; the top panel shows considerable asymmetries across regions. These patterns are only partially explained differences in the prevalence of depressive disorders (Figure 2.5 of the appendix). Certain regions with low prevalence have higher rates of prescription, notably some regions in the great Lisbon area. The maps also attest to the persistence of the antidepressant prescription over time. The average number of psychologists in the local groups increased during the study period, from 1.8 (2015) to 2.3 (2018) FTE/100,000 patients registered in the local groups. However, the bottom panel of Figure 2.1 shows that this increase was not consistent across the country, with some regions still having 0 psychologists in their local group in 2018, or other regions in which the number decreased.

[18] Please note that all the GP practices were subjected to a reform in 2006, which is therefore considered as starting year.

Table 2.1: Descriptive statistics of the study population.

	Mean	Standard deviation			Min	Max
		Overall	between	within		
GP-level variables						
Antidepressant prescription share	26.876	15.072	14.541	3.969	0.000	96.429
Male [0,1]	0.371	0.483	0.483	0.000	0.000	1.000
Age in years	53.331	11.215	11.826	1.026	30.000	80.000
GP Tenure (1-2y) [0,1]	0.210	0.407	0.379	0.247	0.000	1.000
GP Tenure (3-5y) [0,1]	0.190	0.392	0.251	0.311	0.000	1.000
GP Tenure (>5y) [0,1]	0.601	0.490	0.451	0.227	0.000	1.000
Patient list size	1710.8	215.4	215.4	56.9	599	2866
% Utilization rate	69.902	8.665	8.825	2.308	0.428	97.801
% Exempt co-payments	47.643	8.197	8.043	2.149	19.048	84.269
% Female patients	52.865	1.337	1.321	0.246	47.259	57.891
% Patients aged 0-4y	4.021	0.846	0.844	0.179	0.637	7.607
% Patients aged 5-14y	8.547	1.411	1.392	0.284	1.402	14.179
% Patients aged 15-44y	37.770	3.474	3.451	0.641	15.233	49.232
% Patients aged 45-64y	27.607	1.890	1.860	0.416	20.657	35.243
% Patients aged 65-74y	11.134	1.802	1.776	0.375	4.565	18.929
% Patients aged >75y	10.921	3.628	3.616	0.400	3.020	41.810
% Patients Dementia	0.872	0.627	0.604	0.146	0.000	8.647
% Patients COPD	1.394	0.896	0.865	0.212	0.000	11.911
% Patients Cancer	3.990	1.355	1.314	0.355	0.000	9.964
% Patients Asthma	2.866	1.556	1.499	0.280	0.000	13.143
% Patients Hypertension	22.629	6.559	6.599	1.156	0.327	66.118
% Patients Obesity	9.214	5.486	5.184	1.621	0.000	33.952
% Patients Diabetes	8.184	2.270	2.299	0.480	0.109	25.802
% Patients Depression	9.703	4.623	4.519	0.898	0.298	36.110
GP practice-level covariates						
PHCU [0,1] #	0.387	0.487	0.486	0.072	0.000	1.000
FHU-A [0,1] #	0.266	0.442	0.437	0.100	0.000	1.000
FHU-B [0,1] #	0.347	0.476	0.455	0.077	0.000	1.000
Number of years in operation	7.300	2.995	2.892	1.133	0.000	12.000
Patients without GP [%]	6.058	12.328	11.790	4.960	0.000	77.870
Ratio Nurses to GP	1.117	0.374	0.348	0.158	0.312	6.250
Incidence of depression [%]	9.235	3.905	3.372	2.137	0.328	56.281
Local group*-level variables						
Psychologists/100,000 patients	1.976	1.314	1.246	0.476	0.000	7.586
Integrated Care (*ULS*) [0,1]	0.126	0.332	0.337	0.000	0.000	1.000
Population density*	1543.426	2149.705	2174.459	177.349	4.154	14780.560
City [0,1]	0.449	0.497	0.496	0.026	0.000	1.000
Town or suburb [0,1]	0.382	0.486	0.483	0.054	0.000	1.000
Rural area [0,1]	0.169	0.374	0.377	0.051	0.000	1.000
Average monthly earnings (employed)*	1051.025	220.702	221.050	30.594	702.200	2331.200
% Unemployment*	7.091	2.536	2.286	1.146	2.400	16.100
Observations**			17,210			

Note: Since a reform in 2006, the previously existing GP practices are called personalized health care units (PHCU) and coexist in the Portuguese ecosystem with two types of performance-incentivised team-based practices: family health units (FHU) of type A and B (FHU-B). In the initial stage, all newly created FHUs were of type A and had a team-based organization and group incentives. FHU-A could then apply to migrate to FHU-B, with GP individual pay-for-performance incentives and more stringent performance targets. *Some of the covariates are reported at the municipality level, which corresponds roughly to the local group's geographical boundaries. **GP-year observations: 2015=4,045; 2016=4,358; 2017=4,612; 2018=4,195; corresponding to 5,359 unique GPs working in 888 GP practices of 55 local groups.

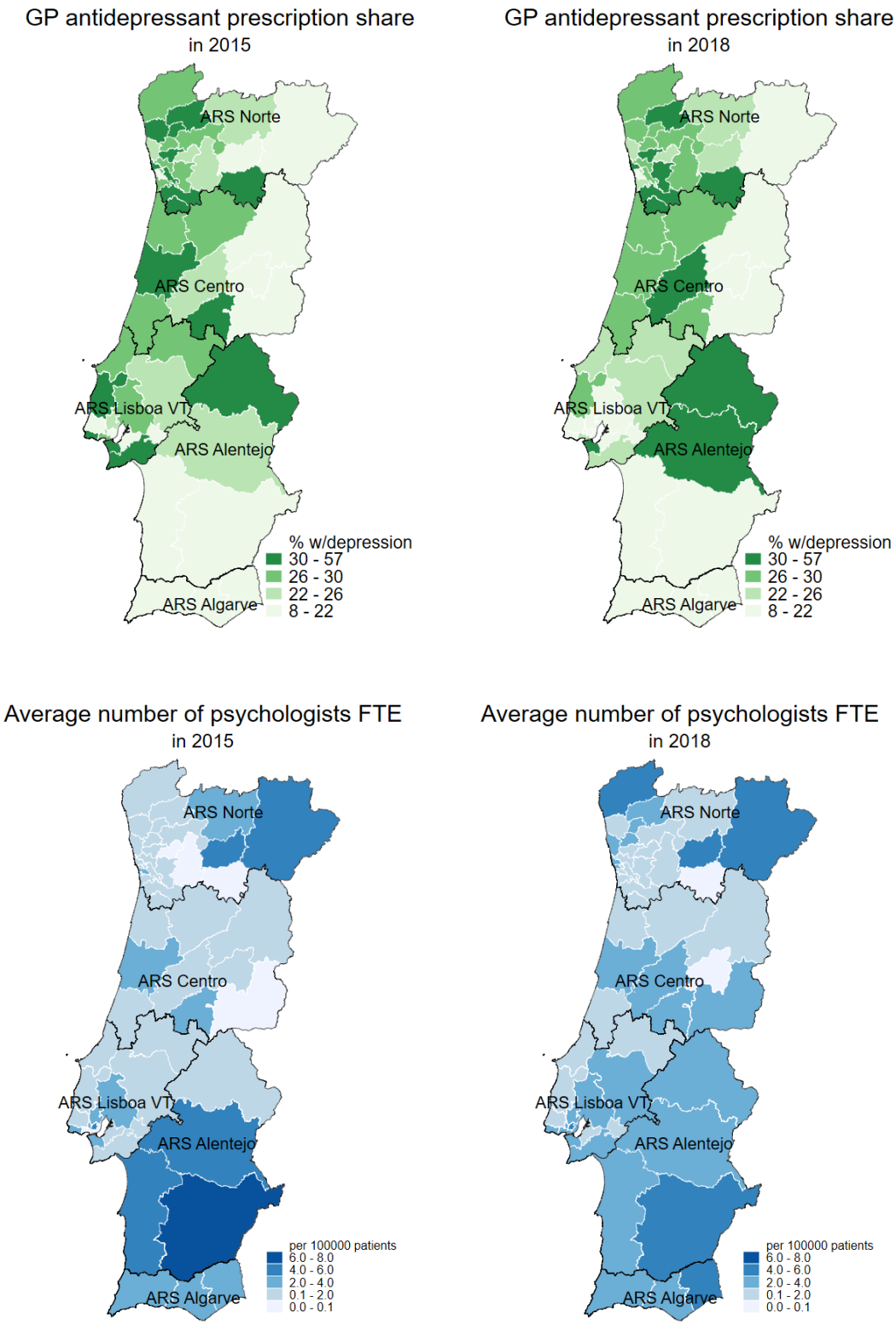


Figure 2.1. Geographical characterization of the GP antidepressant prescription share of patients with depression (top panel) and the average number of psychologists FTE in the local group per 100,000 enrolled patients (bottom panel) in 2015 (left) and 2018 (right).

Figure 2.2 sheds some initial insight into the relationship between GP antidepressant prescription and the number of psychologists in the local group. It shows that there is considerable variation in antidepressant prescription across local groups and that local groups where GP prescribe antidepressants to less than 20% of the patients diagnosed with depression are also more often in the highest quartile of average psychologist supply.

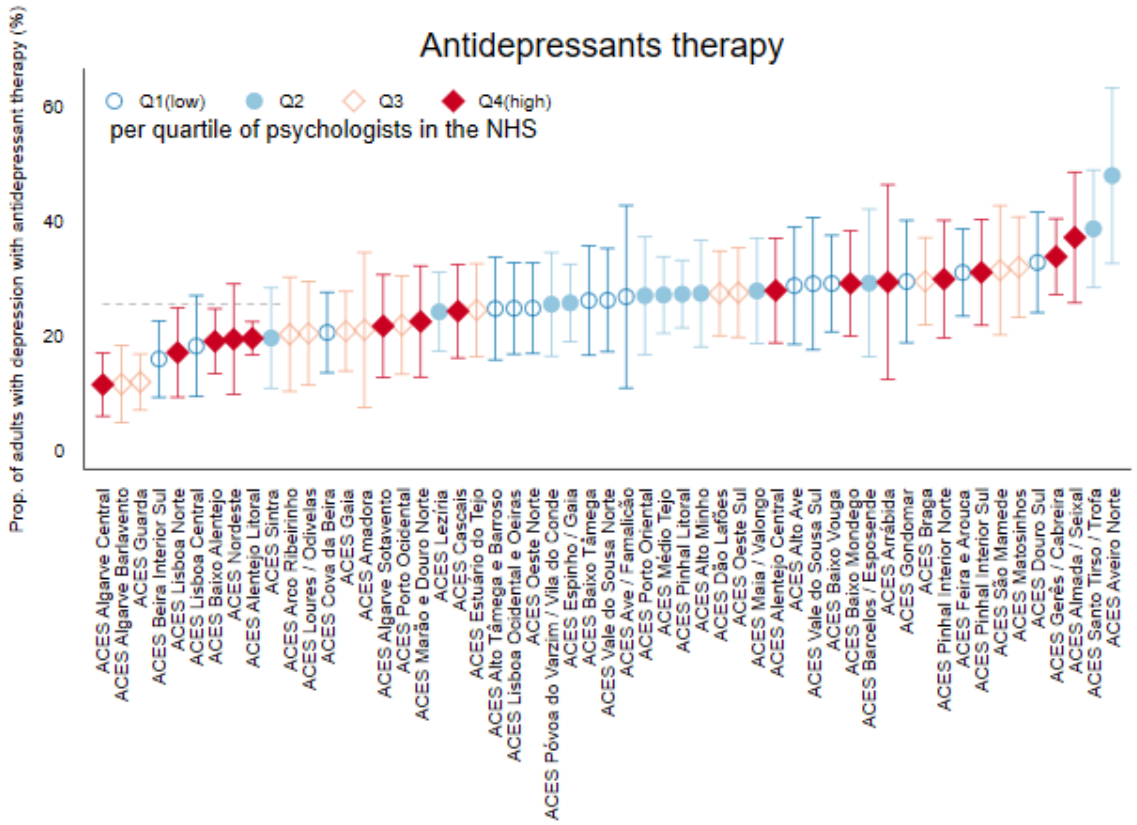


Figure 2.2: Relationship between the share of depressed adults with antidepressants prescription and the average number of psychologists available at the local group level.

Notes: The y axis distinguishes between high and low prescribing local groups (average and standard deviation sorted), while the number of psychologists is represented by the shape and colour of the marker (in quartiles).

2.5.2 Within-between random effects model

Table 2.2 reports results from the REWB model decomposing the relationship between antidepressant prescription and psychological therapy within and between variations. Different

sets of covariates were sequentially added to the estimations, ranging from the unadjusted (1) to the fully adjusted (5) and our preferred specification.

Table 2.2: Within-between random effect models studying the association of psychotherapy supply and antidepressant prescription share.

	No controls	Year dummies	GP	GP practice	Main specification
	(1)	(2)	(3)	(4)	(5)
Psychologists/100,000 patients					
Within-term	-0.943***	-0.268	-0.263	-0.270	-0.305
	[-1.334,-0.551]	[-0.737,0.202]	[-0.688,0.161]	[-0.688,0.148]	[-0.706, 0.0954]
Between-term	-1.819***	-1.828***	-1.467***	-1.323***	-0.955**
	[-2.810,-0.828]	[-2.810,-0.845]	[-2.132,-0.803]	[-2.014,-0.632]	[-1.641, -0.269]
Year dummies	No	Yes	Yes	Yes	Yes
GP covariates	No	No	Yes	Yes	Yes
GP practice covariates	No	No	No	Yes	Yes
Local context * covariates	No	No	No	No	Yes
Constant	30.26***	30.14***	25.23	47.97	22.45
	[27.95,32.57]	[26.97,33.31]	[-35.24,85.71]	[-16.03,112.0]	[-48.56,93.46]
N	17,210	17,210	17,210	17,210	17,210
R^2 within	0.0128	0.0525	0.0697	0.0721	0.0756

Notes: The dependent variable is the share of adults with depression in the GP list that have been prescribed antidepressants. Covariates are included in the model demeaned and centered (e.g., for each covariate two terms are included, one with the mean over the study period and other with the difference from the mean in each year). *Some of the covariates are reported at the municipality level, which corresponds roughly to the local group geographical boundaries. 95% confidence intervals in brackets. Standard errors clustered at the local group level (55 local groups in all models). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Both within and between coefficients in all 5 models suggest an association between higher psychotherapy supply and lower GP prescription of antidepressants. Nevertheless, coefficients for the within and between terms differ considerably in terms of magnitude. On the one hand, the within coefficients are small[19] and mostly statistically insignificant at a 5% level, suggesting that GPs are not changing the prescription behaviour overtime by being exposed to one more psychologist/100,000 patients in their local group. On the other hand, the between term shows a significant association between working in local groups with one unit higher

[19] The interpretation of the within coefficient for model 5 would be: the increase in one psychologist FTE per 100,000 patients in the local group where is associated with a decrease of 0.28 pp (CI 95% -0.699, 0.138) in the share of patients identified with depression that are prescribed antidepressant medication. When compared to the average of 26.88% of adults with depression prescribed antidepressants, this reduction translates into a 1.04% decrease.

average number of psychologists/100,000 patients and having a lower antidepressant prescription share, by 0.955 pp (CI 95% -1.641, -0.269). When compared to the average prescription share of 26.88 %, this result indicates that GPs in local groups with a higher average number of psychologists by 1/100,000 patients have a prescription share that is lower by 3.6%.

Contrary to the within term, the results for the between coefficient are impacted by the characteristics of GPs, GP practices and local groups, as supported by the reduction of the coefficients with the subsequent inclusion of controls in models (2) and (5). Despite this reduction, fully adjusted model depicts a between coefficient that is statistically significant and larger than the within term, suggesting that even after extensively accounting for patients' list characteristics and GP/practice observables unmeasured characteristics and processes are dictating that GPs prescribe less in local groups with higher average number of psychologists.

2.5.3 Heterogeneity analyses

Table 2.3 reports the heterogeneity analysis conducted by estimating our main model stratified for quartiles of the average number of psychologists in the local group. Results confirm that the relationship of interest differs depending on the average psychological therapy supply available to GPs.

Table 2.3: Heterogeneity in the association of psychological therapy supply and antidepressants prescription rate by average number of psychologists in the local group

	Average number of psychologists in the local group				Main model
	1st Quartile (lowest) (1)	2nd Quartile (2)	3rd Quartile (3)	4th Quartile (4)	(5)
Psychologists/100,000 patients					
Within-term	-0.826***	-0.120	0.277	-1.115*	-0.305
	[-1.090,-0.562]	[-0.480,0.241]	[-0.393,0.946]	[-2.170,-0.0601]	[-0.706, 0.0954]
Between-term	-0.869	-9.906**	-1.804	-2.274**	-0.955**
	[-3.995,2.256]	[-17.40,-2.410]	[-11.00,7.395]	[-3.684,-0.864]	[-1.641, -0.269]
Year dummies	Yes	Yes	Yes	Yes	Yes
GP, GP practice and Local context * covariates	Yes	Yes	Yes	Yes	Yes Yes Yes
Constant	-19.63	19.78	161.8*	88.22	
	[-141.0,101.7]	[-97.61,137.2]	[32.14,291.4]	[-47.88,224.3]	22.45
N	4,474	4,257	4,191	4,288	[-48.56,93.46]
N Clusters	21	27	26	24	17,210
R2 within model	0.0966	0.0778	0.117	0.125	0.0756
Mean dep. variable	28.15	28.17	25.80	25.32	-0.305
Min-max average psychologists	0-1.15	1.16-1.73	1-73-2.42	2.41-7.37	0-7.37

Notes: The dependent variable is the share of adults with depression in the GP list that has been prescribed antidepressants. Covariates are included in the model demeaned and centered (e.g. for each covariate two terms are included, one with the mean over the study period and other with the difference from the mean in each year). *Some of the covariates are reported at the municipality level, which corresponds roughly to the local group geographical boundaries. 95% confidence intervals in brackets. Standard errors clustered at the local group level (55 local groups in all models). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For those in the local groups with lowest (1st quartile) and highest (4th quartile) supply, the within term is statistically significant and larger than in the main model ((5) of

Table 2.3). This means that the addition of one psychologist/100,000 patients leads GPs to reduce their prescription in 0.826 pp (95% CI -1.090, -0.562) in local groups deprived from psychologists, and in 1.115 pp (95% CI -2.170, -0.0601) in local groups with the highest supply. These are, respectively, reductions of 2.9% and 4.4% when compared to the average prescription in each subgroup (28.15% and 25.32%).[20]

We also examined whether our results varied based on GP and practice characteristics presented in the literature as determinants of prescription. Within coefficients in Figure 2.3. are mostly not statistically significant except for the analysis by GP age category, which shows that younger GPs were more likely to substitute antidepressant prescriptions with psychotherapy when the supply of psychologists increased overtime. Interestingly, the between terms are also smaller and less precise for female GPs, the eldest GPs, and GPs operating in cities and rural areas, compared to the main model. This suggests that the prescription behaviour of these groups might be less likely to be associated with unmeasured characteristics and processes driving the association between a higher number of psychologists and lower prescription. When we stratified GPs by their average prescription rate, the results were mostly non statistically significant (Table 2.4). Nevertheless, we noted two interesting patterns. First, we observed that GPs with the highest average prescription rate were more responsive to an increase in psychologists over time (3rd and 4th quartiles). Additionally, the between terms were smaller – or even positive – and non-statistically significant, suggesting that grouping GPs by their prescription share might take away some differences in unmeasured variables driving the previously observed association.

[20] We refrain from interpreting the between term in this stratified analysis because there is very little variation in the average number of psychologists in each subgroup, with exception of the 4th quintile.

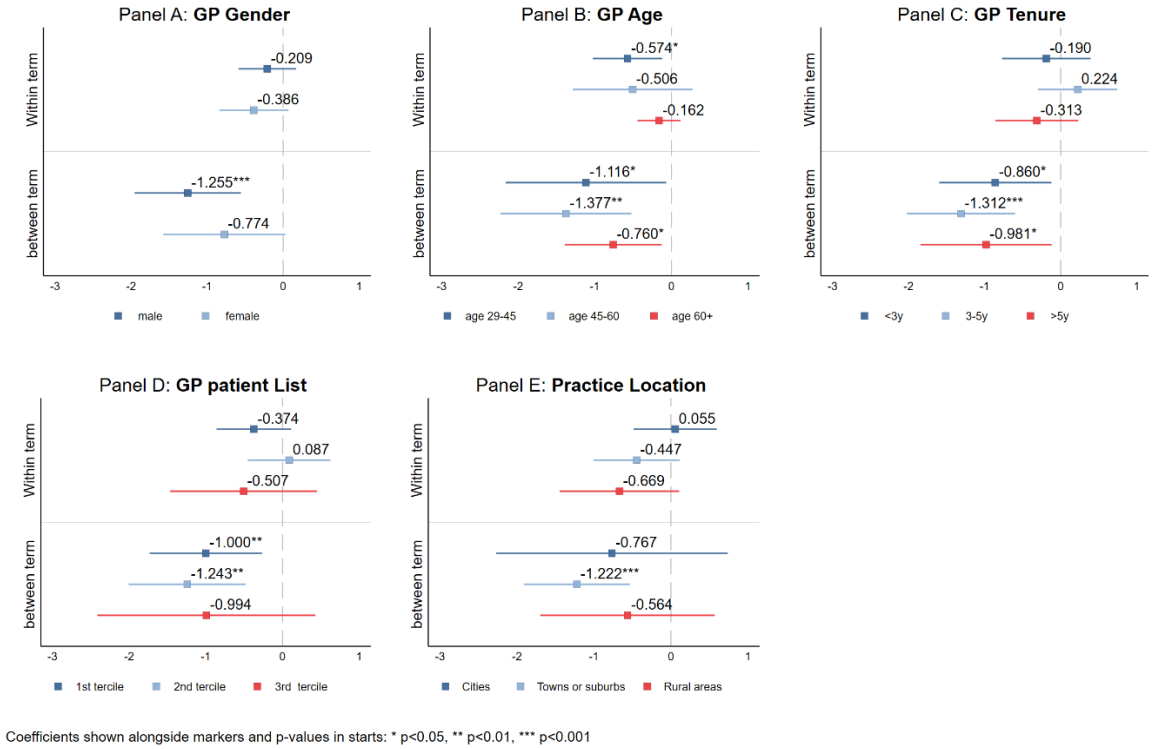


Figure 2.3 Heterogeneity in the association of psychological therapy supply and antidepressants prescription rate by GP and GP practice characteristics

Notes: Coefficients and 95% confidence intervals. The dependent variable is the share of adults with depression in the GP list that has antidepressant therapy covariates are included in the model demeaned and centered (e.g. for each covariate two terms are included, one with the mean over the study period and other with the difference from the mean in each year). Standard errors clustered at the local group level (55 local groups in all models).

Table 2.4: Heterogeneity in the association of psychological therapy supply and antidepressants prescription rate by average antidepressant prescription rate of GPs.

	Average GP antidepressant prescription rate				Main model
	1st Quartile (lowest) (1)	2nd Quartile (2)	3rd Quartile (3)	4th Quartile (4)	(5)
Psychologists/100,000 patients					
Within-term	0.0808	-0.0677	-0.622	-0.464	-0.305
	[-0.0835,0.245]	[-0.563,0.427]	[-1.277,0.0333]	[-1.516,0.589]	[-0.706, 0.0954]
Between-term	-0.0896	-0.193**	-0.0608	0.489	-0.955**
	[-0.385,0.206]	[-0.335,-0.0506]	[-0.226,0.104]	[-0.303,1.280]	[-1.641, -0.269]
Year dummies	Yes	Yes	Yes	Yes	Yes
GP, GP practice and Local group* covariates	Yes	Yes	Yes	Yes	Yes
					Yes
					Yes
Constant	18.62	13.17	39.00***	-5.108	
	[-5.275,42.51]	[-4.464,30.80]	[21.48,56.51]	[-77.27,67.06]	22.45
N	4305	4300	4305	4300	[-48.56,93.46]
N Clusters	55	55	55	55	17,210
R2 within model	0.0562	0.102	0.102	0.126	0.0756
Mean dep. variable					
Min-max average psychologists					

Notes: The dependent variable is the share of adults with depression in the GP list that has been prescribed antidepressants. Covariates are included in the model demeaned and centered (e.g. for each covariate two terms are included, one with the mean over the study period and other with the difference from the mean in each year). *Some of the covariates are reported at the municipality level, which corresponds roughly to the local group geographical boundaries. 95% confidence intervals in brackets. Standard errors clustered at the local group level (55 local groups in all models). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.5.4 Robustness analysis

To exclude the potential effect of an additional psychologist affecting the GP diagnosis capacity, we regressed the incidence rate of depression at the GP level on the local group psychologists' supply. Additional diagnosis capacity could have implications in our ability to measure changes in antidepressant prescription (numerator), because our prescription share outcome is standardized for the prevalence of depression (denominator). It could also influence the prescription patterns by changing the average severity level of the patients diagnosed with depression. The results suggest that an increased supply of mental health professionals was not associated with changes in GP capacity to diagnose depressive disorders (non-statistically significant coefficients, Table 2.5 in the appendix). Figure 2.6 supports this conclusion by

providing a visual clue of a non-systematic relationship between the incidence of depression and the number of psychologists in the local group.

Secondly, we explored whether the local groups would hire psychologists as a response to (changes in) GPs antidepressant prescription. Table 2.5 shows non-statistically significant coefficients for both 1 and 2 year-lags of prescription in psychologists supply, addressing reverse causality-type of concerns. Thirdly, the similarity between findings in the main model and the model without potentially endogenous variables (list size and utilization share) in Table 2.6 suggests that potential bias from time-varying utilization of GP visits and list size is not a salient issue. This supports our choice of a preferred specification with the full set of covariates. Last, we confirmed that our results remain the same when estimated in a balanced panel of GPs over the 4 years (Table 2.6).

2.6 Discussion

Our study aimed to investigate the relationship between the number of psychologists working in Portuguese primary care and the prescription of antidepressants by GPs between 2015 and 2018. We found no overall effect of increasing one psychologist/100,000 patients on GP prescription overtime (within term). Additional analysis uncovered heterogeneity in this finding: adding one psychologist/100,000 patients led to statistically significant reductions in prescription overtime for the GPs in the lowest and highest quartiles of average psychologist's supply. Our study also found that GPs working in local groups with one-unit higher average number of psychologists/100,000 patients have a lower antidepressant prescription share (between term). We thus conclude that on average local groups that have more psychologists have their GPs prescribing less antidepressant drugs but hiring psychologists do not necessarily lead to reduction in antidepressant prescription across all levels of psychologists' supply. The larger magnitude and the significance of the between estimate, compared with the within one, suggests a relevant role of unmeasured characteristics and processes that bring together GPs

prescribing less and local groups with higher average number of psychologists. These characteristics could be at the GP level, if physicians prone to prescribe less often antidepressants were able to self-select into local groups with higher supply of psychological support; or at the local group level, if those groups with more psychologists were also the local groups where mental health services are better integrated, and local protocols, training or clinical leadership favour the practices reducing the prescription of antidepressants.

Our study adds to the body of literature exploring the relationship between pharmacological and non-pharmacological approaches to treat depression, which is mostly composed by GP surveys and qualitative studies susceptible to limitations as self-reporting, desirability bias or real-life contradictions between GPs intentions and non-pharmacological resources available to them (Verdoux et al., 2014). We provide two main contributions to this field of research, which are relevant to several other countries that face similar challenges in access to non-pharmacological treatment approaches, such as financial barriers or long waiting lists (Verdoux et al., 2014). First, we observed that the effect of supplying additional psychological therapy resources to GPs might depend on what the baseline level of these services is, especially in a context scarcity. Adding one psychologist/100,000 patients reduced GP prescription of antidepressants by 2.9% when none or just one psychologist worked at the local group (lowest quartile); and by 4.4% for GPs working in local groups with the highest supply of psychological resources (highest quartile). No effect was found in middle quartiles. These findings suggest that (some) GPs are willing to substitute the prescription of antidepressants by referrals to psychologists when faced with the opportunity of shifting care for the first time, but that additional supply might be used based on patients' severity, possibly by prioritizing combined therapy for moderate patients and only using psychotherapy alone as a substitute again when higher severity patient needs would be covered by larger teams of psychologists. While our data does not allow us to investigate this hypothesis further, literature suggests that most GPs

still prefer to use psychotherapy in a combined treatment approach, despite their positive opinions about psychotherapy effectiveness (Dumesnil et al., 2012; Haddad et al., 2012; Verdoux et al., 2014).

A second contribution is not directly related to the effects of increasing psychotherapy in prescription behavior, but to the confounders associated with both. Our between term suggested that GPs prescribing less are also those exposed to the higher average number of psychologists/100,000 patients, for reasons we cannot account for. Confounders could be at the GP-level and patient-level, for example unobservable beliefs and personal experiences of GPs if they self-select into local groups with larger number of psychologists, or influence the decision of hiring psychologists in their local group (for example through their prescription patterns). However, both our institutional setting and robustness analyses[21] exclude GP and patient-level confounders as the main contributors behind the increased between term. GPs are civil servants hired by public tender- to state owned primary care practices. The hiring in the Portuguese national health system is highly regulated by a national recruitment program and physicians apply mostly through the national tender to open positions in existing practices. On the patient side, there are limited options to choosing or moving GPs. Patients must enrol in the practices corresponding to their catchment area, and the lack of GPs to cover the Portuguese population is one biggest weakness of NHS, widely discussed in the society. Approximately 7.5% of the Portuguese residents did not have a GP in August 2018[22]), with some regions being characterized by a greater number of patients waiting to be registered with a GP. Therefore, we hypothesize that local group level processes are most likely the confounders behind the

[21] Table A2 shows that lagged antidepressant prescription behaviour is not associated with the changes in the number of psychologists.

[22] Data retrieved from the Transparency Portal from the Portuguese NHS https://transparencia.sns.gov.pt/explore/dataset/utentes-inscritos-em-cuidados-de-saude-primarios/ (accessed in 1 May 2023).

magnitude and significance of the between term. Local groups with more psychologists might also be those where mental health services are better integrated; with local protocols and experience supporting the use of psychotherapy alone; or in which clinical leadership and education provided to GPs favours lower prescription of antidepressants. These findings should direct policymakers that aim at reducing antidepressant use to act beyond increasing psychologist's supply; and effectively understand what are the organizational factors that maximize the likelihood of substitution. In fact, there is only one national guideline on treating depression, which dates from 2012 and focuses on major depression (Direção-Geral da Saúde, 2012); each local group has their own (internal) referral pathways and guidance on what patients and cases should be referred to psychologists, which can differ considerably according to anecdotal evidence. Identifying and scaling up best practices, developing nationwide evidence-based guidelines on referral to psychotherapy in primary care and promoting awareness and educational interventions for GPs might be as important as increasing the number of psychologists.

There are important considerations to be made when interpreting our findings. First, the causal interpretation of our findings is limited, even for the within term. Despite not being affected by time-invariant variables, the within coefficient may still be biased due to time-variant unmeasured variables. It is possible that the local group-level confounders we discussed earlier may change over time, (e.g., maturity of mental health services integration or clinical leadership preferences), although this is unlikely to vary substantially over a short panel of four years. Second, as with most literature, we are unable to determine the appropriateness of either treatment approach. Antidepressant drugs are a crucial therapeutic tool in managing depression, and increases in their use are not necessarily inappropriate, but may be a consequence of increased detection, reduced stigma, improved patient acceptance, among others. The effectiveness of antidepressants, particularly serotonin re-uptake inhibitors (SSRIs), has

recently been subjected to extensive public debate (‘How to make better use of antidepressants’, 2022; Moncrieff et al., 2022). It is important to clarify that this study does not support arguments presented in discussions against the use of antidepressants.

Studying the causal effect of psychologist supply on antidepressant prescription in this context requires sophisticated research designs. While traditional randomized control trials, involving random patient assignment, may often face ethical, logistical, and financial constraints, alternative methodologies can be employed. A compelling approach is to leverage natural experiments, which can examine changes in antidepressant prescription rates over time. By contrasting regions that have witnessed an expansion in psychologist supply with areas where no such expansion occurred, a difference-in-differences design becomes feasible. It is essential to ensure that this expansion is not solely driven by changes in disease prevalence, which could also impact antidepressant prescriptions. Alternatively, instrumental variables can be used to identify variables that are correlated with psychologist supply but unrelated to patient characteristics. For instance, by considering the distance to the universities with a psychology program as an instrument, it might be possible to assess the influence of proximity to psychologists on antidepressant prescription.

Additionally, due to data limitations, the interpretation of the results is constrained, resulting in several recommendation for further research. The number of psychologists is only a proxy of actual psychological therapy supply, as psychologists may also be involved in other tasks such as palliative care or long-term care, or they might decide to prioritize some tasks such as patient assessment and counselling vs. psychotherapy, due to the number of patients in the waiting list. To validate our findings and better test the hypotheses of substitution and combined therapy, data on psychologists' consultations and GPs referrals are required. Additionally, we were unable to capture the numbers or activity of psychologists operating on a private basis outside of NHS. Portugal has many self-employed clinical psychologists, and

GPs may recommend these to patients who can afford the out-of-pocket costs. The limited availability of psychologists in the public NHS is therefore likely to contribute to inequalities in access to appropriate mental health treatment and resulting treatment outcomes, which also deserves further investigation. Last, additional research is needed to understand the role of patient beliefs and preferences in the relationship between psychologists and antidepressant use. While some literature shows that patients prefer psychotherapy to antidepressants (Van Schaik et al., 2004), other studies suggest that GPs failure to recommend non-pharmacological options is not an explanation for high antidepressant prescription rate (Read, Gibson, & Cartwright, 2016), and that simply removing barriers in access to these approaches would lead to small changes in the actual use of psychotherapy (Cronin, Forsstrom, & Papageorge, 2020).

2.7 Conclusion

Treating depressive disorders is considered a pressing issue in primary care, with pandemic-related lockdowns and economic downturns increasing the demand for mental health care and ongoing public debate about the appropriateness of antidepressants use. Our study provides an important contribution to understanding and quantifying the relationship between pharmacological and psychological treatment approaches of depression. Our results for the Portuguese primary care setting do not support the theoretical prediction that increased numbers of psychologists reduces antidepressants prescription, at least in such a context of short supply of psychologists. Instead, findings suggest that the extent to which increasing the number of psychologists will impact antidepressants prescription depends on the supply levels of psychological treatment approaches, but also on physician and organizational characteristics that bring together GPs with lower prescription rates and the local groups with higher levels of mental health integration.

2.8 References

American Psychiatric Association. 2009. "Practice Guideline for the Treatment of Patients with Major Depressive Disorder (3rd)." *http://psychiatryonline. org/guidelines. aspx*.

Anderson, I. M. et al. 2008. "Evidence-Based Guidelines for Treating Depressive Disorders with Antidepressants: A Revision of the 2000 British Association for Psychopharmacology Guidelines." *Journal of psychopharmacology (Oxford, England)* 22(4): 343–96. https://pubmed.ncbi.nlm.nih.gov/18413657/ (March 17, 2023).

Andersson, Stig J., Margareta Troein, and Gunnar Lindberg. 2001. "Conceptions of Depressive Disorder and Its Treatment among 17 Swedish GPs. A Qualitative Interview Study." *Family Practice* 18(1): 64–70. https://academic.oup.com/fampra/article/18/1/64/486076 (March 20, 2023).

Ayuso-Mateos, Jose Luis et al. 2001. "Depressive Disorders in Europe: Prevalence Figures from the ODIN Study." *British Journal of Psychiatry* 179(4): 308–16. https://www.cambridge.org/core/product/identifier/S0007125000267330/type/journal_article.

Barros, Pedro Pita. 2012. "Health Policy Reform in Tough Times: The Case of Portugal." *Health Policy* 106(1): 17–22. http://linkinghub.elsevier.com/retrieve/pii/S0168851012001133.

Bell, Andrew, Malcolm Fairbrother, and Kelvyn Jones. 2019. "Fixed and Random Effects Models: Making an Informed Choice." *Quality & Quantity* 53(2): 1051–74. http://link.springer.com/10.1007/s11135-018-0802-x.

Bell, Andrew, and Kelvyn Jones. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3(1): 133–53. https://www.cambridge.org/core/product/identifier/S2049847014000077/type/journal_article.

Bosque-Mercader, Laia, and Luigi Siciliani. 2023. "The Association between Bed Occupancy Rates and Hospital Quality in the English National Health Service." *European Journal of Health Economics* 24(2): 209–36. https://doi.org/10.1007/s10198-022-01464-8 (March 13, 2023).

Caldas, A C et al. 2020. "Sem Mais Tempo a Perder-Saúde Mental Em Portugal: Um Desafio Para a Próxima Década." *CNS. http://www. cns. minsaude. pt/wp-content/uploads/2019/12/SEM-MAIS-TEMPO-A-PERDER. pdf. Accessed* 11.

Caldas de Almeida, José Miguel et al. 2013. "Estudo Epidemiológico Nacional de Saúde Mental." *Lisboa. Faculdade de Ciências Médicas, da Universidade Nova de Lisboa* 1.

Comissão Técnica de Acompanhamento da Reforma da Saúde Mental. 2017. *Relatório Da Avaliação Do Plano Nacional de Saúde Mental 2007-2016 e Propostas Prioritárias Para a Extensão a 2020*.

Cronin, Christopher J, Matthew P Forsstrom, and Nicholas W Papageorge. 2020. *What Good Are Treatment Effects without Treatment? Mental Health and the Reluctance to Use Talk Therapy*.

Cuijpers, Pim et al. 2021. "Psychotherapies for Depression: A Network Meta-Analysis Covering Efficacy, Acceptability and Long-Term Outcomes of All Main Treatment Types." *World psychiatry : official journal of the World Psychiatric Association (WPA)* 20(2): 283–93. https://pubmed.ncbi.nlm.nih.gov/34002502/ (June 27, 2023).

Cuijpers, Pim, Clara Miguel, Mathias Harrer, Constantin Yves Plessen, Marketa Ciharova, David Ebert, et al. 2023. "Cognitive Behavior Therapy vs. Control Conditions, Other Psychotherapies, Pharmacotherapies and Combined Treatment for Depression: A Comprehensive Meta-Analysis Including 409 Trials with 52,702 Patients." *World Psychiatry* 22(1): 105–15. https://onlinelibrary.wiley.com/doi/full/10.1002/wps.21069 (June 27, 2023).

Cuijpers, Pim, Clara Miguel, Mathias Harrer, Constantin Yves Plessen, Marketa Ciharova, Davide Papola, et al. 2023. "Psychological Treatment of Depression: A Systematic Overview of a 'Meta-Analytic Research Domain.'" *Journal of affective disorders* 335: 141–51. https://pubmed.ncbi.nlm.nih.gov/37178828/ (June 27, 2023).

DGS. 2012. *Norma Nº 034/2012 Terapêutica Famacológica Da Depressão Major e Da Sua Recorrência No Adulto*. Lisbon, Portugal. https://www.dgs.pt/directrizes-da-dgs/normas-e-circulares-normativas/norma-n-0342012-de-30122012.aspx.

Direção-Geral da Saúde. 2012. *Terapêutica Farmacológica Da Depressão Major e Da Sua Recorrência No*

Adulto. https://nocs.pt/terapeutica-depressao-major.

Druss, Benjamin G. 2006. "Rising Mental Health Costs: What Are We Getting For Our Money?" *Health Affairs* 25(3): 614–22. http://www.healthaffairs.org/doi/10.1377/hlthaff.25.3.614 (April 18, 2023).

Dumesnil, Hélène et al. 2012. "General Practitioners' Choices and Their Determinants When Starting Treatment for Major Depression: A Cross Sectional, Randomized Case-Vignette Survey" ed. James G. Scott. *PLoS ONE* 7(12): e52429. https://dx.plos.org/10.1371/journal.pone.0052429 (February 20, 2023).

ERS. 2023. "Acesso a Serviços de Saúde Mental Nos Cuidados de Saúde Primários."

Estrela, Marta, Maria Teresa Herdeiro, Pedro Lopes Ferreira, and Fátima Roque. 2020. "The Use of Antidepressants, Anxiolytics, Sedatives and Hypnotics in Europe: Focusing on Mental Health Care in Portugal and Prescribing in Older Patients." *International Journal of Environmental Research and Public Health* 17(22): 8612. https://www.mdpi.com/1660-4601/17/22/8612.

Fairbrother, Malcolm. 2016. "Trust and Public Support for Environmental Protection in Diverse National Contexts." *Sociological Science* 3: 359–82. https://www.sociologicalscience.com/articles-v3-17-359/.

Fournier, Jay C. et al. 2010. "Antidepressant Drug Effects and Depression Severity." *JAMA* 303(1): 47. https://pubmed.ncbi.nlm.nih.gov/20051569/ (March 20, 2023).

Haddad, M. et al. 2012. "Clinicians' Attitudes to Depression in Europe: A Pooled Analysis of Depression Attitude Questionnaire Findings." *Family Practice* 29(2): 121–30. https://academic.oup.com/fampra/article/29/2/121/461275 (March 20, 2023).

Harkness, Elaine F., and Peter J. Bower. 2009. "On-Site Mental Health Workers Delivering Psychological Therapy and Psychosocial Interventions to Patients in Primary Care: Effects on the Professional Practice of Primary Care Providers." *Cochrane Database of Systematic Reviews* (1). https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD000532.pub2/full (September 17, 2022).

Haute Autorité de santé. 2017. *Épisode Dépressif Caractérisé de l'adulte: Prise En Charge En Premier Recours*. https://www.has-sante.fr/jcms/c_1739917/fr/episode-depressif-caracterise-de-l-adulte-prise-en-charge-en-premier-recours.

"How to Make Better Use of Antidepressants." 2022. *The Economist*: Science & technology | Drugs for depression. https://www.economist.com/science-and-technology/2022/10/19/how-to-make-better-use-of-antidepressants.

Kennedy, Sidney H. et al. 2016. "Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder." *The Canadian Journal of Psychiatry* 61(9): 540–60. http://journals.sagepub.com/doi/10.1177/0706743716659417.

Kirsch, Irving et al. 2008. "Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration." *PLoS medicine* 5(2): 0260–68. https://pubmed.ncbi.nlm.nih.gov/18303940/ (March 20, 2023).

Kjosavik, S. R., S. Hunskaar, D. Aarsland, and S. Ruths. 2011. "Initial Prescription of Antipsychotics and Antidepressants in General Practice and Specialist Care in Norway." *Acta psychiatrica Scandinavica* 123(6): 459–65. https://pubmed.ncbi.nlm.nih.gov/21401533/ (March 20, 2023).

Kovess-Masfety, Viviane et al. 2007. "Differences in Lifetime Use of Services for Mental Health Problems in Six European Countries." *Psychiatric Services* 58(2): 213–20. http://psychiatryonline.org/doi/abs/10.1176/ps.2007.58.2.213 (March 17, 2023).

Lewer, Dan, Claire O'reilly, Ramin Mojtabai, and Sara Evans-Lacko. 2022. "Antidepressant Use in 27 European Countries: Associations with Sociodemographic, Cultural and Economic Factors." https://www.cambridge.org/core. (January 13, 2022).

Van Marwijk, Harm W.J., Dick Bijl, Herman J. Adèr, and Marten De Haan. 2001. "Antidepressant Prescription for Depression in General Practice in The Netherlands." *Pharmacy world & science : PWS* 23(2): 46–49. https://pubmed.ncbi.nlm.nih.gov/11411443/ (March 20, 2023).

Mercier, Alain et al. 2011. "Understanding the Prescription of Antidepressants: A Qualitative Study among French GPs." *BMC Family Practice* 12(1): 1–9.

Mojtabai, Ramin, and Mark Olfson. 2011. "Proportion Of Antidepressants Prescribed Without A Psychiatric Diagnosis Is Growing." *Health Affairs* 30(8): 1434–42. http://www.healthaffairs.org/doi/10.1377/hlthaff.2010.1024.

Moncrieff, Joanna et al. 2022. "The Serotonin Theory of Depression: A Systematic Umbrella Review of the Evidence." *Molecular Psychiatry 2022*: 1–14. https://www.nature.com/articles/s41380-022-01661-0 (April 18, 2023).

Morrison, Jill et al. 2009. "Factors Influencing Variation in Prescribing of Antidepressants by General Practices in Scotland." *British Journal of General Practice* 59(559): 88–93.

Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46(1): 69. https://www.jstor.org/stable/1913646?origin=crossref.

Munoz-Arroyo, Rosalia, Matt Sutton, and Jill Morrison. 2006. "Exploring Potential Explanations for the Increase in Antidepressant Prescribing in Scotland Using Secondary Analyses of Routine Data." *The British Journal of General Practice* 56(527): 423. /pmc/articles/PMC1839016/ (March 20, 2023).

National Institute for Health Care Excellence. 2022. *National Clinical Practice Guideline 222. Depression in Adults: Treatment and Management*. London, UK. https://www.nice.org.uk/guidance/ng222/resources/depression-in-adults-treatment-and-management-pdf-66143832307909.

Olfson, Mark, Kurt Kroenke, Shuai Wang, and Carlos Blanco. 2014. "Trends in Office-Based Mental Health Care Provided by Psychiatrists and Primary Care Physicians." *The Journal of Clinical Psychiatry* 75(3): 21732. https://www.psychiatrist.com/jcp/medical/primary-care/trends-office-based-mental-health-care-provided-psychiatrists (February 20, 2023).

Ordem dos Psicologos. 2018. *Guia Para Implementação Dos Serviços Unidades Núcleos de Psicologia No SNS*. Lisbon, Portugal.

Ordem dos Psicólogos Portugueses. 2015. *Desenvolvimento Sustentável e Sustentabilidade Dos Cuidados de Saúde Primários*. Lisbon, Portugal. https://www.ordemdospsicologos.pt/ficheiros/documentos/opp_cuidadosdesaudeprimarios_documento.pdf.

Park, Lawrence T., and Carlos A. Zarate. 2019. "Depression in the Primary Care Setting" ed. Caren G. Solomon. *New England Journal of Medicine* 380(6): 559–68. http://www.nejm.org/doi/10.1056/NEJMcp1712493.

Plessen, Constantin Yves et al. 2023. "Exploring the Efficacy of Psychotherapies for Depression: A Multiverse Meta-Analysis." *BMJ mental health* 26(1): e300626. https://pubmed.ncbi.nlm.nih.gov/36914209/ (June 27, 2023).

Read, John, Kerry Gibson, and Claire Cartwright. 2016. "Do GPs and Psychiatrists Recommend Alternatives When Prescribing Anti-Depressants?" *Psychiatry Research* 246: 838–40. https://pubmed.ncbi.nlm.nih.gov/27825789/ (February 20, 2023).

Van Schaik, Digna J.F. et al. 2004. "Patients' Preferences in the Treatment of Depressive Disorder in Primary Care." *General Hospital Psychiatry* 26(3): 184–89.

Schneider, Frank, Martin Härter, and Susanne Schorr. 2017. *S3-Leitlinie/Nationale Versorgungsleitlinie Unipolare Depression*. Springer-Verlag. https://www.leitlinien.de/themen/depression/version-3.

Sibbald, Bonnie, Julia Addington-Hall, Douglas Brenneman, and Paul Freeling. 1996. "Investigation of Whether On-Site General Practice Counsellors Have an Impact on Psychotropic Drug Prescribing Rates and Costs." *British Journal of General Practice* 46(403): 63–67.

Verdoux, Hélène et al. 2014. "Psychotherapy for Depression in Primary Care: A Panel Survey of General Practitioners' Opinion and Prescribing Practice." *Social Psychiatry and Psychiatric Epidemiology* 49(1): 59–68. http://link.springer.com/10.1007/s00127-013-0717-8 (February 20, 2023).

Vigo, Daniel, Graham Thornicroft, and Rifat Atun. 2016. "Estimating the True Global Burden of Mental Illness." *The lancet. Psychiatry* 3(2): 171–78. https://pubmed.ncbi.nlm.nih.gov/26851330/ (May 23, 2023).

Vos, Theo et al. 2020. "Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019." *The Lancet* 396(10258): 1204–22. https://linkinghub.elsevier.com/retrieve/pii/S0140673620309259.

Wang, Philip S et al. 2011. "Treated and Untreated Prevalence of Mental Disorder Worldwide." *Oxford textbook of community mental health*: 50–66.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Young, Alexander S., Ruth Klap, Cathy D. Sherbourne, and Kenneth B. Wells. 2001. "The Quality of Care for Depressive and Anxiety Disorders in the United States." *Archives of General Psychiatry* 58(1): 55–61. https://jamanetwork.com/journals/jamapsychiatry/fullarticle/481702 (March 17, 2023).

2.9 Appendices

2.9.1 Psychology supply on diagnostic of depression and reverse causality analyses

Table 2.5: Psychology supply on diagnostic of depression and reverse causality analyses

Dependent variable: Incidence of depression (%)		Dependent variable: Variation of psychologists, overtime (within term*)		
Psychologists/100,000 patients		**Lag of antidepressant share**		
Within-term	0.133 [-0.0713,0.338]	1 year lag	0.000419 [-0.000338, 0.00118]	
Between-term	-4.67e-08 [-0.000000103, 9.51e-09]	2 year lag		0.00104 [-0.000164, 0.00224]
Year dummies	Yes		Yes	Yes
GP covariates	Yes		Yes	Yes
GP practice covariates	Yes		Yes	Yes
Local context * covariates	Yes		Yes	Yes
Constant	0.00000175 [-0.00000105, 0.00000455]		0.0598 [-1.977,2.096]	-0.717 [-5.189,3.755]
N	17,210		11459	7032
R^2 Within	0.100		0.177	0.169

Notes: The dependent variable is the share of adults with depression in the GP list that has antidepressant therapy *Covariates are included in the model demeaned and centered (e.g. for each covariate two terms are included, one with the mean over the study period and other with the difference from the mean in each year). **Some of the covariates are reported at the municipality level, which correspond roughly to the local group geographical boundaries. 95% confidence intervals in brackets. Standard errors clustered at the local group level (55 local groups in all models). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.9.2 Comparison of within-between random effects model (main) and model without GP utilization rate and patient list size.

Table 2.6: Comparison of within-between random effects model (main) and model without GP utilization rate and patient list size

	REWB Main without GP utilization rate and patient list size (1)	REWB Main (2)
Psychologists/100,000 patients		
Within-term	-0.301	-0.305
	[-0.706,0.103]	[-0.706,0.0954]
	-0.301	-0.305
Between-term	**-0.909***	**-0.955****
	[-1.620,-0.197]	**[-1.641,-0.269]**
Year dummies	Yes	Yes
GP covariates	Yes	Yes
GP practice covariates	Yes	Yes
Local context covariates**	Yes	Yes
Constant	27.89	22.45
	[-44.47,100.3]	[-48.56,93.46]
N	17210	17210
R^2 Within	0.0720	0.0756

Notes: The dependent variable is the share of adults with depression in the GP list that has antidepressant therapy *Covariates are included in the model demeaned and centered (e.g. for each covariate two terms are included, one with the mean over the study period and other with the difference from the mean in each year). **Some of the covariates are reported at the municipality level, which correspond roughly to the local group geographical boundaries. 95% confidence intervals in brackets. Standard errors clustered at the local group level (55 local groups in all models). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. REWB – within between random effects model;

2.9.3 Comparison of within between random effects model with unbalanced (main) and balanced model

Table 2.7: Comparison of within between random effects model with unbalanced (main) and balanced model

	No controls		Year dummies		GP		GP practice		Main specification	
	Unbalanced (1)	Balanced (2)	Unbalanced (3)	Balanced (4)	Unabalanced (5)	Balanced (6)	Unbalanced (7)	Balanced (8)	Unbalanced (9)	Balanced (10)
Psychologists/100,000 patients										
Within-term	-0.943*** [-1.334, -0.551]	-0.996*** [-1.455, -0.536]	-0.268 [-0.737, 0.202]	-0.305 [-0.837, 0.227]	-0.263 [-0.688,0.161]	-0.275 [-0.756, 0.205]	-0.270 [-0.688,0.148]	-0.280 [-0.752, 0.192]	-0.305 [-0.706, 0.0954]	-0.327 [-0.767, 0.113]
Between-term	-1.819*** [-2.810, -0.828]	-1.676** [-2.885,-0.468]	-1.828*** [-2.810, -0.845]	-1.676** [-2.885,-0.468]	-1.467*** [-2.132,-0.803]	-1.561*** [-2.326,-0.796]	-1.323*** [-2.014,-0.632]	-1.408** [-2.294,-0.522]	-0.955** [-1.641, -0.269]	-1.183* [-2.104,-0.262]
Year dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GP covariates	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
GP practice covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Local context covariates*	No	No	No	No	No	No	No	No	Yes	Yes
Constant	30.26*** [27.95,32.57]	28.74*** [27.95,29.53]	30.14*** [26.97,33.31]	28.36*** [27.67,29.05]	25.23 [-35.24,85.71]	225.1* [37.85,412.3]	47.97 [-16.03,112.0]	223.5* [43.71,403.2]	22.45 [-48.56,93.46]	216.6* [47.68,385.6]
N	17,210	11,244	17,210	11,244	17,210	11,244	17,210	11,244	17,210	11,244
R^2 Within	0.0128	0.0144	0.0525	0.0623	0.0697	0.0797	0.0721	0.0816	0.0756	0.0889

Notes: The dependent variable is the share of adults with depression in the GP list that has antidepressant therapy Covariates are included in the model demeaned and centered (e.g. for each covariate two terms are included, one with the mean over the study period and other with the difference from the mean in each year). *Some of the covariates are reported at the municipality level, which correspond roughly to the local group geographical boundaries. 95% confidence intervals in brackets. Standard errors clustered at the local group level (55 local groups in all models). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. REWB – within between random effects model; FE – fixed effects.

2.9.4 Diagram of the study population

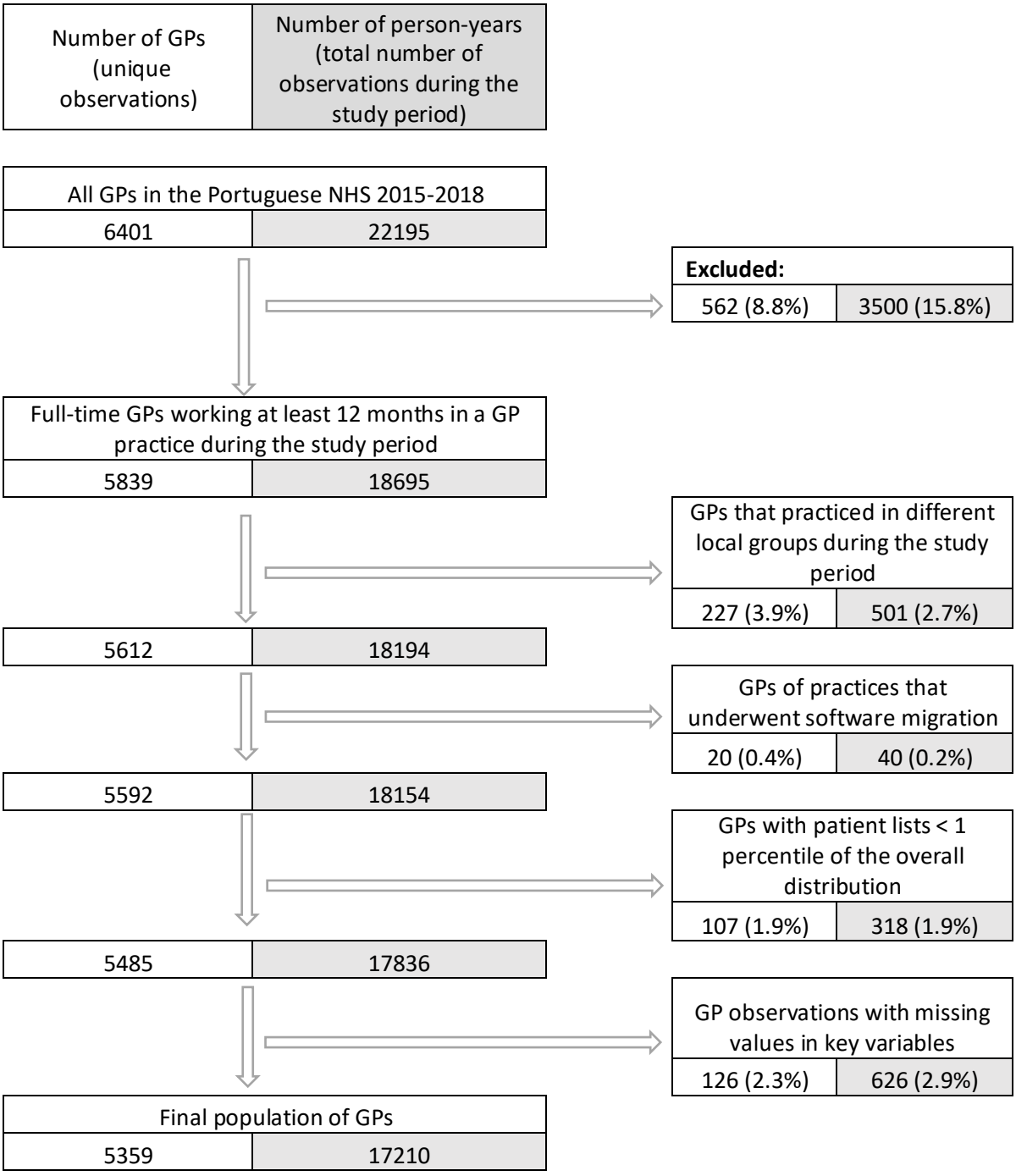


Figure 2.4: Diagram of the study population

2.9.5 Geographical distribution of the prevalence of depressive disorders diagnosis

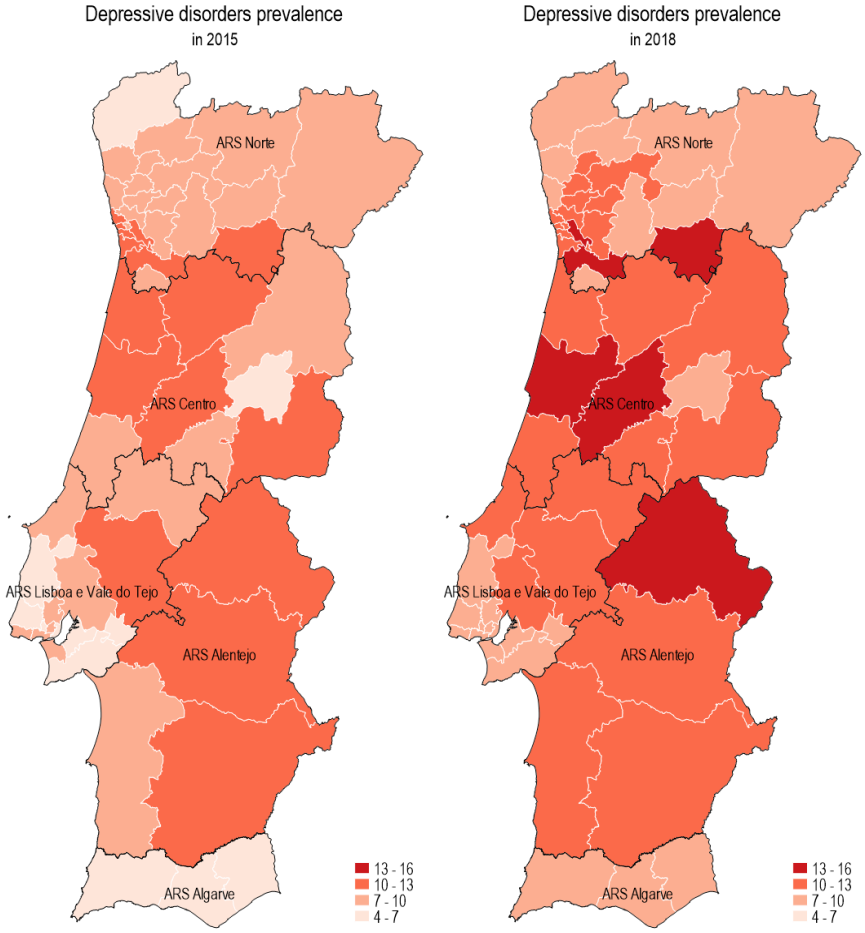


Figure 2.5: Geographical characterization of the prevalence of depressive disorders diagnosis (diagnosed according to P76 of ICPC-2)

2.9.6 Relationship between the incidence of depression in the GP patient list and number of psychologists/100,000 patients in the local group.

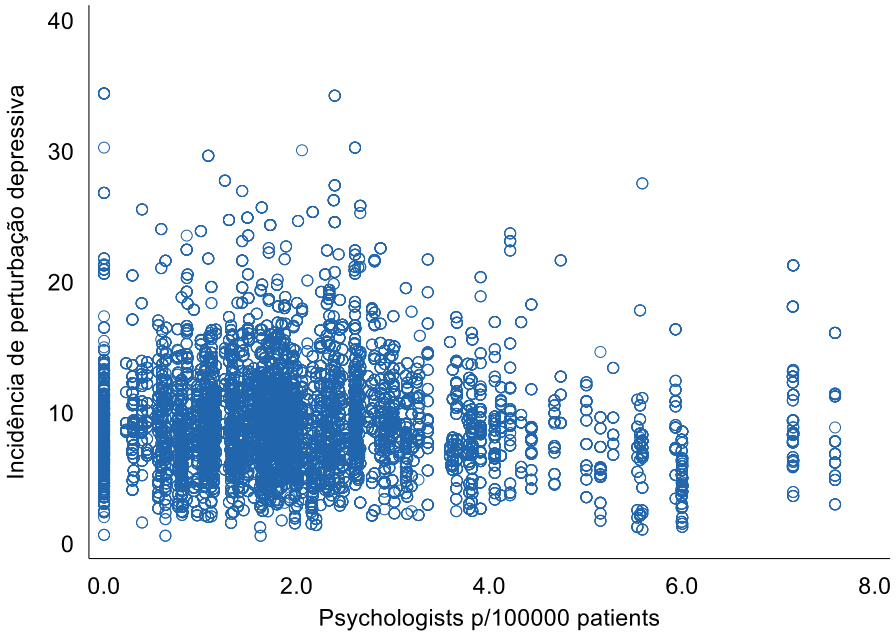


Figure 2.6: Relationship between the incidence of depression in the GP patient list and number of psychologists/100,000 patients in the local group

Chapter 3

3 Heterogeneity in physician's job preferences in a dual practice context - Evidence from a DCE[23]

Abstract

Many countries are facing challenges in recruiting and retaining physicians, particularly in regions where the public and private sectors compete for doctors. Understanding the factors influencing physicians' job choices can help inform policies aimed at attracting and retaining this valuable workforce. This study aims to elicit the strength of physicians' preferences regarding various job-related aspects, including EARNINGS, TIME FLEXIBILITY, DISCUSSION of clinical cases, frequency of FACILITIES and equipment updates, TRAINING opportunities and AUTONOMY in decision making.

To achieve this, a Discrete Choice Experiment (DCE) was administered to 697 physicians. Each participant completed a series of eight choice tasks, where they had to choose between two hypothetical jobs differing in these attributes with levels mirroring positions in the public and private sectors in Portugal. The resulting choices were analysed using mixed logit, generalized multinomial logit and latent classes models to account for diverse unobserved variations in physicians' preferences and to explore preference heterogeneity across different observable characteristics.

Jobs that offered more AUTONOMY and TRAINING opportunities were strongly preferred, as physicians would require additional compensation to work with reduced AUTONOMY

23 This chapter is co-authored with João Frutuoso, Eduardo Costa, Filipa Breia Fonseca.

(equivalent to 28.62% of gross income) or less frequent TRAINING (equivalent to 22.75%). This study also shows that the ranking of the job characteristics is similar between physicians working exclusively in the public sector and those engaged in dual practice. Nevertheless, public sector physicians place more emphasis on the availability of frequent TRAINING possibilities and frequent updates of FACILITIES and equipment compared to their counterparts in dual practice.

These findings contribute to existing knowledge by highlighting the significance of non-monetary attributes and shedding light on the preferences of physicians across various employment scenarios. They offer valuable insights for policy development aimed at influencing physicians' allocation of time between sectors.

3.1 Introduction

Shortage of health professionals is a pressing issue in many high and low-middle income countries. According to the World Health Organization (2016), a global shortfall of 15 million health workers is projected by 2030, with low-middle income countries being disproportionately affected. (World Health Organization, 2016). In Europe, as of 2020, 18 out of 30 countries reported shortages of nursing professionals, while 13 countries reported shortages of General Practitioners (GP) (McGrath, 2021). These professional shortages are situations where there are adequate skilled individuals, but an insufficient number of them are available to fill positions in the specific occupation and required location. Developed countries face challenges in meeting the growing demand for health workers due to population ageing and economic growth (Liu, Goryakin, Maeda, Bruckner, & Scheffler, 2017). Moreover, increasingly tight healthcare budgets have deteriorated working conditions for health professionals (World Health Organization, 2022). This deterioration can decrease the competitiveness of health professions relative to job alternatives and increase competition between the private and public sectors for the limited supply of qualified practitioners.

In contexts where both the public and private sectors provide healthcare services, many doctors engage in what's known as dual practice or moonlighting. This means they work in both the public and private sectors simultaneously. This practice has a long history and is common in Western European countries (Berman & Cuizon, 2004; Garattini & Padula, 2018; García-Prado & González, 2011). For example, in the United Kingdom (UK), over 60% of doctors working in public hospitals conduct private practice alongside their duties in the UK National Health Service (NHS) (McPake et al., 2016). In Portugal, while exact figures are unavailable, a recent survey (M. Ferreira, Lopes, Guimarães, & Barros, 2018) revealed that despite doctors' close ties to the NHS, where most of the doctors complete their medical residency, over 66% of surveyed physicians did not work full-time in the public sector, and more than 50% also engage

in private sector practice. Despite the prevalence of dual practice, our understanding of the factors influencing its prevalence and its consequential impact on healthcare labour supply remains limited (Socha & Bech, 2011). This knowledge gap poses challenges for policymakers striving to address labour shortages and optimize the distribution of healthcare professionals in such contexts (García-Prado, González, & Gonzalez, 2007).

To bridge this knowledge gap, this study employs a Discrete Choice Experiment (DCE). We aim to contribute to a better understanding of physicians' preferences and test hypotheses related to preference heterogeneity among healthcare professionals working in different sectors. Empirical work has increasingly relied on choice experiments, such as DCE techniques to identify the job attributes that are most valued by physicians, going beyond income to include professional, educational, and lifestyle considerations and quantify the trade-offs between them (Soekhai, de Bekker-Grob, Ellis, & Vass, 2019). Studies in high-income countries have examined mostly factors affecting healthcare worker recruitment in rural or underserved areas (e.g. Grobler et al., 2015; Scott et al., 2013; Verma et al., 2016), while less is known about the preference of physicians working in dual practice (Clark, Determann, Petrou, Moro, & de Bekker-Grob, 2014; de Bekker-Grob, Ryan, & Gerard, 2012; Scott, Holte, & Witt, 2020). This research aims to address this gap by incorporating contract variables and employing robust econometric techniques, allowing us to understand sub-group preferences and provide guidance for tailored policies.

The theoretical framework used in this research describes how physicians choose to allocate their labour time as a utility maximization decision. Physicians make decisions regarding their allocation of effort to optimize a utility function that maximizes income and leisure, within the constraints of time availability and market conditions (Reinhardt, 1972). This framework enabled researchers to explore the impact of wage increments and other financial incentives on labour-related decisions (Armour et al., 2001; McGuire, 2000b). This framework has been

extended to incorporate the distinctive characteristics of the physician's profession, including their capacity to influence demand (Newhouse, 1970). Special emphasis has been placed on incorporating additional non-monetary benefits stemming from the agency relationship of physicians into the income-leisure framework (Labelle, Stoddart, & Rice, 1994). Examples of non-monetary attributes include prestige, the intellectual content of an occupation, the types of consumers and colleagues they interact with, and the flexibility of the work schedule (Nicholson & Propper, 2011).

Physicians holding multiple jobs in both private and public sector might be different from physicians working exclusively in one sector when it comes to which job characteristics they most prefer (Socha & Bech, 2011). Evidence on the role of job characteristics on job choice in the context of dual practice has not been fully studied (Eggleston & Bir, 2006; Socha & Bech, 2011). The existing body of literature examining the determinants of moonlighting among physicians consistently highlights the significant influence of economic factors and employment conditions on their decision-making process (Culler & Bazzoli, 1985; García-Prado & González, 2011; Saether, 2005). Personal characteristics (such as gender and age) and family variables are also significantly associated with physicians entering into dual practice (Cheng, Kalb, & Scott, 2018; Johannessen & Hagen, 2014). Research conducted in developing countries, characterized by more permissive work environments, reveals that physicians engage in private or dual practice for a multifaceted array of reasons, ranging from seeking a more independent and flexible work arrangement to pursuing additional income, leveraging opportunities for professional influence, or simply capitalizing on available time (García-Prado & González, 2011; Russo, de Sousa, Sidat, Ferrinho, & Dussault, 2014; Russo, McPake, Fronteira, & Ferrinho, 2014). A study for Australia, a country with public service and private practices, has shown that overall the preference of physicians for working one extra hour in the public sector was weak, and heterogeneous, revealing that the preference for private sector jobs

was stronger for more risk-averse doctors and those enjoying higher salaries (Scott et al., 2020). In Norway, the engagement in dual practice by public hospital physicians is associated with economic factors (such as debt, interest payments, and income expectations) followed by family variables (such as cohabiting status and the number of children) (Johannessen & Hagen, 2014) though it depended according to the gender and medical specialty. According to a recent literature review, the most common reason for working in the private sector in high-income countries was low government salaries, poor working conditions and high workload in public hospitals, while the most frequent reasons for working in the public sector were recruitment of patients for private-practice, access to public resources followed by social responsibility (Hoogland et al., 2022). It is important to note that attraction and retention strategies must address the underlying factors that make employment unattractive from monetary to non-monetary factors. Public policies that solely relied on financial incentives to attract physicians, for instance to rural areas or to prevent emigration, have often yielded underwhelming results (Brugha, Clarke, Hendrick, & Sweeney, 2021; Hamouzadeh, Akbarisari, Olyaeeemanesh, Yekaninejad, et al., 2019; Kumar & Clancy, 2021).

3.2 Setting

Universal access to healthcare in Portugal is ensured through the NHS, supplemented by health insurance schemes for specific professions (e.g., public servants, military, policemen) and private voluntary health insurance. The NHS is responsible for primary and hospital care, while dental care, diagnostics, and rehabilitation are often provided by the private sector with public funding support. Social care, including long-term care and palliative care, is provided through a national network that involves public, for-profit, and not-for-profit sectors (please consult Simões et al., (2017) for more details).

Physicians and other healthcare professionals in public health services are hired mostly through public hiring procedures with fixed wages as statutory public servants. The fixed salary is established according to a salary matrix linking professional category and duration of service and is independent of any productivity measure (Barros & Simões, 2007). Hospital workers receive variable income for their overtime work. Physicians can choose from three arrangements in the public sector: full-time (not exclusive) (35 hours/week), extended full-time (exclusive NHS) (42 hours/week), and part-time (not allowed for a head of service). Most of them chose either the full-time or extended full-time schemes (Barros & Simões, 2007). In the private market, healthcare professionals act as independent providers and/or as employees of healthcare institutions. Physicians without exclusive contracts in either the public or private sector may combine an NHS contract with part-time work in the private sector (Barros & Costa, 2022).

However, despite the higher number of physicians per 1,000 population in Portugal (5.3) compared to the OECD average (3.6), the NHS has faced a shortage of healthcare professionals (OECD, 2021). In 2011, 58% of 43,247 registered physicians were working for the NHS in mainland Portugal (Santana, Peixoto, & Duarte, 2014). Since then, the increase in the number of physicians has been offset by various factors – the growth of the private sector; public budget constraints affecting income and working conditions (E. Costa, Santos, & Barros, 2021); competition from other European countries (P. L. Ferreira, Raposo, Tavares, & Correia, 2020); the ageing workforce (in 2016, more than 50% of physicians in Portugal had more than 55 years) (D. Lopes, Castro, & Simões, 2018) among others.

This shortage of health professionals is particularly evident in primary healthcare, where the number of patients lacking a GP reached 1.5 million in January 2023 (15 per cent of the population). Despite efforts by the NHS to attract GPs through various policies, such as performance-based incentives, exclusive employment status, and incentives for rural and

deprived areas, the effectiveness of these measures remains uncertain (T. Correia, Gomes, Nunes, & Dussault, 2020; Ministério da Saúde, 2015). Moreover, according to the OECD, physicians in Portugal experienced a 21% decrease in real remuneration in euros between 2010 and 2021 (OECD, 2023). These conditions make the private sector and emigration more appealing options.

3.3 Methods

3.3.1 Choice experiment design

Discrete choice experiments have been widely employed in health economics research to assess physicians' preferences (Clark et al., 2014; de Bekker-Grob et al., 2012; Mandeville, Lagarde, & Hanson, 2014). Following best practices (Coast et al., 2012; Coast & Horrocks, 2007; Ryan, Gerard, & Amaya-Amaya, 2008) the DCE parameters were derived from both theoretical and empirical literature, as well as from physician's opinions, using qualitative methods as outlined below. Firstly, we identified the attributes of interest and their levels through a comprehensive literature review and semi-structured interviews with physicians. Secondly, we used statistical methods to construct choice sets based on the insights gained from the interviews. Lastly, to ensure the face and content validity of the questionnaire, a pre-test was conducted with a group of volunteer physicians (n=28) to validate the phrasing and effort. Following this step, a pilot study was conducted with other volunteers (n=32) to update the choice sets. The pilot estimated parameters were used as prior information for generating the final choice sets (Carlsson & Martinsson, 2003; Crabbe & Vandebroek, 2012). Observations from the pilot were not included in the final analysis.

The most relevant job attributes and levels were identified initially thorough a review of the empirical literature research on health workforce choice experiments conducted in high-income

countries (Ammi and Peyron, 2016; Holte et al., 2015; Scott, 2001; Witt, 2017). To further refine our attribute selection and adapt to the national reality, we conducted 17 face-to-face semi-structured in-depth interviews with a diverse group of physicians representing various practice settings (details in appendix section 3.8.4). Finally, we conducted a comprehensive analysis of the policy relevance of these attributes and levels to arrive at our final parsimonious selection, providing a well-rounded foundation for our study. The six attributes selected, and their levels are described in Table 3.1.

Table 3.1: Attributes, levels, description, and expected sign of the regression parameters, E(sign).

Attributes	Attribute levels	Abbreviation E(sign)	Description
Earnings	Base, Base +20%, Base + 40%	EARN $\beta1 > 0$	Physician's remuneration level, alternatively a 20% increase or a 40% top-up
Time Flexibility	Fixed, Personalized, By objectives	TIME $\beta2 > 0$	The opportunities to adapt working times, either having a fixed number of hours, a personalized number of hours, or objectives-oriented work
Discussion of clinical cases	Remote, In-person	DISCUS $\beta3 > 0$	Clinical cases can be discussed with peers and other specialties either remotely or in person
Facilities and equipment update	Infrequent, Frequent	EQUIP $\beta4 > 0$	The frequency of updates to the facilities and equipment used, which can be either frequent or infrequent
Training possibilities	Infrequent, Frequent	TRAIN $\beta5 > 0$	The possibility for training in different fields and balancing clinical practice with academic work and research, which can occur either frequently or infrequently
Autonomy in decision-making	Low, High	AUTON $\beta6 > 0$	The degree of autonomy in decision-making capacity, including the paperwork necessary for decision-making within the institution

The first attribute, "earnings", describes relative increases (+20%, +40%) to the fixed payment according to the professional category and career stage. The 20% and 40% increases in income are realistic and aligned with a policy aimed at attracting physicians to underserved areas,

where a 40% wage increase is offered to those who relocate (Ministério da Saúde, 2015). In DCE it is appropriate to measure preferences relative to changes in the existing reference point (i.e. physician's current income) (Holte, Sivey, Abelsen, & Olsen, 2016; Tversky & Kahneman, 1991). The second attribute "Time Flexibility" studied working schedules. We considered the regular fixed weekly schedules. The other possible levels were "flexible schedules" that allow for the negotiation of different hours per week according to individual preferences; and "by objective" in which physicians can set goals they must perform independently of work hours. The third attribute focused on the update of "facilities and equipment", with choices between outdated or frequently renovated or replaced materials. The fourth attribute, "training possibilities", studied the acquisition of technical knowledge within and outside the medical field, that could occur more frequently or less frequently. The last attribute, "autonomy", analyses the level of autonomy that physicians have in their job and whether they must navigate complex bureaucratic processes to practice.

The survey consisted of two parts: a discrete choice experiment with 8 choice sets (the 16 choice sets were split into two blocks, and physicians were randomly allocated to answer one block), and a questionnaire with characterization questions related to the physician's job situation, intention to leave their job, and other personal characteristics (gender, age, speciality).

Each choice set displayed two unlabelled job alternatives described by six attributes. Respondents were asked to choose their preferred job and were informed that the job offers were similar in all aspects except for the attributes displayed (see appendix 3.8.1 for an example choice task, as presented to the respondents).

3.3.2 Choice sets

The experimental design followed a two-stage procedure. First, the attributes and levels were organised in scenarios using a factorial experimental design, which consists of all possible combinations of the levels of the attributes and allows for the estimation of main effects and interactions, rendering 10,368 different possible choice tasks. From the set of possible scenarios, 16 were generated using a fractional factorial design (D-efficient design performed in Stata 17) and split randomly into 2 blocks. This method provides the minimum number needed to maintain an orthogonal design. The initial parameter information was based on insights gained from the interviews, assuming a uniform parameter distribution. These priors were later updated using the results from the pilot.

3.3.3 Pre-test and pilot study

To assess the wording and interpretation of the attributes we run an initial pre-test. Physicians (n=16) were asked to review the instructions and choice sets (see appendix section 3.8.5). They reported their understanding of the attributes, and any doubts were noted. After revising the text, 12 other clinicians were asked to read and report once more, until only minimal changes were suggested.

Finally, the DCE was piloted by 32 physicians in May 2021. The DCE administered in the pilot had eight choice sets created using a fractional factorial design. The pilot results were used to set the priors used in the final efficient design, using a D-efficient design. Results from the pilot were used as priors in the final efficient design. The values of the new priors were slightly higher than our initial priors, but the order of the relative importance of the attributes was kept. The final choice sets focused only on the estimation of the main effects. Additional quality

control measures such as dominance checks or repeat tasks were not included due to the length of the survey and the risk of participants dropping out.

3.3.4 Data

An online survey was conducted between June 2022 and September 2022. The survey was shared with physicians in Portugal, regardless of their speciality or employment status. The survey was distributed using multiple channels, including social networks and institutional mailing lists from the major medical doctors' union. The online survey was administrated with Qualtrics software.

Before starting the DCE, respondents were informed of the scope of the survey and required to provide their informed consent. Only physicians were permitted to proceed with the questionnaire. experimental procedure was explained, emphasizing that respondents were tasked with selecting alternatives from hypothetical scenarios, and simulating real-life choices. Respondents were given the option to quit the experiment at any time. The study was pre-registered and ethics approval was granted before data collection.

3.3.5 Analytical approach

We use a DCE to evaluate physician job-stated preferences, explore heterogeneity in these preferences, and estimate willingness to accept (WTA) values for different job attributes.

The utility from doctor *n* is derived from choosing alternative *j* in choice situation *t* is given by a deterministic utility (V_{njt}) and an unknown random component ε_{njt}, and is specified as:

$$U_{njt} = V_{njt}(X_{njt}, \beta) + \varepsilon_{njt}; n = 1, \ldots, N; j = A, B; t = 1, \ldots, 8 \quad (7)$$

We assume that the utility derived from a given job alternative is determined by a linear combination of six attributes described above. Hence, a linear addictive utility function was estimated as follows:

$$V_{njt} = \beta_1 EARN_{njt} + \beta_2 \text{TIME}_{\text{personalized}_{\text{njt}}} + \beta_3 \text{TIME}_{\text{by objectives}_{\text{njt}}} + \beta_4 DISCUSS_{\text{njt}} + \beta_5 EQUIP_{\text{njt}} + \beta_6 TRAIN_{\text{njt}} + \beta_7 AUTON_{\text{njt}} \quad (8)$$

The outcome variable is binary and shows whether physician *n* chose job alternative *j* in choice *t*. Since this is an unlabelled choice design, there is no reason to believe the respondents treat Job A and Job B differently. Constant terms were not included in our models because participants were not required to choose against a constant scenario. The models were estimated using 2,000 replications or Halton draws, which provide sufficient simulation accuracy (Czajkowski & Budziński, 2019).

Several estimation methods were used to calculate the likelihood of selecting a job in the series of choices by physicians. The initial model estimated (Model 1) used the standard conditional or Multinomial Logit Model (MNL) widely employed in this research field (Merlo, van Driel, & Hall, 2020). However, MNL models are more restrictive on the stochastic terms and assume that all participants have the same preferences for a given attribute (Hensher, Rose, & Greene, 2015). To explore potential heterogeneity in physician preferences and assess the role of scale heterogeneity in our sample, we examined different models.

In Models 2 and 3 we employed mixed logit models (MXL), known for their enhanced flexibility in approximating diverse random utility models (McFadden & Train, 2000). By incorporating a stochastic element into the model attributes to account for unobserved physician preference heterogeneity (Hensher & Greene, 2003), we could improve the model's fit compared to the standard logit model (Hole, 2008). In model 2 we estimate a MXL model with all attributes treated as random and independently normally distributed. In model 3, we

treated the EARNINGS attribute as fixed. This approach enabled us to later estimate the monetary value of the marginal willingness to accept compensations for other non-monetary attributes.

To better understand the meaning of the utility weights estimated for each attribute, we examined the approximate thresholds that establish the willingness to trade non-monetary attributes for a change in earnings. These are calculated as the Marginal Willingness To Accept (MWTA) lower compensation in exchange for something else (de Bekker-Grob et al., 2012). The estimation of the MWTA for a non-monetary attribute, *x*, is calculated as follows:

$$MWTA(X) = \frac{\partial U/\partial x}{\partial U/\partial EARN} = \frac{\beta_{x_n}}{\beta_{EARN}} \quad (9)$$

A negative MWTA indicates that physicians are willing to forgo some of their earnings to upgrade from the reference level of an attribute to a higher level (e.g., transitioning from infrequent to frequent training).

We assessed heterogeneity in respondents' preferences through two distinct approaches.

Firstly, we examined the presence of scale heterogeneity, that is, whether the idiosyncratic error term exhibits a greater scale for some participants than others indicating variations in choice behaviour due to factors like the seriousness with which choices are made or the strength of expressed preferences (see for e.g. Louviere et al. (1999)). To measure the level of heterogeneity present among respondents, we estimated a Generalized Multinomial Logit model (G-MNL) (Fiebig, Keane, Louviere, & Wasi, 2010). Additionally, we employed Latent Class Modelling (LCM) to capture preference heterogeneity among unobserved subgroups of the population, confirming whether distinct preferences existed across these homogenous latent groups. Based on each respondent's probability of belonging to a class, it was possible to provide insight into the classes' membership based on the physicians' characteristics.

Secondly, to account for heterogeneity that stems from observable differences (i.e., explained heterogeneity) we introduced interactions between each attribute variable and the observable characteristics of the respondent in the model 2 (MXL1) specifications. This approach allowed us to explore the hypothesis that preferences may differ based on the sector in which physicians work, specifically whether they are exclusively employed in the public sector or engaged in dual practice. These hypotheses were informed by earlier studies (Ramos, Alves, Guimarães, & Ferreira, 2017; Scott et al., 2020) and insights gathered from semi-structured interviews with Portuguese doctors.

Lastly, another potential source of preference heterogeneity can be driven by differences across specialities. It is well established that the personal preferences of the students in terms of their future working life influence their choice of speciality (see for e.g. (Levaillant, Levaillant, Lerolle, Vallet, & Hamel-Broza, 2020; Ramos et al., 2017). Moreover, the scope for supplying hours in the private sector may vary widely according to a physician's speciality and experience (Carlsen, Hole, Kolstad, & Norheim, 2012). For example, GPs may be generally less likely to engage with the private sector compared to certain specialists like dermatologists or surgeons. A study from 2004, indicated that 50% of the hospital physicians in Portugal had a second job, compared with 23% of public sector workers in primary care (Ferrinho, Van Lerberghe, Fronteira, Hipólito, & Biscaia, 2004). Therefore, we conducted a separate heterogeneity analysis aimed at investigating potential variations in preference structures across different medical specialties.

To estimate the models, we used STATA software version 17 (StataCorp., 2021).

3.4 Results

There were 697 physicians who completed the survey, yielding 11,152 choice responses. Evaluating the sample's representativeness is complicated due to discrepancies in official data from Portugal. The data accounts for all doctors licensed to practice, which overestimates the actual number of practicing doctors by approximately 30%, (OECD, 2021). Participant characteristics are presented in Table 3.2. About one in four respondents were male (27.6%) and most were aged 31-55 (71.7%) According to the WHO Human Resources for Health profile, 36.4% of physicians are male, with a majority of them (71.3%) falling within the age range of 25 to 54 (World Health Organization, 2022). General practitioners account for 40.7% of the respondents, specialists without surgical activities for 42.2%, specialists with surgical activities for 8.5%, and 8.6% were other specialities (e.g., public health). Almost half of the physicians worked exclusively in the public sector (47.5%), while the other half were dual practitioners (42.6%) (40.5% had the NHS as their main employer, while only 2.15% of physicians had a private sector institution as their main employer).

Table 3.2. Characteristics of the physicians

Characteristics of physicians in the sample (n=697)	Absolute value	Percentage
Gender		
Female	500	71.74
Male	192	27.55
Other	5	0.72
Age		
<30	135	19.37
31-55	503	72.17
>56	59	8.46
Employee		
Only in the private sector	56	8.03
Only in the public sector	331	47.49
Not employed	8	1.15
Not working in Health care	5	0.72
Dual - mostly in the private sector	15	2.15
Dual - Mostly in the public sector	282	40.46
Speciality		
Primary Care - General Practice	284	40.75
Hospital Care - Medical	294	42.18
Hospital Care - Surgical	59	8.46
Other	60	8.61

In Table 3.2, we present the main results for four models: multinomial logit (MNL), mixed logit (MXL1, MXL2), and generalized multinomial logit (GMNL). In MXL1 we consider all coefficients, including earnings, as random and normally distributed, thereby accounting for potential preference heterogeneity. In MXL2, the earnings coefficient is treated as fixed while treating the coefficients for the other attributes as random, enabling the calculation of the marginal rate of substitution in the subsequent section.

The goodness of fit measures (BIC, AIC, and log-likelihood), consistently demonstrate that the MXL models outperform both the MNL and GMNL models. Notably, Model MXL1, where the earnings coefficient is allowed to vary randomly, provides the best fit. However, the estimated scale parameter in the GMNL model is significant, suggesting the presence of scale heterogeneity within the dataset. Nevertheless, the signs, significance, and magnitudes of all random coefficients remain stable across all four models, indicating the robustness of the results.

Table 3.3. Regression results from preference estimation using multinomial, mixed logit and generalized mixed logit models.

	(1)	(2)		(3)		(4)	
	MNL	MXL1		MXL2		GMNL	
	Mean (SE)	Mean (SE)	Std. Dev. (SE)	Mean (SE)	Std. Dev. (SE)	Mean (SE)	Std. Dev. (SE)
Earnings	0.739***	1.672***	1.071***	1.220***		1.528***	
	(0.0306)	(0.123)	(0.0993)	(0.0626)		(0.223)	
Time flexibility - personalized	0.594***	1.546***	0.971***	1.159***	1.002***	1.255***	0.873***
	(0.0553)	(0.144)	(0.211)	(0.0965)	(0.141)	(0.202)	(0.244)
Time flexibility - objectives	0.465***	1.086***	1.257***	0.835***	0.811***	0.984***	0.906***
	(0.0497)	(0.117)	(0.159)	(0.0812)	(0.123)	(0.173)	(0.201)
Facilities and equipment	0.453***	1.006***	1.008***	0.764***	0.697***	0.939***	0.674***
	(0.0392)	(0.0959)	(0.127)	(0.0647)	(0.0950)	(0.151)	(0.205)
Discuss clinical cases	0.0762*	0.145*	0.960***	0.134*	0.532***	0.0464	0.595***
	(0.0325)	(0.0710)	(0.133)	(0.0521)	(0.109)	(0.0665)	(0.172)
Training	0.881***	1.797***	1.382***	1.352***	0.956***	1.744***	1.288***
	(0.0404)	(0.139)	(0.142)	(0.0873)	(0.0994)	(0.252)	(0.225)
Autonomy	1.004***	2.277***	1.707***	1.701***	1.378***	2.021***	1.827***
	(0.0519)	(0.172)	(0.156)	(0.106)	(0.104)	(0.305)	(0.287)
N Observations	11152	11152		11152		11152	
N Physicians	697	697		697		697	
Scale parameter (τ)						0.764***	
						(0.136)	
Log-likelihood	-3092.5	-2872.7		-2898.3		-2941.9	
AIC	6125.4	5667.5		5822.7		5911.8	
BIC	6176.6	5769.9		5917.8		6014.3	

Note: Standard errors are shown in parentheses.
Significance levels denoted as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
All estimates are relative to the reference level for each attribute, as follows: "base" for Earnings, "fixed schedule" for Time flexibility, "remote" for the Discussion of clinical cases, "frequent update" for Facilities and equipment, "frequent" for Training, and "low" for Autonomy.
Models estimated using 2000 Halton draws.

Based on these findings, the preferred model, MXL1, is interpreted as follows: all attributes' mean coefficients were statistically significantly different from zero and their signs were positive, as expected. Overall, physicians were more likely to choose alternatives granting more autonomy and training possibilities, followed by earnings.

The highest utility gain comes from increasing AUTONOMY from low to high. The utility gain of having frequent TRAINING possibilities instead of infrequent is approximately equivalent to the utility gain from increasing EARNINGS from base salary to a salary topped

with a 20% and 40% increase and nearly twice the utility gain from increasing the frequency of updating the FACILITIES and equipment which are used in healthcare provision from an infrequent to a frequent update. Unsurprisingly, the coefficient associated with the TIME FLEXIBILITY was positive and larger for personalised time, followed by objectives-oriented work compared with working with fixed schedules. This implies, ceteris paribus, that the utility associated with changing from a fixed schedule to the personalization of working hours is greater than the utility of changing from a fixed schedule to working by objectives. The possibility of DISCUSSing clinical cases and better FACILITIES and equipment had the lowest impact on physicians' choices.

The magnitude of the estimated standard deviation of each coefficient showed a considerable and significant amount of preference heterogeneity for each attribute, in particular autonomy and training.

The estimates of the marginal willingness to accept compensations are derived from Model 3 and are presented in Table 3.4. The MWTA were converted into monetary amounts in euros based on the net payment to physicians working in the public sector[24].

[24] We used the official gross salary for physicians entry level of their careers in the National Health Service, which is based on a 40-hour contract (gross income 2.746,24 €), the most common contractual arrangement. This gross salary translates to a net income ranging from 1841.09 € for unmarried physicians to 1,874.09 € for married physicians with two children. Available at: https://www.acss.min-saude.pt/wp-content/uploads/2016/09/Tabela-remuneratoria_carreira-medica.pdf. The gross salary was converted to the net salary considering current taxation and social security contributions in Portugal.

Table 3.4. Marginal willingness to accept compensation estimates.

Attribute (Levels)	Marginal WTA	95% CI	%	Monetary value in €	Hours worked
Time flexibility (Personalized vs Fixed)	-0.950	[-1.08,-0.82]	19.50	359.01	34
Time flexibility (Objectives vs Fixed)	-0.684	[-0.80,-0.579	14.04	258.49	24
Facilities and Equipment update (Frequent vs Infrequent)	-0.626	[-0.71,-0.54]	12.85	236.58	22
Discuss clinical cases (In person vs Remote)	-0.109	[-0.19,-0.03]	2.24	41.24	4
Training (Frequent vs Infrequent)	-1.108	[-1.22,-1.00]	22.75	418.85	39
Autonomy (High vs Low)	-1.394	[-1.52,-1.27]	28.62	526.92	50

Note: MWTA is based on equation 3 following model 3 specifications.
The variable EARN is considered continuous. The confidence intervals are estimated by the delta method. The Marginal WTA is calculated by the ratio between the coefficients of the attribute and the earnings attribute and can be interpreted as a monetary value the physician is willing to accept for the utility decrease associated with the attribute.
Monetary value calculated at the mean of net monthly income 1841.09€. The corresponding number of hours is calculated from a Net hourly wage of 10.62 €/hour.
The percentage represents the proportion of the MWTA for each attribute.

The negative sign of these estimates means that physicians are willing to trade a portion of their income for other valuable benefits. Physicians would require an extra 28.62% of their monthly net income (equivalent to €526.92) as compensation for working in a role with limited autonomy, indicating that autonomy is more important than other characteristics. Alternatively, this could be interpreted as committing to an extra 50 hours of work. Physicians are also willing to accept a reduction of 23% in their income (€ 418.85) in return for opportunities for frequent training. The coefficient for time flexibility shows that physicians are willing to forego € 359.01 or €258.49 from their earnings when transitioning from a fixed schedule to a personalised schedule or objectives-based work, respectively. When comparing the exchange between facilities/equipment and earnings, the marginal rate of substitution indicates that physicians are willing to sacrifice 12.85% of their monthly income, to approximately 236.58 €, for an increase in the frequency of facility renovation, or work an additional 22 hours based on public sector remuneration.

3.4.1 Are these preferences different whether professionals work in the public sector only?

In this section, we explore the heterogeneity among physicians, considering their distinct work arrangements. Specifically, we examine whether there are differences in the preferences of those who are exclusively employed in the public sector or engaged in dual practice. To investigate the heterogeneity of preferences we used the mixed logit model, enriched with interaction terms related to physicians' contracts (Public = 1 if physicians work exclusively in the public sector). This approach allows us to uncover variations in preferences among different physician groups, offering valuable insights into how these preferences may diverge based on their current work arrangements.

For this analysis, we exclusively focus on the group of physicians working in the public sector and those who engaged in dual practice. This specific comparison provides meaningful insights into the variations we aim to explore. It's important to note that our ability to draw sustained results for the group of physicians exclusively allocating their time to private sector jobs is limited by the relatively small number of observations (N=56).

Table 3.5: Results of the mixed logit model with single interactions presenting preference heterogeneity according to the physicians' contractual circumstances.

	MXL - Public vs. Dual			
	Mean	(SE)	Std. Dev.	(SE)
Earnings	1.267***	(0.0848)		
Time flexibility - personalized	1.158***	(0.138)	0.920***	(0.182)
Time flexibility - objectives	1.011***	(0.115)	-0.427	(0.239)
Facilities and equipment	0.669***	(0.0834)	0.471**	(0.156)
Discuss clinical cases	0.123	(0.0766)	-0.421**	(0.157)
Training	1.269***	(0.116)	0.867***	(0.142)
Autonomy	1.780***	(0.152)	1.351***	(0.133)
Earnings × Public	0.288	(0.150)	1.018***	(0.117)
TF- personalized × Public	0.264	(0.226)	0.553	(0.364)
TF- objectives × Public	-0.311	(0.183)	1.329***	(0.237)
Facilit. and equip. × Public	0.393**	(0.151)	0.952***	(0.187)
Discuss Clin. Cases × Public	0.0625	(0.125)	0.747***	(0.190)
Training × Public	0.580**	(0.206)	1.023***	(0.243)
Autonomy × Public	0.365	(0.235)	1.161***	(0.294)
N Observations	10048			
N Physicians	628			
Log-likelihood	-2551.4			
AIC	5156.7			
BIC	5351.5			

Note: The estimation includes only physicians holding posts in both public and private practice or working exclusively in the public sector. Standard errors are shown in parentheses. Significance levels denoted as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All estimates are relative to the reference level for each attribute, as follows: "base" for Earnings, "fixed schedule" for Time flexibility, "remote" for the Discussion of clinical cases, "frequent update" for Facilities and equipment, "frequent" for Training, and "low" for Autonomy. Models estimated using 2000 Halton draws.

The results in Table 3.5 present no substantial differences in how physicians in the public sector and those in dual practice rank the job attributes presented. However, there are subtle nuances in the strength of these preferences. Public sector physicians appear to place heightened importance on the prospects of more frequent TRAINING and access to updated FACILITIES and equipment. This could be attributed to the ample training opportunities and advanced

equipment typically available in the public sector, facilitating their handling of complex medical cases.

The findings derived from the latent class models, which are detailed in appendix 3.8.2, provide additional support for these findings. These latent classes reveal the presence of distinct groups with preferences closely aligned with the sector in which physicians are employed. Specifically, public sector physicians appear to exhibit a stronger inclination towards valuing TRAINING opportunities, especially at early career stages (Class 1), and place a higher premium on having increased AUTONOMY, as well as access to frequently updated FACILITIES and equipment, particularly as they advance in their careers (Class 2). In contrast, dual public-private jobholders are more inclined to prioritize job opportunities that offer higher EARNINGS and more flexible TIME as important components for their choices (Class 3). The preferences of this last group appear to be aligned with those who choose family medicine or general practice as their specialty.

Finally, we conducted a separate analysis to investigate preferences within specific medical disciplines. The results in Table 3.8, available in the appendix 3.8.3, display the estimations with interactions for GPs versus hospital specialists. Among the physicians in our sample, GPs tend to place a higher value on EARNINGS and the frequency of FACILITIES and equipment updates compared to their counterparts in other hospital specialties. This observation aligns with our initial concern that an overrepresentation of GPs in the sample of public workers could influence the preference coefficients estimated for this group. This difference in preferences may be attributed to the growing recognition of family practice as a well-established specialty. GPs have the opportunity to work in public units organized into multi-professional teams known as "Family Health Units." In these units, they can earn pay-for-performance financial incentives based on their performance, which serves as a motivating factor for selecting this

specialty. These organizational structures often offer better remuneration and career prospects compared to non-surgical hospital specialties (Ramos et al., 2017).

3.5 Discussion

This study examines the preferences of physicians working in Portugal. The results demonstrate that autonomy emerged as the most valued attribute, while on-site clinical case discussions were barely considered. In addition to autonomy, physicians valued highly frequent training, earnings increases and more flexible schedules.

In terms of policy implications, our analysis of preference heterogeneity highlights the varying importance of job attributes across different sectors, focusing on two distinct groups — those exclusively working in the public sector and those in dual private-public job participation. Interestingly, these two groups share similar preferences for job attributes, with some nuanced differences. Notably, public sector workers place a higher value on training opportunities and the frequency of facilities and equipment updates compared to their counterparts that supply some hours in the private sector. Given that physicians exclusively employed in the public sector highly prioritize training, modern equipment, and facilities, implementing initiatives aimed at enhancing their professional development and providing access to up-to-date facilities and equipment could serve to boost their motivation and job satisfaction. This targeted approach may help address retention challenges and improve the overall healthcare workforce in the public sector.

These findings contribute to the existing literature on physicians' job preferences by suggesting that non-monetary attributes may play a significant role in shaping their preferences. Contrary to previous studies that emphasize the importance of income increases on job preferences, especially in developing countries (Bao, Huang, Wang, Yan, & Chen, 2023; Hamouzadeh,

Akbarisari, Olyaeeemanesh, & Yekaninejad, 2019; Kolstad, 2011), our results align with the literature suggesting that autonomy (Zweifel, 2011), hours worked (Scott et al., 2013), control over working hours (Holte et al., 2015), training (Lagarde, Pagaiya, Tangcharoensathian, & Blaauw, 2013), opportunities for professional development (Holte et al., 2015) and learning environment (Ubach, Scott, French, Awramenko, & Needham, 2003), are among the most valued attributes by physicians. This research highlights the importance of creating work environments that foster physicians' motivation by emphasizing the intrinsic rewards of the medical profession (decision autonomy), promoting professional development (training), work-life balance (personalized schedules) and best working and patient treatment conditions (facilities and equipment). This study was conducted in a single country with a healthcare training and hiring system that is largely uniform across the country. Physicians in the Portuguese public sector receives similar base salaries. However, salaries can differ based the amount of out-of-hours work. Respondents highly valued the time flexibility attribute. This suggests that compensating physicians adequately for time rigid contracts may be a way to address the detrimental effects it may have on work-life balance. Regarding the regulation of dual practice, there is a wide international variation. In Portugal, dual practice was discouraged by offering allowances or other work benefits to physicians who work exclusively for the public sector (García-Prado et al., 2007).

While this DCE approach has many advantages, it also suffers from some limitations. Firstly, the utilization of an online survey, while advantageous for reaching a broad spectrum of respondents and providing comparable face validity to face-to-face DCE administration (Jiang et al., 2023), may have constraints in terms of representativeness. For instance, only 8.5% of our sample consisted of physicians aged 55 or older, constraining our conclusions for this age group. Moreover, the use of convenience sampling through online distribution channels, such as medical professionals' social networks and newsletters, facilitated access to a diverse group

of physicians. However, it is important to acknowledge the potential for physician self-selection bias. Respondents' decisions to participate may have been influenced by factors such as proximity to such networks, availability at a given time, or their willingness to engage in the research. It's worth noting that we did not analyse respondents' engagement metrics or study the response rates, although these have recently recommended practices (A. Pearce et al., 2021; Watson, Becker, & de Bekker-Grob, 2017) in this field. Exploring these aspects could be addressed in future research. Despite our efforts to ensure the sample aligns with the Portuguese physician population (by disseminating the survey across various networks and extending the deadline to collect a larger sample), sample selection bias may still be present. Therefore, caution should be exercised when generalizing the results to the entire Portuguese physician population.

Secondly, the design of the DCE was intentionally kept simple to avoid churn and cognitive burden on the respondents. We presented them with a forced dual choice (between Job A or B), without a choice to choose neither. While this simplification may not fully capture real-life decision-making, preventing us from mapping all possible career choices, it was necessary to simplify the choice task and minimize the proportion of respondents that choose to opt out because they find the choice tasks too complex or difficult (Veldwijk, Lambooij, de Bekker-Grob, Smit, & de Wit, 2014). During pre-testing, only a small number of respondents (n=3) had difficulty choosing between the presented scenarios, suggesting that most found suitable options. On the other hand, we used non-numerical attributes like "high," "low," "frequent," and "infrequent." This choice aligns with common practice in similar studies focusing on non-monetary job attributes. These levels were selected because no universally accepted numerical benchmarks exist for attributes like training frequency. Using these terms aimed to make the survey more accessible and reduce the cognitive load for respondents. However, we recognize that these terms may have varied interpretations and do not capture the full complexity of

attribute definitions. Thirdly, hypothetical bias should also be considered, where individuals' stated preferences in experimental settings may not accurately reflect their actual behaviour. In our study, we took steps to minimize this bias by excluding attributes that physicians may consider unacceptable, specifically those that could potentially harm patients or compromise the quality of care.

Despite these limitations, this research expands the current evidence on physician decision-making, which has predominantly focused on practice location (preference for rural area jobs) or choice of specialty preferences. By considering other important "push-pull" factors that influence job choices, this study broadens our understanding of physician job preferences.

Future research should seek to overcome these limitations by expanding the sample size while ensuring a broader representation of age groups, and various work arrangements spanning different sectors and medical specialties. Expanding the participant pool in these ways would enhance the generalizability of the findings, making them more applicable to a wider range of healthcare contexts. Furthermore, future studies could consider employing more complex experimental designs. For instance, including an "opt-out" option in the choice set that represents the participants' current job situation would allow for a more dynamic exploration of job preferences. Combining DCE studies with real-world evidence of choices using administrative data, among other sources, can complement the findings from experimental studies and provide a comprehensive understanding of physicians' decision-making mechanisms.

3.6 Conclusions

Retaining high-performing physicians is crucial for maintaining high-quality and efficient healthcare systems. Understanding the factors influencing health workforce labour decisions is

essential for developing effective policies aimed at attracting and retaining skilled professionals. Our findings emphasize that non-monetary factors, such as autonomy and training opportunities, significantly influence physicians' job choices, in addition to financial considerations. These results highlight the importance of addressing the possible root causes behind poor autonomy, training and scientific development within healthcare institutions when aiming at retaining and attracting physicians to public healthcare providers.

While physician placement has grown in both the public and private sectors in recent years, further growth is necessary to address the unmet healthcare needs of the ageing population and to replace those who will soon retire. We hope that this research can provide insights into physicians' responsiveness to a combination of financial and non-financial factors, offering potential mechanisms to guide their allocation towards sectors experiencing greater demand and need.

With the changing landscape of healthcare providers, decision-makers can use DCEs like the one conducted in our study to inform their policies. These DCEs can be tailored to specific services and conditions, aiding in the development of more effective attraction and retention policies. Furthermore, similar studies can be utilized to understand the preferences of other health professionals and determine the circumstances under which their preferences play a decisive role. Such insights can be of interest to policymakers and researchers seeking to understand the drivers behind health workforce labour decisions.

3.7 References

Ammi, M., & Peyron, C. (2016). Heterogeneity in general practitioners' preferences for quality improvement programs: a choice experiment and policy simulation in France. *Health Economics Review*, 6(1), 44. Retrieved from https://doi.org/10.1186/s13561-016-0121-7

Armour, B. S., Pitts, M. M., Maclean, R., Cangialose, C., Kishel, M., Imai, H., & Etchason, J. (2001). The effect of explicit financial incentives on physician behavior. *Archives of Internal Medicine*, 161(10), 1261–1266. Retrieved from https://doi.org/10.1001/archinte.161.10.1261

Bao, M., Huang, C., Wang, L., Yan, G., & Chen, G. (2023). Eliciting primary healthcare physicians' preferences for job characteristics in rural China: a discrete choice experiment. *BMJ Open*, 13(3), e056741. Retrieved from https://doi.org/10.1136/bmjopen-2021-056741

Barros, P. P., & Costa, E. (2022). *Recursos Humanos em Saúde*.

Barros, P. P., & Simões, J. D. A. (2007). Portugal: health system review. *Health Systems in Transition*.

Berman, P., & Cuizon, D. (2004). Multiple public-private jobholding of health care providers in developing countries. *An Exploration of Theory and Evidence. Issue Paper-Private Sector. London: Department for International Developmnent Health Systems Resource Centre Publication*.

Brugha, R., Clarke, N., Hendrick, L., & Sweeney, J. (2021). Doctor retention: A cross-sectional study of how ireland has been losing the battle. *International Journal of Health Policy and Management*, 10(6), 299–309. Retrieved from https://doi.org/10.34172/ijhpm.2020.54

Carlsen, B., Hole, A. R., Kolstad, J. R., & Norheim, O. F. (2012). When you can't have the cake and eat it too. A study of medical doctors' priorities in complex choice situations. *Social Science and Medicine*, 75(11), 1964–1973. Retrieved from https://doi.org/10.1016/j.socscimed.2012.08.005

Carlsson, F., & Martinsson, P. (2003). Design techniques for stated preference methods in health economics. *Health Economics*, 12(4), 281–294. Retrieved from https://doi.org/10.1002/hec.729

Cheng, T. C., Kalb, G., & Scott, A. (2018). Public, private or both? Analyzing factors influencing the labour supply of medical specialists. *Canadian Journal of Economics/Revue Canadienne d'économique*, 51(2), 660–692. Retrieved from https://doi.org/10.1111/caje.12334

Clark, M. D., Determann, D., Petrou, S., Moro, D., & de Bekker-Grob, E. W. (2014). Discrete Choice Experiments in Health Economics: A Review of the Literature. *PharmacoEconomics*, 32(9), 883–902. Retrieved from https://doi.org/10.1007/s40273-014-0170-x

Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., & Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Economics*, 21(6), 730–741. Retrieved from https://doi.org/10.1002/hec.1739

Coast, J., & Horrocks, S. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of Health Services Research and Policy*, 12(1), 25–30. Retrieved from https://doi.org/10.1258/135581907779497602

Correia, T., Gomes, I., Nunes, P., & Dussault, G. (2020). Health workforce monitoring in Portugal: does it support strategic planning and policy-making? *Health Policy*. Retrieved from https://doi.org/https://doi.org/10.1016/j.healthpol.2019.12.014

Costa, E., Santos, R., & Barros, P. P. (2021). The Financial Sustainability of the Portuguese Health System. In *The Sustainability of Health Care Systems in Europe* (Vol. 295, pp. 209–229). Emerald Publishing Limited.

Crabbe, M., & Vandebroek, M. (2012). Using appropriate prior information to eliminate choice sets with a dominant alternative from D-efficient designs. *Journal of Choice Modelling*, 5(1), 22–45. Retrieved from https://doi.org/10.1016/S1755-5345(13)70046-0

Culler, S. D., & Bazzoli, G. J. (1985). The moonlighting decisions of resident physicians. *Journal of Health Economics*, 4(3), 283–292. Retrieved from https://doi.org/10.1016/0167-6296(85)90034-7

Czajkowski, M., & Budziński, W. (2019). Simulation error in maximum likelihood estimation of discrete choice models. *Journal of Choice Modelling*, 31, 73–85. Retrieved from https://doi.org/10.1016/j.jocm.2019.04.003

de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a

review of the literature. *Health Economics*, 21(2), 145–72. Retrieved from https://doi.org/10.1002/hec.1697

Eggleston, K., & Bir, A. (2006). Physician dual practice. *Health Policy*, 78(2–3), 157–166. Retrieved from https://doi.org/10.1016/j.healthpol.2005.09.007

Ferreira, M., Lopes, A., Guimarães, M., & Barros, H. (2018). The Medical Career and the Key Factors Driving the Exodus of Doctors from the National Health Service in Portugal. *Acta Médica Portuguesa*, 31(9), 483–488. Retrieved from https://doi.org/10.20344/amp.10121

Ferreira, P. L., Raposo, V., Tavares, A. I., & Correia, T. (2020). Drivers for emigration among healthcare professionals: Testing an analytical model in a primary healthcare setting. *Health Policy*, 124(7), 751–757. Retrieved from https://doi.org/10.1016/j.healthpol.2020.04.009

Ferrinho, P., Van Lerberghe, W., Fronteira, I., Hipólito, F., & Biscaia, A. (2004). Dual practice in the health sector: Review of the evidence. *Human Resources for Health*, 2, 1–17. Retrieved from https://doi.org/10.1186/1478-4491-2-14

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science*, 29(3), 393–421. Retrieved from https://doi.org/10.1287/mksc.1090.0508

Garattini, L., & Padula, A. (2018). Dual practice of hospital staff doctors: hippocratic or hypocritic? *Journal of the Royal Society of Medicine*, 111(8), 265–269. Retrieved from https://doi.org/10.1177/0141076818783332

García-Prado, A., & González, P. (2011). Whom Do Physicians Work For? An Analysis of Dual Practice in the Health Sector. *Journal of Health Politics, Policy and Law*, 36(2), 265–294. Retrieved from https://doi.org/10.1215/03616878-1222721

García-Prado, A., González, P., & Gonzalez, P. (2007). Policy and regulatory responses to dual practice in the health sector. *Health Policy*, 84(2–3), 142–152. Retrieved from https://doi.org/10.1016/j.healthpol.2007.03.006

Grobler, L., Marais, B. J., & Mabunda, S. (2015). Interventions for increasing the proportion of health professionals practising in rural and other underserved areas. *Cochrane Database of Systematic Reviews*. Retrieved from https://doi.org/10.1002/14651858.CD005314.pub3

Hamouzadeh, P., Akbarisari, A., Olyaeeemanesh, A., Yekaninejad, M.-S. S., Olyaeemanesh, A., & Yekaninejad, M.-S. S. (2019). Physician preferences for working in deprived areas: A systematic review of discrete choice experiment. *Medical Journal of the Islamic Republic of Iran*, 33(1), 83. Retrieved from https://doi.org/10.34171/mjiri.33.83

Hamouzadeh, P., Akbarisari, A., Olyaeeemanesh, A., & Yekaninejad, M. S. (2019). Physician preferences for working in deprived areas: A systematic review of discrete choice experiment. *Medical Journal of the Islamic Republic of Iran*, 33(1), 83. Retrieved from https://doi.org/10.34171/mjiri.33.83

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, 30, 133–176.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied Choice Analysis* (2nd ed.). Cambridge: Cambridge University Press. Retrieved from https://doi.org/DOI: 10.1017/CBO9781316136232

Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*, 27(4), 1078–1094. Retrieved from https://doi.org/10.1016/j.jhealeco.2007.11.006

Holte, J. H., Kjaer, T., Abelsen, B., & Olsen, J. A. (2015). The impact of pecuniary and non-pecuniary incentives for attracting young doctors to rural general practice. *Social Science and Medicine*, 128, 1–9. Retrieved from https://doi.org/10.1016/j.socscimed.2014.12.022

Holte, J. H., Sivey, P., Abelsen, B., & Olsen, J. A. (2016). Modelling Nonlinearities and Reference Dependence in General Practitioners' Income Preferences. *Health Economics*, 25(8), 1020–1038. Retrieved from https://doi.org/10.1002/hec.3208

Hoogland, R., Hoogland, L., Handayani, K., Sitaresmi, M., Kaspers, G., & Mostert, S. (2022). Global Problem of Physician Dual Practices: A Literature Review. *Iranian Journal of Public Health*, 51(7). Retrieved from https://doi.org/10.18502/ijph.v51i7.10079

Jiang, R., Pullenayegum, E., Shaw, J. W., Mühlbacher, A., Lee, T. A., Walton, S., … Pickard, A. S. (2023).

Comparison of Preferences and Data Quality between Discrete Choice Experiments Conducted in Online and Face-to-Face Respondents. *Medical Decision Making*, 0272989X2311719. Retrieved from https://doi.org/10.1177/0272989X231171912

Johannessen, K.-A., & Hagen, T. P. (2014). Physicians' engagement in dual practices and the effects on labor supply in public hospitals: results from a register-based study. *BMC Health Services Research*, 14(1), 299. Retrieved from https://doi.org/10.1186/1472-6963-14-299

Kolstad, J. R. (2011). How can rural jobs be made more attractive to Tanzanian health workers ? Results from a discrete choice experiment . *Health Economics*, 20(September), 196–211.

Kumar, S., & Clancy, B. (2021). Retention of physicians and surgeons in rural areas—what works? *Journal of Public Health*, 43(4), e689–e700.

Labelle, R., Stoddart, G., & Rice, T. (1994). A re-examination of the meaning and importance of supplier-induced demand. *Journal of Health Economics*, 13(3), 347–368. Retrieved from https://doi.org/10.1016/0167-6296(94)90036-1

Lagarde, M., Pagaiya, N., Tangcharoensathian, V., & Blaauw, D. (2013). One size does not fit all: Investigating doctors' stated preference heterogeneity for job incentives to inform policy in Thailand. *Health Economics (United Kingdom)*, 22(12), 1452–1469. Retrieved from https://doi.org/10.1002/hec.2897

Levaillant, M., Levaillant, L., Lerolle, N., Vallet, B., & Hamel-Broza, J.-F. (2020). Factors influencing medical students' choice of specialization: A gender based systematic review. *EClinicalMedicine*, 28, 100589. Retrieved from https://doi.org/10.1016/j.eclinm.2020.100589

Liu, J. X., Goryakin, Y., Maeda, A., Bruckner, T., & Scheffler, R. (2017). Global Health Workforce Labor Market Projections for 2030. *Human Resources for Health*, 15(1), 11. Retrieved from https://doi.org/10.1186/s12960-017-0187-2

Lopes, D., Castro, E., & Simões, J. (2018). Saúde 2040. Planeamento de médicos e enfermeiros em Portugal. Coimbra: Edições Almedina.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., … Irwin, J. (1999). Combining sources of preference data for modeling complex decision processes. *Marketing Letters*, 10, 205–217.

Mandeville, K. L., Lagarde, M., & Hanson, K. (2014). The use of discrete choice experiments to inform health workforce policy: A systematic review. *BMC Health Services Research*, 14(1), 1–14. Retrieved from https://doi.org/10.1186/1472-6963-14-367

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470. Retrieved from https://doi.org/10.1002/1099-1255(200009/10)15:5<447::aid-jae570>3.0.co;2-1

McGrath, J. (2021). *Report on labour shortages and surpluses*. *European Labour Authority*. Retrieved from https://doi.org/10.2883/746322

McGuire, T. G. (2000). Physician agency (Vol. 1, pp. 461–536). Elsevier.

McPake, B., Russo, G., Hipgrave, D., Horta, K., Campbell, J., Hort, K., & Campbell, J. (2016). Implications of dual practice for universal health coverage. *Bulletin of the World Health Organization*, 94(2), 142.

Merlo, G., van Driel, M., & Hall, L. (2020). Systematic review and validity assessment of methods used in discrete choice experiments of primary healthcare professionals. *Health Economics Review*, 10(1), 1–9.

Ministério da Saúde. (2015). Decree-Law 101/2015. Lisbon, Portugal.

Newhouse, J. P. (1970). A model of physician pricing. *Southern Economic Journal*, 174–183.

Nicholson, S., & Propper, C. (2011). Medical Workforce. *Handbook of Health Economics*, 2, 873–925. Retrieved from https://doi.org/10.1016/B978-0-444-53592-4.00014-1

OECD. (2021). *Health at a Glance 2021* (Health at a Glance). Retrieved from OECD: 10.1787/ae3016b9-en

OECD. (2023). *OECD Economic Surveys: Portugal 2023* (Vol. 2023). Organisation for Economic Co-operation and Development. Retrieved from https://doi.org/10.1787/2b8ee40a-en

Pearce, A., Harrison, M., Watson, V., Street, D. J., Howard, K., Bansback, N., & Bryan, S. (2021). Respondent Understanding in Discrete Choice Experiments: A Scoping Review. *The Patient - Patient-Centered*

Outcomes Research, 14(1), 17–53. Retrieved from https://doi.org/10.1007/s40271-020-00467-y

Ramos, P., Alves, H., Guimarães, P., & Ferreira, M. A. (2017). Junior doctors' medical specialty and practice location choice: simulating policies to overcome regional inequalities. *European Journal of Health Economics*, 18(8), 1013–1030. Retrieved from https://doi.org/10.1007/s10198-016-0846-6

Reinhardt, U. E. (1972). A production function for physician services. *The Review of Economics and Statistics*, 54(1), 55–66.

Russo, G., de Sousa, B., Sidat, M., Ferrinho, P., & Dussault, G. (2014). Why do some physicians in Portuguese-speaking African countries work exclusively for the private sector? Findings from a mixed-methods study. *Human Resources for Health*, 12(1), 1–9.

Russo, G., McPake, B., Fronteira, I., & Ferrinho, P. (2014). Negotiating markets for health: An exploration of physicians' engagement in dual practice in three African capital cities. *Health Policy and Planning*, 29(6), 774–783. Retrieved from https://doi.org/10.1093/heapol/czt071

Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008). Discrete choice experiments in a nutshell. In *Using discrete choice experiments to value health and health care* (pp. 13–46). Springer.

Saether, E. M. (2005). Physicians' Labour Supply: The Wage Impact on Hours and Practice Combinations. *Labour*, 19(4), 673–703. Retrieved from https://doi.org/10.1111/j.1467-9914.2005.00317.x

Santana, P., Peixoto, H., & Duarte, N. (2014). Demography of physicians in Portugal: prospective analysis. *Acta Medica Portuguesa*, 27(2), 246–251.

Scott, A. (2001). Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. *Journal of Health Economics*, 20(3), 329–347. Retrieved from https://doi.org/10.1016/S0167-6296(00)00083-7

Scott, A., Holte, J. H., & Witt, J. (2020). Preferences of physicians for public and private sector work. *Human Resources for Health*, 18(1), 59. Retrieved from https://doi.org/10.1186/s12960-020-00498-4

Scott, A., Witt, J., Humphreys, J., Joyce, C., Kalb, G., Jeon, S.-H. H., & McGrail, M. (2013). Getting doctors into the bush: General practitioners' preferences for rural location. *Social Science and Medicine*, 96, 33–44. Retrieved from https://doi.org/10.1016/j.socscimed.2013.07.002

Simões, J. de A., Augusto, G. F. G. F., Fronteira, I. I., Hernandez-Quevedo, C., Simões, J. de A., Augusto, G. F. G. F., … Hernandez-Quevedo, C. (2017). Portugal: Health System Review. *Health Systems in Transition*, 19(2), 1–184.

Socha, K. Z., & Bech, M. (2011). Physician dual practice: a review of literature. *Health Policy*, 102(1), 1–7.

Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete Choice Experiments in Health Economics: Past, Present and Future. *PharmacoEconomics*, 37(2), 201–226. Retrieved from https://doi.org/10.1007/s40273-018-0734-2

StataCorp. (2021). Stata Statistical Software: Release 17. *College Station: StataCorp LLC*.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.

Ubach, C., Scott, A., French, F., Awramenko, M., & Needham, G. (2003). What do hospital consultants value about their jobs? A discrete choice experiment. *British Medical Journal*, 326(7404), 1432–1435. Retrieved from https://doi.org/10.1136/bmj.326.7404.1432

Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E. W., Smit, H. A., & de Wit, G. A. (2014). The Effect of Including an Opt-Out Option in Discrete Choice Experiments. *PLoS ONE*, 9(11), e111805. Retrieved from https://doi.org/10.1371/journal.pone.0111805

Verma, P., Ford, J. A., Stuart, A., Howe, A., Everington, S., & Steel, N. (2016). A systematic review of strategies to recruit and retain primary care doctors. *BMC Health Services Research*, 16(1). Retrieved from https://doi.org/10.1186/s12913-016-1370-1

Watson, V., Becker, F., & de Bekker-Grob, E. (2017). Discrete Choice Experiment Response Rates: A Meta-analysis. *Health Economics*, 26(6), 810–817. Retrieved from https://doi.org/10.1002/hec.3354

Witt, J. (2017). Physician recruitment and retention in Manitoba: results from a survey of physicians' preferences for rural jobs. *Canadian Journal of Rural Medicine : The Official Journal of the Society of Rural Physicians of Canada = Journal Canadien de La Medecine Rurale : Le Journal Officiel de La Societe de Medecine Rurale Du Canada*, 22(2), 43–53.

World Health Organization. (2016). *Global strategy on human resources for health: workforce 2030*. Retrieved from World Health Organization:

World Health Organization. (2022). *Health and care workforce in Europe: time to act*.

Zweifel, P. (2011). Swiss experiment shows physicians, consumers want significant compensation to embrace coordinated care. *Health Affairs*, 30(3), 510–518. Retrieved from https://doi.org/10.1377/hlthaff.2010.0954

3.8 Appendices

3.8.1 DCE design

Example of the questionnaire and choice seta (translated to English from the original).

Q1.1

Dear participant,

Thank you for participating in this survey on **physician mobility.** This study aims to understand the factors that are associated with workplace choice.

Before you begin:

 Make sure you have about 10 minutes of uninterrupted time;

 Maximize your internet browser window;

 If using a mobile device, put it in landscape mode;

 Answer according to your opinion and knowledge without resorting to the internet or other people's answers.

CONSENT FORM I voluntarily participate in this study and understand that I may stop at any time. I understand that my responses will be collected anonymously, the data will be treated following the regulations of the European Data Protection Regulation (GDPR EU) and will be used exclusively to produce scientific research and teaching. As I move forward, I agree with these statements.

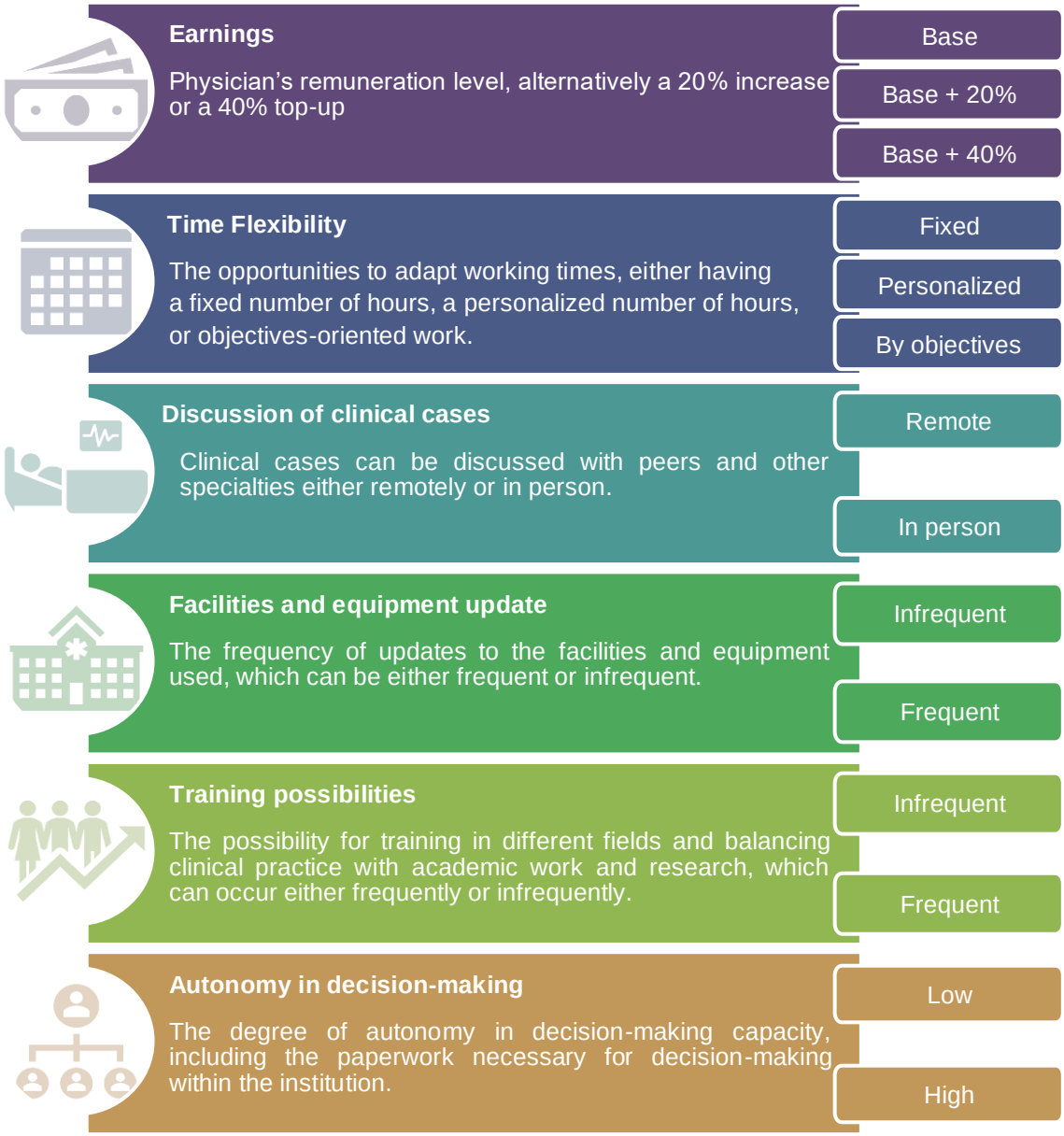
- ○ I agree to voluntarily participate in this study
- ○ I do not wish to participate

Q2.1

Important Instructions - Please Read Carefully Before Proceeding:

1. In this questionnaire we will present **two descriptions of job offers** so that you can reflect on **your preferences**.

2. The job offer descriptions provided are intentionally very general. The **jobs presented ONLY vary in the attributes listed** below, and each attribute can have either two or three possible levels:



Please proceed with the questionnaire keeping these instructions in mind.

Q2.2 We begin with an illustrative task.

The image displays two job offers. Imagine – regardless of your current employment situation – that you have two job offers to choose from.

Which Job (A or B) would you prefer?

	Job A	**Job B**
Earnings	Base	Base + 40%
Time Flexibility	Personalized	Fixed
Discussions of Clinical Case	In-person	Remote
Facility, Equipment, and Material Updates	Frequent	Infrequent
Training possibilities	Infrequent	Frequent
Autonomy in decision-making	High	Low
Choice	○ **Job A**	○ **Job B**

Q2.3

You chose ${Q2.2/ChoiceGroup/SelectedChoices}, which means you prefer the attributes of this offer over the attributes from the offer ${Q2.2/ChoiceGroup/UnselectedChoices}.

We then begin the questionnaire in which you will have to make this decision in eight different scenarios.

3.8.2 Latent Classes

Table 3.6 displays the coefficients derived from a latent class logit model with three latent classes, while Table 3.7 presents a mean comparison of physicians' characteristics across these three classes. In this model, maximum likelihood estimation was conducted using the expected maximization algorithm, along with a numerical search algorithm, Newton-Raphson. The selection of the number of latent classes was based on a trade-off between model fit (AIC and BIC) and the parameter estimate precision. Increasing the number of classes beyond 3 classes slightly improved the model fit but resulted in excessively large standard errors for the attributes coefficients, therefore we opted for a three-latent-class model.

Table 3.6: Estimation of the latent class logit model – 3 classes

	Class 1	Class 2	Class 3
Attribute	Coefficient (t-Stats)	Coefficient (t-Stats)	Coefficient (t-Stats)
Earnings	0.515***	0.308**	1.333***
	(5.29)	(3.29)	(12.37)
Time flexibility - personalized	0.849***	0.161	1.221***
	(5.86)	(0.80)	(9.70)
Time flexibility - objectives	1.054***	-0.127	0.719***
	(4.61)	(-0.82)	(7.53)
Facilities and equipment	0.105	1.033***	0.635***
	(0.69)	(7.05)	(8.90)
Discuss clinical cases	0.278*	0.394*	-0.160
	(2.19)	(2.28)	(-1.92)
Training	1.935***	0.885***	0.602***
	(8.85)	(6.18)	(7.20)
Autonomy	0.709***	2.205***	1.148***
	(5.56)	(9.27)	(7.80)
The posterior probability of belonging to a class	23.15%	22.59%	54.27%
N Observations	10944		
N Physicians	684		
Log-likelihood	-2803.9		
AIC	5653.88		
BIC	5758.03		

Note: The model is estimated with Stata's gllamm command using the modified Newton–Raphson algorithm. The model converged in 13 iterations. Full sample of physicians working in the private-, public-sector or dual practice. Significant coefficient estimates are indicated with *p < 0.1, **p < 0.05 or ***p < 0.01.

Table 3.7: Characteristics of physicians, by latent classes

	Class 1	Class 2	Class 3	1 vs 2	1 vs 3	2 vs 3
Attribute	Mean			P-value of differences test		
Public sector	0.46	0.51	0.48	0.00***	0.17	0.02**
Dual practice	0.43	0.41	0.45	0.22	0.05*	0.00***
GP	0.33	0.39	0.44	0.00***	0.00***	0.00***
Female	0.80	0.70	0.69	0.00***	0.00***	0.42
Aged under 30	0.24	0.14	0.19	0.00***	0.00***	0.00***
Aged 30 to 55	0.70	0.70	0.75	0.93	0.00***	0.00***
Aged over 55	0.06	0.16	0.06	0.00***	0.47	0.00***
N	155	159	370			

Note: Mean comparison of physician's characteristics across the three classes. P-value from the T-tests on the equality of means of each class (1 to 3) versus the comparison class (1 to 3). The significance of the tests is indicated with *p < 0.1, **p < 0.05 or ***p < 0.01.

Across all latent classes, jobs with higher earnings, training, and autonomy are associated with higher utility. Within the first class, which constitutes 23.15% of the sample, physicians display a positive and significant preference for nearly all job attributes, except for FACILITIES. Notably, they highly value attributes such as TRAINING, personalized or objective-based TIME FLEXIBILITY, as opposed to fixed schedules, AUTONOMY, and EARNINGS. This group of physicians is more likely to have hospital specialties rather than general practitioners. Additionally, they tend to be female and younger in age. This class may correspond to "early-career professionals or hospital interns".

In the second class, comprising 22.59% of the sample, the significance of attributes varies slightly. Physicians in this class prioritize uppermost AUTONOMY, as evidenced by the attribute's sign and significance. While the importance of continuing TRAINING remains

significant, having FACILITIES and equipment updated more frequently also holds importance in their job choice. Class 2 physicians are more inclined to work exclusively in the public sector, possess hospital specialties, and tend to be older. We classify this class as "experienced public hospital physicians".

The third latent class, encompassing 54.27% of survey respondents, is distinct in that all attributes are statistically significant except for the attribute related to DISCUSSIONS of clinical cases. Physicians in this class assign substantial value to EARNINGS and TIME FLEXIBILITY, slightly ahead of AUTONOMY and TRAINING in their preferences. Class 3 physicians are more likely to engage in dual practice, have a higher representation of general practitioners, and fall within the age range of 30 to 55. This class likely encompasses the preferences of "dual practitioners and general practitioners".

In summary, the goodness-of-fit statistics indicate only a slight improvement compared to the mixed logit models, affirming the stability of our primary conclusions across various model specifications. Specifically, these findings reinforce the notion that physicians prioritize AUTONOMY and TRAINING ahead of EARNINGS. However, some heterogeneity is observed, particularly between physicians working in dual practice and those exclusively in the public sector. Public sector physicians tend to value TRAINING opportunities, especially when younger, as well as having more frequently updated FACILITIES and equipment, particularly as they progress in their careers.

3.8.3 Preference heterogeneity according to specialty

Table 3.8: Coefficients and significance levels for Mixed Logit Model (MXL1) specification, for all attributes interacted with specialities dummy variable (GP=1 for general practitioners, GP=0 for hospital specialists).

	Mixed Logit Model - Interactions GP vs. Hospital			
	Mean	(SE)	Std. Dev.	(SE)
Earnings	1.176***	(0.0799)		
Time flexibility - personalized	1.257***	(0.130)	0.887***	(0.211)
Time flexibility - per objectives	0.986***	(0.110)	0.607**	(0.185)
Facilities and equipment	0.750***	(0.0867)	0.765***	(0.121)
Discuss clinical cases	0.220**	(0.0680)	-0.470**	(0.151)
Training	1.516***	(0.122)	1.035***	(0.118)
Autonomy	1.775***	(0.144)	1.385***	(0.135)
Earnings × GP	0.666**	(0.208)	1.194***	(0.174)
TF- personalized × GP	0.177	(0.259)	1.086*	(0.440)
TF- objectives × GP	-0.102	(0.213)	1.504***	(0.300)
Facilit. and equip. × GP	0.359*	(0.182)	0.588	(0.370)
Discuss Clin. Cases × GP	0.469	(0.293)	1.193**	(0.402)
Training × GP	-0.171	(0.135)	0.927***	(0.270)
Autonomy × GP	0.0876	(0.236)	0.688*	(0.332)
N Observations	10944			
N Physicians	684			
Log-likelihood	-2799.4			
AIC	5648.106			
BIC	5845.22			

Note: Standard errors are shown in parentheses.
Significance levels denoted as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
The estimation includes only physicians holding posts in both public and private practice or working exclusively in the public sector or private sector.
All estimates are relative to the reference level for each attribute, as follows: "base" for Earnings, "fixed schedule" for Time flexibility, "remote" for the Discussion of clinical cases, "frequent update" for Facilities and equipment, "frequent" for Training, and "low" for Autonomy.
Models estimated using 2000 Halton draws.

3.8.4 Semi-structured interviews

To refine our selection of attributes and align with the national context, we conducted face-to-face semi-structured, in-depth interviews with a diverse group of physicians, representing various practice settings. In total, we interviewed 17 physicians, encompassing those working in the public sector (6), the private sector (4), and engaging in dual practice (7). We also considered different age groups, including individuals aged 30-55 (9) and those over 56 (8), ensuring a comprehensive spectrum of perspectives.

These interviews were conducted via MS Teams by two independent researchers, following a well-defined protocol presented in Table A4.

From the interviews, we derived a list of attributes that emerged as important considerations. These attributes include salary, time flexibility (autonomy in managing working hours), contract flexibility, working conditions (logistics and processes), working conditions (technology and facilities), access to further education (technical/scientific growth), promotion possibilities, team management/leadership, and good leadership (promotion of leadership based on meritocracy). Notably, other attributes mentioned in the literature, such as location, staffing levels/skill mix, teaching opportunities, and proximity to family and friends, were not considered salient based on the insights gathered in the interviews.

Subsequently, the research team reviewed this attribute list to ensure the policy relevance of each attribute, leading to our final, parsimonious selection.

Semi-structured interviews Protocol

Introduction

We are conducting a study on "job mobility and its causes, whether it be between systems (public and private) or even within the same system". In this regard, the interview aims to discuss your experience as a physician, focusing on aspects of mobility and/or continuity in the profession. The interview is expected to last approximately 30 minutes. There are no right or wrong answers; we are interested in your opinion. The information provided will be strictly confidential and completely anonymous.

Interview Guide

1. Do you enjoy the profession you are in? Now, what are your roles or professional categories, and where do you work?
2. How long have you been performing your current job functions at your current workplace?
3. Are you satisfied with your work and the progression of your career? What are the main professional changes you have experienced throughout your career?
4. What would be the main reason or reasons that would lead you to leave your profession or role? What would weigh the most in that decision?
5. Do you feel that the profession is still valued by patients?
6. Do you feel that you participate in shaping the organization's future strategic objectives where you practice your profession? Do you feel recognized, and do your values, culture, and organization's goals coincide or align with yours? Are you satisfied with task delegation?
7. Do you consider research an important aspect of a physician's career? Does it currently play a significant role in your role and work?
8. In terms of career progression, how important is it to your professional and personal fulfilment and potential aspirations?
9. How do you envision performing your role in 10 years? How would you like it to be?

Conclusion of the interview

General characterization questions (age, gender, marital status, number of children, place of residence, academic degree, career stage in medicine)

3.8.5 Pre-test Questions

1. Were the attributes and their corresponding levels easily understandable, and did they cover the appropriate range? Were there attributes deemed irrelevant when making decisions?
2. Was the task comprehensible, and were the instructions sufficiently clear?
3. How did you find the experience of responding to the questions? Were there an excessive number of choices to make?