How Primary Care Practice Closures Affect Remaining Physicians: Evidence from Swiss Health Insurance Data.

Master Thesis

Department of Economics
University of Zurich

Xavier Schärer

Supervisor: Prof. Dr. Rainer Winkelmann

Industry Partner: tarifsuisse

Universität Zürich[UZH]

Course of Studies:	Economics
Student ID:	16-711-921
Address:	Winterthurerstrasse 11 8180 Bülach
E-Mail:	xavier.schaerer@uzh.ch
Date:	14.01.2026

Abstract

Primary care physician shortages threaten healthcare access, yet how practice closures affect supply and specifically remaining providers across different market conditions remains poorly understood. Using Swiss mandatory health insurance data (2018–2024) and non-traditional difference-in-differences design, this study analyzes how primary care closures impact regional supply and nearby physicians, stratified by regional provider density. The results show pronounced heterogeneity. In labor market regions with low general practitioner density operating near capacity, closures reduce aggregate primary care supply of consultations and billed volume by one percent per percentage point of closure intensity. Remaining physicians increase consultations by 0.5 percent but cannot fully compensate. Medium physician density regions exhibit no measurable effects, indicating spare capacity. High physician density regions show volume declines at the regional and provider level, potentially reflecting organizational transitions rather than care losses. These findings highlight that with different regional supply conditions, uniform policy approaches lead to undesirable outcomes in disadvantages regions. Undersupplied areas require targeted succession planning, practice takeover incentives to ensure an adequate health service supply as well as care continuity for complex patients.

1 Introduction

The Swiss healthcare system faces mounting pressures from demographic and structural shifts that threaten continuity of primary care provision. According to the Swiss Medical Association physician statistics 2024, every second physician is older than fifty years, and every fourth already exceeds sixty (FMH, 2025). This demographic developement foreshadows a significant wave of retirements in coming years. Simultaneously, structural trends, including growing prevalence of part-time work, salaried employment models, and persistent rural physician shortages complicate primary care delivery precisely where it remains most vulnerable (Zeller & Giezendanner, 2020).

Prior research has documented the patient-level consequences of these disruptions. When primary care practices close, patients reduce general practitioner (GP) consultations and shift toward specialists and hospital outpatient services (Bischof & Kaiser, 2021). This fragmentation increases costs and inefficiencies. Yet the supply-side response remains poorly understood: How do remaining providers adjust when competitors exit? Do they absorb displaced patients, and if so, how does this vary with local market conditions?

This thesis addresses this gap by examining provider-side dynamics. We ask: *How do general practitioner practice closures affect the number of patients, consultations and revenues of nearby general practitioners on a provider- and regional-level, and how do these effects vary with regional general practitioner density?*

Using Swiss mandatory health insurance claims data (2018–2024) and a difference-in-differences framework with staggered, repeated and continuous treatment intensity, we find pronounced heterogeneity. In low-density regions operating near capacity, closures reduce aggregate primary care supply by one percent per percentage point of closure intensity. Remaining GPs expand consultations by 0.5 percent per percentage point but cannot fully compensate, leading to selective patient intake and exclusion of the most complex cases. In medium-density regions, effects are insignificant—spare capacity enables frictionless redistribution. In high-density regions, service volumes decline, presenting puzzling results which we interpret as organisational transitions rather than genuine care losses. These patterns reveal a critical policy gap: uniform approaches fail when regional supply conditions diverge so sharply. In under-

supplied areas, succession planning and targeted takeover incentives become essential, alongside care continuity protections for multi-morbid patients. Coordinated workforce planning accounting for regional heterogeneity is imperative to prevent systematic primary care access erosion.

This thesis builds on recent literature examining physician exits as structural market shocks. Bischof and Kaiser (2021) analyze Swiss insurance claims data and find that patients affected by a GP closure experience a 12% decline in GP visits, partially offset by increases in specialist consultations (+11%) and hospital outpatient visits (+6%). While they find no deterioration in mortality or hospitalization rates, their results point to increased fragmentation and higher per-visit costs, particularly in regions with low primary care density. In a related study, Simonsen et al. (2021) show that although GP reassignment in Denmark does not reduce overall primary care usage, it leads to significantly higher rates of diagnostic activity and treatment initiation, especially for chronic conditions such as hypertension and diabetes. These findings suggest that interpersonal discontinuities in care may reveal latent medical needs, as newly assigned physicians re-evaluate patients' health status. Both studies underscore the role of continuity in care and reveal important heterogeneity depending on local supply conditions and patient characteristics.

Building on this patient-centered evidence, a more recent strand of literature has begun to examine provider-side responses and systemic dynamics. Zocher (2024) investigates GP exits in Austria and documents anticipatory behavior: soon-exiting physicians reduce treatment intensity while increasing referrals. After the transition, successors are found to increase preventive care and specialist referrals, yet the mortality rate among affected patients rises—likely due to insufficient follow-up care after hospital discharge. In contrast to complete practice closures, Hjalmarsson et al. (2024) analyze handovers in Switzerland where the practice remains open but the physician changes. Their findings show both short-term increases in visits and costs and persistent changes in utilization patterns, driven by differences in GP practice style. These studies position physician exits as structural market shocks that affect both the demand and supply sides of healthcare provision. By focusing on the supply-side behavioral responses of remaining primary care providers, this study contributes to the literature by revealing the heterogeneity of local supply conditions in response to exogenous practice closures. The differential regional responses permits the em-

pirical identification of primary care market conditions characterized by supply constraints versus excess capacity. This approach transcends conventional static definitions of under- and oversupply, instead employing market mechanisms to classify supply states.

The remainder of this thesis proceeds as follows. Section 2 presents the institutional background, develops a conceptual framework linking regional supply conditions to provider responses, describes the data and closure identification and empirical methodology. Section 3 presents results by GP density to examine treatment effect heterogeneity on the regional and provider levels. Section 4 discusses findings in light of the conceptual framework, existing literature and methodological limitations. It works out directions for future research as well as relevant policy implications. Section 5 concludes the thesis by summarizing the main results and takeaways.

2 Material and Methods

2.1 Institutional Background and Primary Care Market

Switzerland operates a system of mandatory health insurance (MHI) in which private insurers are legally required to accept all applicants. Coverage is extensive and includes most physician services, pharmaceuticals, inpatient care, physiotherapy, and long-term care (De Pietro et al., 2015). Each insurer must offer a basic plan with a standard deductible of CHF 300 and unrestricted provider choice. In exchange for lower premiums, individuals may opt for higher deductibles (up to CHF 2,500) or more restrictive provider choice models such as preferred provider organizations, health maintenance organizations, or telemedicine plans. While MHI accounts for approximately 40% of total healthcare expenditure in Switzerland, it represents the main income source for outpatient providers. In 2022, services reimbursed through MHI (including cost-sharing from patients) accounted for more than 80% of the total volume in primary care practices (Swiss Federal Statistical Office, 2025). This implies that MHI claims data provide a comprehensive picture of primary care utilization.

Patients' provider choice is linked to their insurance plan. The standard plan entails no mandatory gatekeeping mechanism, allowing patients unrestricted access to both GPs and specialists without referrals. Despite this freedom, approx-

imately 90% of Swiss residents report having a regular source of primary care (Merçay, 2016), suggesting that established patient-provider relationships remain the norm. In preferred provider organizations plans, patients designate a regular GP when signing the insurance contract; the GP acts as gatekeeper and issues referrals to specialists and hospitals. Patients can request a transfer to another GP by notifying their insurer, for instance upon relocation or following a GP's retirement. In health maintenance organizations plans, patients select from a list of contracted providers.

GPs constitute the most important source of ambulatory care in Switzerland (De Pietro et al., 2015). Historically, most GPs operated as self-employed physicians in solo practices. In 2014, 51% of all GPs worked in solo practices, while 84% of ambulatory physicians owned or co-owned their practice (Hostettler & Kraft, 2018; Senn et al., 2016). In recent years, however, group practices have become increasingly common, accompanied by a rise in salaried employment. Ambulatory providers are predominantly reimbursed under a national fee-for-service schedule (TARMED), with fees regulated at the federal and cantonal level such that a particular service is reimbursed identically for all physicians adjusted for price level of their canton (De Pietro et al., 2015). While access to primary care is generally considered satisfactory (OECD & WHO, 2011), GP density varies considerably across regions and tends to be lower in rural areas.

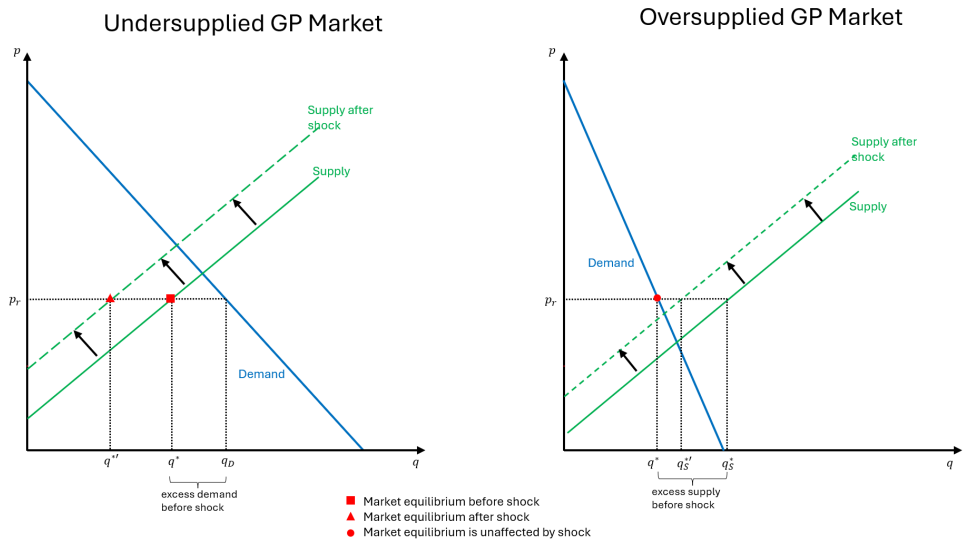
The GP workforce is characterized by a median age of 55 and a predominance of male physicians (FMH, 2025). This demographic profile foreshadows a substantial wave of retirements, raising concerns about continuity of care and access to primary health services. The potential supply shortage is further exacerbated by structural changes in workforce composition, including increased preference for part-time work and salaried employment in larger practices. These trends are amplified by the feminization of the profession (Gisler et al., 2017) and complicate solo practice succession. Zeller and Giezendanner (2020) report that GP density in Switzerland remains below the commonly used benchmark of one full-time equivalent per 1,000 inhabitants, though this measure imprecisely accounts for demographic factors such as age structure and morbidity. This supply constraint is accompanied by growing demand due to the healthcare needs of an aging population (Senn et al., 2016). The combination creates a context in which practice closures without succession may increasingly become a structural feature of the primary care landscape, with potentially

significant consequences for remaining providers and patients.

2.2 Conceptual Framework

This section develops a conceptual framework for analyzing how practice closures affect remaining providers and local healthcare markets. While prior research has predominantly examined patient-level responses to physician exits documenting shifts toward specialist care, hospital outpatient services and changes in treatment patterns (Bischof & Kaiser, 2021; Hjalmarsson et al., 2024; Simonsen et al., 2021; Zocher, 2024),less is known about supply-side adjustments among remaining physicians. We address this gap by conceptualizing practice closures as market shocks operating simultaneously at two levels: as a negative supply shock on the regional level, reducing aggregate healthcare capacity. And as a positive demand shock on the individual provider level, as displaced patients seek care elsewhere. This dual perspective builds on theoretical insights from health economics regarding capacity constraints and market equilibrium conditions under uncertainty (Arrow, 1963). Arrow's seminal work established that healthcare markets fundamentally differ from standard competitive markets due to uncertainty in demand and supply-side constraints including barriers to entry and quality controls. In healthcare systems with regulated pricing (such as Switzerland's TARMED fee schedule) price adjustments are infeasible, forcing all market-clearing to occur through quantity responses. Consequently, the impact of practice closures depends critically on local capacity conditions: regions operating near capacity limits may face binding constraints that prevent full absorption of displaced demand. While regions with spare capacity can accommodate shocks through provider service expansion. By explicitly modeling these capacity constraints and their heterogeneous effects across regional supply conditions, our framework generates testable predictions about how remaining providers adjust their service volumes, treatment intensity, and patient composition in response to nearby practice closures. In this subsection we first address the regional perspective, then examine individual providers and finally derive testable hypotheses.

Figure 1: Practice closures as a supply shock on the regional level



2.2.1 Regional Level: Aggregate Supply Shock

At the regional level, the impact of a closure is determined by the aggregate spare capacity of the remaining market. Figure 1 illustrates this mechanism using a stylized representation of the regional healthcare market. The horizontal axis measures the quantity of primary care services (q), while the vertical axis represents price (p). The demand curve (*Demand*) reflects the aggregate healthcare needs of the regional population, while the supply curve (*Supply*) represents the total capacity of all active GPs in the region before the closure.

When a practice closes, the regional supply curve shifts leftward from *Supply* to *Supply after shock*, reflecting the reduction in aggregate capacity. The magnitude and consequences of this shift depend critically on the initial market conditions. In the Swiss context, where prices are fixed through the regulated TARMED (p_r) fee-for-service schedule, markets cannot clear through price adjustments. Instead, adjustment must occur along quantity margins, with the nature of this adjustment differing fundamentally between undersupplied and oversupplied regions.

In regions characterized by undersupply (depicted in the left panel of Figure 1), the market operates at or near maximum capacity limits prior to the closure. The regional supply curve is steep or even vertical in the relevant range, indicat-

ing that remaining providers have full patient lists and limited ability to expand their service provision. As illustrated in the left panel of Figure 1, the leftward shift of the supply curve in this scenario leads to a significant reduction in the equilibrium quantity of services from q^* to $q^{*'}$. The local market cannot absorb the displaced demand because remaining providers are already operating at their capacity constraints. This creates excess demand at the regulated price level, forcing patients to either forego care, delay consultations, or incur higher search and travel costs by seeking services in neighbouring regions. The net result of a GP practice closure is a genuine reduction in the volume of primary care services provided locally. Although patients can potentially substitute with care in neighbouring regions.

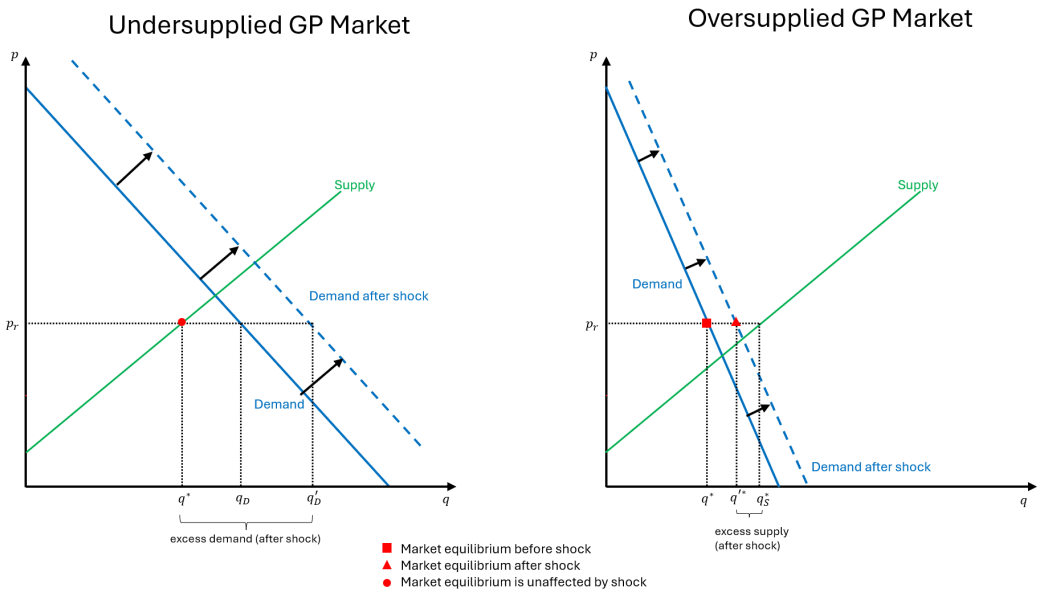
Conversely, regions characterized by oversupply have sufficient aggregate spare capacity among remaining providers(as depicted in the right panel of Figure 1). When a practice closes in such a region, the leftward shift of the supply curve from $Supply$ to $Supply\ after\ shock$ results in no reduction of the equilibrium quantity. Other GPs in the vicinity possess the resources and capacity to fill the gap left by the exiting GP. Patients displaced by the closure can be readily absorbed into existing practices without creating binding capacity constraints elsewhere in the market. Consequently, we expect to observe no net effect on the total volume of services provided in the region, as the supply shock can be fully compensated by the remaining providers spare capacity.

This regional-level analysis highlights a fundamental asymmetry: practice closures in undersupplied regions represent a genuine loss of healthcare capacity that cannot be compensated locally, whereas closures in oversupplied regions constitute a reallocation of patients among providers without affecting aggregate service provision. This distinction has important implications for access to care and motivates our empirical investigation of heterogeneous treatment effects by regional physician density.

2.2.2 Provider Level: Individual Demand Shock

From the perspective of an individual remaining provider, the closure of a nearby practice represents a positive demand shock. Figure 2 depicts this mechanism at the level of a single GP practice. The horizontal axis measures the quantity of services provided by the individual practice (q), while the vertical axis again

Figure 2: Practice closures as a demand shock on the provider level



represents price (p). The initial demand curve facing the provider (*Demand*) reflects the baseline patient population seeking care from this particular practice. The provider's individual supply curve (*Supply*) captures their production function and capacity constraints.

When a practice closure occurs in the same region, displaced patients seek alternative sources of care, shifting the demand curve facing remaining providers outward from *Demand* to *Demand after shock*. However, the provider's ability to respond to this increased demand is constrained by their individual capacity and production function. The shape of the individual supply curve plays a crucial role in determining the provider's response, and this shape differs systematically between undersupplied and oversupplied regions.

In a scenario of undersupply, remaining providers are likely already operating at or near their capacity limits prior to the closure (depicted in the left panel of Figure 2). . While demand for their services increases substantially after a nearby closure, their ability to increase the total quantity of care (q) is strictly limited by binding capacity constraints such as available consultation hours, practice infrastructure, and administrative capacity. So we expect to see little or no response under the condition of undersupply.

This imbalance between rising demand and fixed supply creates non-price

rationing pressure. Providers facing this situation may respond through several mechanisms. First, they may exert additional effort to accommodate some new patients, driven by ethical obligations or professional norms, but such expansion is necessarily modest given physical and temporal constraints. Second, and more significantly, providers may engage in triage, selectively accepting new patients based on clinical need, complexity, or compatibility with existing practice routines. This selective accommodation can lead to compositional shifts in the patient collective. According to Bischof and Kaiser (2021), chronically ill patients are particularly affected by practice closures due to their high dependence on continuity of care. If remaining GPs prioritize these vulnerable patients out of ethical obligations, the average complexity and morbidity of their patient collective will rise, potentially increasing the intensity of services per patient. Alternatively, if capacity constraints force providers to reject time-intensive chronic cases in favour of simpler acute consultations to maximize throughput, the patient mix may shift towards healthier, less complex cases. In either scenario, the binding capacity constraints dictate that an adjustment could occur through a compositional shift in the type of patients served rather than through a large expansion in the number of patients or total volume of services.

In a scenario of oversupply, remaining providers operate with spare capacity prior to the closure (depicted in the right panel of Figure 2). When the demand curve shifts from *Demand* to *Demand after shock* through a nearby closure, the provider can respond by increasing their quantity of services from q^* to $q^{*'}$ without encountering binding capacity constraints. They can take on a significant number of displaced patients by utilizing previously underutilized consultation slots, extending practice hours modestly, or reallocating administrative resources. Consequently, we expect to observe substantially larger effects on patient volume, consultation numbers, and revenues for providers in oversupplied regions compared to their constrained counterparts in undersupplied regions. Furthermore, because these providers are not forced to ration care, we do not anticipate compositional shifts in their patient collective. New patients can be accommodated without displacing existing patients or altering the treatment intensity.

This provider-level analysis complements the regional perspective by highlighting that even when aggregate regional supply is sufficient to absorb a closure (as in oversupplied regions), the distribution of that absorption across in-

dividual providers can vary considerably. It also underscores that in undersupplied regions, the combination of aggregate capacity constraints (regional level) and individual capacity constraints (provider level) creates a dual bottleneck that limits both the total volume of services and the flexibility of remaining providers to respond.

2.2.3 Hypotheses

Based on these mechanisms, we derive four testable hypotheses regarding the effects of practice closures. The hypotheses distinguish between aggregate regional effects and individual provider responses:

Regional Hypotheses

- **H1 (Service Volume):** Practice closures in undersupplied regions lead to a reduction in the total volume of primary care services provided locally, whereas no such negative effects are observed in oversupplied regions.
- **H2 (Cross-Regional Substitution):** In undersupplied regions, a lack of local capacity drives patients to consume more healthcare services in other regions.

Provider Hypotheses

- **H3 (Capacity Constraints):** The increase in patient numbers and volume for remaining providers is larger in oversupplied regions than in undersupplied regions, where providers are inhibited by binding capacity constraints.
- **H4 (Patient Collective):** In undersupplied regions, capacity constraints lead to compositional shifts in the patient collective of remaining GPs as they selectively accommodate new demand.

2.3 Data

Our primary data sources are the SASIS data pool (DP) and the Swiss National Provider Registry (*Zahlstellenregister*, ZSR), maintained by SASIS AG on behalf of all insurers. A summary of all used data sources is provided in Appendix A. We analyze annual data from 2018 to 2024, including all outpatient

healthcare providers authorized to bill services under mandatory health insurance in Switzerland. The DP contains information on billing volume (number of consultations, number of patients, and total MHI-billed revenue) and the ZSR on provider characteristics including location, physician gender, age, and practice age. Additionally, the data include aggregated patient characteristics such as age distribution, deductible levels, and hospitalization indicators. The datasets exclusively capture services billed and processed through MHI, thereby excluding privately financed healthcare services and out-of-pocket payments.

A critical definition issue concerns the unit of observation. A provider is defined as a legal entity or individual physician that has applied for billing authorization, identified by a unique registry number (ZSR-ID). Crucially, this does not correspond one-to-one with physical practice locations. Group practices exhibit two distinct organizational models affecting data structure. First, individual physicians within a group practice may each obtain separate ZSR-IDs, resulting in multiple observations per physical location; under this arrangement, physicians maintain separate financial accounts while sharing infrastructure and support staff. Second, physicians may establish a legal entity (*Aktiengesellschaft* or *Gesellschaft mit beschränkter Haftung*) that applies for a single practice authorization, where one ZSR-ID represents the entire practice with multiple employed physicians.

Upon registration, applicants are categorized into provider groups according to their medical specialties. Following established literature (Bischof & Kaiser, 2021; Hjalmarsson et al., 2024), we identify GPs using the categories "general internal medicine" (*Allgemeine Innere Medizin*) and "general practicing physician" (*Praktischer Arzt / Praktische Ärztin*). This classification introduces a challenge for mixed group practices: when a legal entity employs physicians with different medical specialties, it is assigned to the residual category "group practices". This heterogeneous category contains mixed specialist group practices, mixed specialist-GP practices, and any group practice with at least one physician holding multiple specialty titles. Consequently, the provider group classification does not permit precise identification of the service portfolio in these mixed practices, representing a limitation for measuring closure impacts in markets where mixed group practices are an important player in the primary care market.

2.4 Definition of Practice Closures

Since the national provider registry only explicitly records practice openings, closures must be identified indirectly. Following the methodological approach established in the literature (Bischof & Kaiser, 2021; Hjalmarsson et al., 2024), we define a practice as closed when its activity level falls below a predefined threshold. For the primary analysis, we classify a ZSR-ID as inactive if it records fewer than 50 patients and less than CHF 100,000 in physician services billed through MHI within a calendar year.

This threshold-based identification entails several methodological considerations. First, a provider may remain active at minimal levels even after crossing the threshold. However, provided the threshold is sufficiently low, this does not compromise validity, as the objective is to capture effective loss of providers who contribute substantially to healthcare supply. The data indicate that many physicians, particularly at old age, continue billing at marginal activity levels, often for personal use or a limited patient base. Second, any threshold choice is inherently arbitrary given the absence of evidence-based criteria for inactivity. We address this limitation through sensitivity analyses with alternative threshold specifications (see Appendix D).

Consistent with prior research demonstrating that practice handovers generally maintain continuity of care (Hjalmarsson et al., 2024; Zocher, 2024), we exclude closures where a handover occurred. We define a handover as the observation of a new GP-ZSR-ID at the same address coinciding with the first inactivity by the closing GP practice. Additionally, we exclude all closures at addresses where at least one other GP remains active, accounting for physician exits from group practices where remaining GPs can provide continuity of care. This address-based identification is adequate for rural areas, where typically no more than one practice occupies a single building. However, the method exhibits limitations in densely populated areas, where multiple independent practices may occupy different floors of the same building, potentially leading to false exclusions.

A further limitation arises from the use of annual data, which introduces imprecision regarding exact closure timing. The data reveal only the year in which a ZSR-ID becomes inactive, leaving ambiguity as to whether actual closure occurred in the preceding year or in the year of first observed inactivity. While

monthly data, as employed by Bischof and Kaiser (2021) and Hjalmarsson et al. (2024), would permit more precise timing, but this is beyond the scope of this thesis. These measurement challenges underscore limited understanding of practice closures and their impact on healthcare supply as well a need for closer monitoring.

2.5 Empirical Strategy

This section outlines our identification strategy for estimating the causal effects of GP practice closures on remaining primary care providers. We employ a generalized difference-in-differences framework exploiting variation in the timing and intensity of practice closures across labor market regions (LMR).

2.5.1 Treatment Definition

The treatment is defined as closure of a GP practice in a given LMR where no handover occurred. To account for heterogeneity in regional market size and better identify shock magnitude, we operationalize treatment as continuous rather than a binary indicator. Specifically, treatment intensity equals the share of patients served by closing practices relative to total primary care patients in the region. As illustrated in Figure 3, patient volume of closing practices declines substantially in years preceding inactivity. To capture the full extent of the supply shock and avoid attenuation bias from anticipatory effects, we measure treatment intensity using patient counts from two years prior to observed inactivity ($t-2$). The number of patients is defined as the number of unique persons who are billed for a service by a ZSR-ID in a calendar year (for further details see Appendix B). We employ patient numbers as the basis for treatment intensity, as this metric is less susceptible to physician-induced variation compared to consultation volume or billed services.

2.5.2 Unit of Analysis and Regional Aggregation

The analysis is conducted at regional and provider level, with providers defined by their unique ZSR-ID. Providers are aggregated to LMRs, which serve as the geographic unit for treatment assignment. Switzerland comprises 101 LMRs, delineated such that approximately 57% of employed individuals both reside and work within the same region (Swiss Federal Statistical Office, 2021). This

functional economic definition provides a more appropriate approximation of healthcare market boundaries than administrative units such as districts or cantons, as it reflects actual patient mobility patterns and provider catchment areas. The 101 LMRs are nested within 16 labor market greater regions, which we use to control for inter-regional differences. As we can see in Figure 7 in the Appendix, LMRs seem to be adequate to analyze the health care market as over 80% of services in the average LMR are provided for inhabitants of that LMR.

2.5.3 Treatment Timing

Due to annual data frequency of our data, precise closure timing cannot be identified. We define the closure year as the first year of observed inactivity, though actual closure may have occurred in the preceding year. To address this concern, we include a leading closure rates coefficient in our specification. This approach allows a closure at time t to affect nearby providers in both year t and year $t-1$. Even with precise timing identification, this specification could remain valid, as Zocher (2024) documents that primary care practice closures involve gradual service decline in the year preceding definitive closure.

2.5.4 Treated Population

The population of interest comprises healthcare providers who may absorb patients from closing GP practices. We estimate effects only for GPs who do not close during the observation period. We do not analyse specialists and hospital outpatient departments for two reasons: first, it is unclear which specialists would be affected (e.g., surgeons are less likely to be affected than dermatologists); second, specialist and hospital services have limited coverage in rural areas, implying greater spillovers across LMR boundaries. A provider is treated if it operates in an LMR that experienced at least one GP practice closure (without handover) in the treatment year or the following year. This specification yields an analysis period of 2020 to 2023.

2.5.5 Estimation Model

We begin by estimating effects at the regional level to understand how practice closures affect aggregate healthcare provision. At the LMR level, we estimate the following specification:

$$Y_{rt} = \beta_0 + \beta_1 \text{closurerate}_{rt} + \beta_2 \text{closurerate}_{r,t+1} + \text{population}_r + \varepsilon_{rt} \qquad (1)$$

where Y_{rt} denotes the outcome for LMR r in year t. closurerate_{rt} represents the treatment intensity, defined as the number of practice closures relative to the existing provider stock in region r and year t. The leading term $\beta_2 \text{closurerate}_{r,t+1}$ accounts potential anticipatory effects. We control for population size in the region and ε_{rt} is the stochastic error term. The coefficients of interest are β_1, identifying the average contemporary effect of the regional closure rate on aggregate outcomes, and β_2, which measures leading effects.

At the provider level, we estimate effects on individual remaining providers' outcomes:

$$Y_{irt} = \beta_0 + \beta_1 \text{closurerate}_{rt} + \beta_2 \text{closurerate}_{r,t+1} + \delta_R + \tau_t + \varepsilon_{irt} \qquad (2)$$

where Y_{irt} denotes the outcome for provider i in LMR r in year t. LMR r is nested within greater region R. closurerate_{rt} represents treatment intensity in region r and year t. The leading term $\text{closurerate}_{r,t+1}$ addresses concerns about closure timing and captures gradual effects. δ_R denotes labor market greater region fixed effects, τ_t denotes year fixed effects, and ε_{irt} is the error term. Standard errors are clustered at the LMR level. We do not include provider level fixed effects which are common with panel data, as we fear that it will absorb too much of the variation of treatment. The coefficients of interest are β_1, identifying the average effect of increased regional closure rate in the year of the first inactivity on provider-level outcomes, as well as β_2, which measures leading effects. For the specification of the provider level extended model see Appendix C.

2.5.6 Outcome Variables

On the regional level we focus on the aggregate number of consultation as well as the aggregate volume of physician services. To measure potential shifts in regional domestic health provision we look at the share of consultations and volume of physician services which is consumed by inhabitants of the LMR. On the provider level we have the following outcome variables for service volume:

number of patients, number of consultations and volume of physician services (for definitions of these measures see Appendix B). Further we measure service intensity with number of consultations per patient and volume of physician services per patient. Lastly we look at the mean age of patients and the share of patients with an inpatient stay in the year before. These two variables are frequently used as morbidity indicators.

2.5.7 Heterogeneity Analysis

We investigate treatment effect heterogeneity along the dimension of regional primary care supply by stratifying the analysis by GP density tertiles at the LMR level. This stratification is motivated by prior evidence suggesting that practice closures have more pronounced effects in areas with limited provider availability (Bischof & Kaiser, 2021). We define GP density as the number of active GP-ZSR-IDs per 1,000 inhabitants within each LMR. More sophisticated density measures could account for patient mobility patterns, regional morbidity levels, and heterogeneous physician activity (as used in Bischof and Kaiser, 2021). The distribution of LMR into GP density tertiles can be found in Figure 6 in the Appendix.

We employ GP density as a proxy for local supply conditions, interpreting it in relative rather than absolute terms. Specifically, regions classified in the lower tertile of GP density are characterized as having more constrained supply relative to regions in the upper tertile. This interpretation aligns with our conceptual framework: lower-density regions face greater capacity constraints and are more likely to exhibit undersupply characteristics, whereas higher-density regions have more spare capacity and are more consistent with oversupply conditions. The tertile classification thus captures the relative position of each region along the supply spectrum, allowing us to test whether provider responses to closures differ systematically based on local market tightness.

2.5.8 Identification assumptions

The empirical strategy is designed to recover causal effects of practice closure intensity on regional and provider-level outcomes. Within a potential-outcomes framework, the coefficients on closure intensity can be interpreted as average causal effects under three main identifying assumptions.

First, the generalized difference-in-differences design relies on a parallel-trends assumption. In the absence of practice closures, LMRs with different closure intensities are assumed to follow similar trends in outcomes once greater-region and year fixed effects are taken into account. This implies that any systematic deviation in trends between regions with higher and lower closure intensity can be attributed to the closure shock rather than to unobserved factors evolving differentially over time.

Second, causal interpretation requires exogeneity of closure timing and, more stringently, closure intensity with respect to unobserved determinants of outcomes. Exogeneity of timing is plausible because most exits are retirement-driven and thus primarily determined by physician age and long-term succession planning rather than by contemporaneous changes in local demand or policy (Bischof & Kaiser, 2021; Hjalmarsson et al., 2024). Descriptive evidence on the age distribution of closing physicians (Table 9 in the Appendix) supports this interpretation. Exogeneity of intensity is harder to prove: regions with structurally weaker primary care supply, or distinct long-run utilization trends may both experience higher closure intensity and diverging outcome trajectories. The inclusion of greater-region and year fixed effects, population controls, and practice characteristics in the extended specifications address, but cannot fully rule out, such residual confounding.

Third, the Stable Unit Treatment Value Assumption is only approximately satisfied. Practice closures in one LMR may induce patient flows to neighbouring regions, thereby affecting outcomes outside the treated region. By using LMRs as functional catchment areas and conditioning on greater-region fixed effects, the empirical design aims to contain these spillovers. Nevertheless, some degree of interference across regions is likely to remain, implying that the estimated coefficients should be interpreted as net effects of closure intensity on outcomes within a region, potentially combined with offsetting adjustments elsewhere.

2.5.9 Limitations of the Difference-in-Differences Specification

Our empirical approach deviates from the canonical two-period, two-group difference-in-differences framework in three dimensions: treatment timing varies across regions (staggered adoption), treatment intensity is continuous rather than bi-

nary, regions can experience closures repeatedly and we specify the model that closures have a leading effect. Recent methodological work demonstrates that two-way fixed effects estimators can produce biased estimates under treatment effect heterogeneity in such settings, as already-treated units serving as controls may introduce bias when effects evolve over time or accumulate with repeated exposure (Callaway et al., 2024; de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021). The repeated nature of treatment is particularly problematic: our estimates may capture contemporaneous effects, cumulative effects of past closures, or some weighted combination. A comprehensive treatment of these challenges exceeds the scope of this thesis. We acknowledge that our estimates should be interpreted with caution. Due to this specification we also will not provide any event studies in this thesis, as there is no simple treatment and control group. Future research could employ estimators designed for staggered, continuous, repeated and non-absorbing treatments (Callaway et al., 2024; de Chaisemartin & D'Haultfœuille, 2023) to address these concerns.

3 Results

This section presents the empirical findings on the association between GP practice closures and remaining GPs. The analysis proceeds in three steps: First, descriptive evidence documents systematic differences across regions with varying levels of GP density and traces the activity of closing practices over time. Second, difference-in-differences estimates at the LMR level quantify associations with aggregate patient redistribution and care utilization. Third, provider-level regressions assess changes in practice size, service volume, treatment intensity, and patient composition due to GP practice closures.

3.1 Descriptive Results

Table 1 summarizes practice-level characteristics by tertiles of GP density for the period 2020–2023. The distribution of observations reflects the spatial concentration of primary care; low density regions account for 2,422 practice-year observations, compared to 5,382 in medium density and 10,197 in high density regions. Practices in sparsely supplied regions are predominantly located in rural or intermediary municipalities, whereas high density regions are largely

urban. Physician age and practice age are similar across tertiles, with an average GP age of approximately 55 to 57 years and an average practice age of roughly 16 years. Gender composition varies by density: in low density regions, approximately two-thirds of GPs are male, whereas the female share reaches about 40 percent in high density regions.

Regarding patient demographics, the mean patient age is higher in low density regions. The hospitalization rate, used here as a proxy for patient morbidity, remains relatively stable across tertiles.

Practice size and service volume display an inverse relationship with GP density. Average patient numbers, annual consultations, and total physician service volumes are highest in low density regions and decline as density increases. Conversely, intensity-normalized outcomes show a positive gradient. Consultations per patient and revenue per patient increase from low to high density regions. In high density regions, GPs treat fewer patients on average but provide more consultations and bill higher volumes per patient.

Table 9 in the Appendix characterizes closing GP practices. Closures occur in all three density groups at an annual rate of approximately 3%, totaling 179 closures in low density, 386 in medium density, and 506 in high density regions between 2020 and 2023. The average physician age at closure ranges from 62.5 to 63.9 years across tertiles, and the average practice age is between 22 and 24 years. Regarding location, the majority of closures in medium and high density regions are in urban municipalities, while closures in low density regions frequently occur in intermediary and rural areas.

Figure 3 depicts the evolution of key outcome variables for active and closing practices over time. Closing practices show a gradual decline in volume metrics in the years preceding recorded inactivity. Patient numbers remain stable until approximately two years before closure, followed by a steep decline. Consultation volumes and billed physician services begin to decrease slightly earlier than patient numbers. Active practices, in contrast, demonstrate stable or moderately growing trends throughout the observation period. The divergence between patient counts and service volumes in closing practices implies a reduction in service intensity (consultations and revenue per patient) during the pre-closure phase.

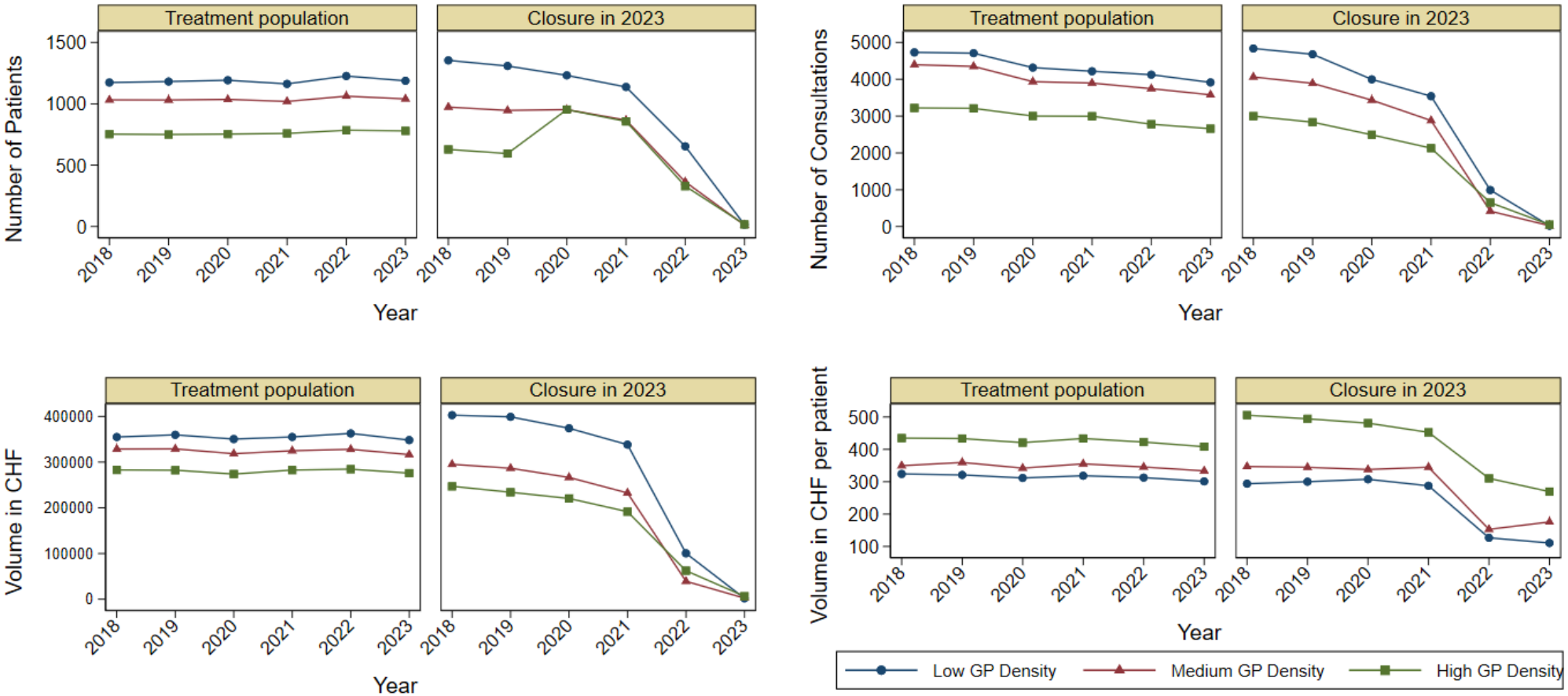
Table 1: Descriptive statistics by GP density tertiles

Variable	low GP density	medium GP density	high GP density	p-value (low vs medium)
Year				
2020	580 (24%)	1,271 (24%)	2,331 (23%)	
2021	596 (25%)	1,294 (24%)	2,500 (25%)	
2022	608 (25%)	1,350 (25%)	2,624 (26%)	
2023	638 (26%)	1,467 (27%)	2,742 (27%)	
Physician & Practice characteristics				
Sex				
institution	175 (7%)	415 (8%)	384 (4%)	
male	1,587 (66%)	3,335 (62%)	5,728 (56%)	
female	660 (27%)	1,632 (30%)	4,085 (40%)	
physician age	56.53 (9.34)	56.39 (8.92)	55.40 (10.00)	0.551
practice age	15.98 (10.43)	15.74 (10.31)	15.70 (10.64)	0.345
FSO classification of practice location				
urban	1,088 (45%)	3,395 (63%)	8,343 (82%)	
intermediary	733 (30%)	1,176 (22%)	1,306 (13%)	
rural	601 (25%)	811 (15%)	548 (5%)	
Volume & intensity measures				
number of patients	1,197 (736)	1,065 (667)	798 (914)	<0.001
number of consultations	4,205 (2,857)	3,892 (2,487)	2,929 (2,116)	<0.001
volume of physician services in CHF	358,719 (252,783)	332,230 (215,532)	288,004 (214,703)	<0.001
number of consultations per patient	3.62 (1.47)	3.92 (1.93)	4.06 (2.09)	<0.001
volume of physician services per patient	315.56 (140.08)	349.55 (217.76)	423.50 (349.31)	<0.001
Patient collective characteristics				
average age of patients	52.09 (6.06)	53.04 (6.81)	54.35 (8.00)	<0.001
share of patients with inpatient stay in t-1	0.11 (0.06)	0.11 (0.08)	0.12 (0.10)	0.003
Treatment variables				
closure rate at t	0.04 (0.06)	0.04 (0.03)	0.03 (0.04)	0.035
closure rate at t+1	0.05 (0.06)	0.04 (0.04)	0.03 (0.03)	0.018
N	2422	5382	10197	

[1] n (%); Mean (SD)

[2] t-test for continuous variables

Figure 3: Development of outcome variables over time in active practices and closing practices by GP density tertiles



These patterns support the methodological decision to measure treatment intensity using patient counts from two years prior to observed inactivity (t-2), as this predates the sharpest decline in practice activity.

3.2 Regional Level Estimation Results

We now examine how GP practice closures are associated with aggregate primary care provision at the LMR level. Table 2 presents the estimation results for the baseline specification controlling for LMR population size, stratified by GP density tertiles to assess whether regional responses vary with initial supply conditions. The regional analysis tests two core hypotheses: whether closures reduce aggregate service volumes (H1), and whether patients substitute lost local services by seeking care in neighboring regions (H2).

3.2.1 Low Density Regions

The results for low density LMRs provide strong evidence of binding capacity constraints at the regional level. Following practice closures, we observe statistically significant negative associations with both aggregate consultation volumes (coefficient: -1.34, 95% CI $[-2.52, -0.16]$, $p = 0.028$) and total physician services (coefficient: -1.35, 95% CI $[-2.56, -0.13]$, $p = 0.031$) in the contemporaneous period. These estimates indicate that a one percentage point increase in closure intensity corresponds to approximately a one percent reduction in aggregate regional supply relative to baseline levels. The near-unity elasticity suggests that remaining providers cannot compensate for the lost capacity, consistent with hypothesis H1 that closures in undersupplied regions lead to genuine reductions in total primary care provision.

Examining cross-regional substitution patterns, the coefficients for the share of consultations and physician services consumed domestically within the LMR reveal no statistically significant deviations from baseline levels (domestic consultations: $p = 0.460$; domestic volume: $p = 0.453$). This null finding provides no empirical support for hypothesis H2 in low density regions. The absence of significant cross-regional substitution implies that patients unable to access local primary care do not systematically seek services in neighbouring LMRs, raising concerns about potential unmet medical needs or delayed care in these undersupplied areas.

Table 2: Estimation results on LMR level: controlled for LMR population size

	log(volume)	log(nr of cons)	% domestic vol	% domestic cons
Panel A: Tertile 1 (Low Density)				
β_1, contemporary	-1.346*	-1.339*	5.957	5.819
	(0.031)	(0.028)	(0.453)	(0.460)
β_2, leading	-0.868	-0.901	-1.602	-1.481
	(0.227)	(0.201)	(0.835)	(0.844)
Observations	136	136	136	136
R-squared	0.618	0.614	0.429	0.442
Panel B: Tertile 2 (Medium Density)				
β_1, contemporary	0.003	0.237	5.041	5.539
	(0.998)	(0.815)	(0.631)	(0.604)
β_2, leading	-0.042	0.110	-6.551	-5.455
	(0.973)	(0.928)	(0.606)	(0.673)
Observations	136	136	136	136
R-squared	0.694	0.698	0.529	0.526
Panel C: Tertile 3 (High Density)				
β_1, contemporary	-4.790*	-3.942*	-31.479	-33.815
	(0.021)	(0.042)	(0.370)	(0.332)
β_2, leading	-3.401*	-2.937*	-1.790	-3.149
	(0.039)	(0.030)	(0.958)	(0.925)
Observations	132	132	132	132
R-squared	0.760	0.742	0.030	0.028

Notes: p-values in parentheses. **Significance levels**: *** p<0.001, ** p<0.01, * p<0.05. **volume**: volume of physician services; **nr of cons**: number of consultations; **% domestic vol**: share of physician services volume in the LMR which is consumed by inhabitants of that LMR; **% domestic cons**: share of consultations in the LMR which is consumed by inhabitants of that LMR.

3.2.2 Medium Density Regions

In medium density LMRs, the estimation yields null results across all regional outcomes, indicating market flexibility and sufficient spare capacity to absorb practice closures without friction. The coefficients for aggregate consultation volumes ($p = 0.815$) and physician services volumes ($p = 0.998$) are statistically indistinguishable from zero in the contemporaneous period. Similarly, the leading period estimates show no significant associations (consultations: $p = 0.928$; volume: $p = 0.973$). These findings suggest that medium density regions possess adequate aggregate capacity among remaining providers to redistribute displaced patients without creating binding constraints.

The null effects extend to cross-regional substitution patterns, where both domestic consultation share ($p = 0.604$) and domestic volume share ($p = 0.631$) remain statistically unchanged. The stability observed across all metrics supports the interpretation that medium density regions represent a functional market equilibrium where supply shocks are buffered by the aggregate flexibility of the provider network. In these areas, closures constitute a reallocation of patients among providers rather than a reduction in total service provision.

3.2.3 High Density Regions

The estimation results for high density LMRs reveal pronounced negative associations that warrant careful interpretation. In the contemporaneous period, regional consultation volumes decline significantly (coefficient: -3.94, 95% CI $[-7.73, -0.15]$, $p = 0.042$), as do physician services volumes (coefficient: -4.79, 95% CI $[-8.80, -0.77]$, $p = 0.021$). The leading period estimates similarly indicate substantial reductions (consultations: -2.94, 95% CI $[-5.57, -0.30]$, $p = 0.030$; volume: -3.40, 95% CI $[-6.62, -0.18]$, $p = 0.039$). The magnitude of these estimates suggests that each percentage point increase in closure intensity is associated with approximately a three to five percent reduction in baseline volumes. These findings present an empirical puzzle, as standard competitive models would predict capacity expansion rather than contraction in high density markets with multiple providers. We will further discuss this matter in Section 4.2.3.

Consistent with other density tertiles, the estimates for cross-regional substitution in high density regions show no statistically significant deviations (do-

mestic consultations: $p = 0.332$; domestic volume: $p = 0.370$), providing no evidence that closure-induced service reductions are offset by increased utilization in neighbouring regions.

3.3 Provider Level Estimation Results

We now examine associations between GP practice closures and outcomes for individual remaining GPs. Table 3 presents the baseline specification controlling for year and labor market greater region fixed effects. This provider-level analysis permits assessment of service volumes, treatment intensity, and patient composition among GPs who absorb displaced demand, testing hypotheses H3 (capacity constraints) and H4 (patient collective shifts).

3.3.1 Low Density Regions

The provider-level results for low density LMRs reveal significant but incomplete absorption of displaced demand, consistent with binding capacity constraints. In the contemporaneous period, practice closures are positively associated with patient numbers among remaining providers (coefficient: 0.45, 95% CI $[0.05, 0.85]$, $p = 0.027$), indicating that a one percentage point increase in regional closure intensity corresponds to approximately a 0.5 percentage point increase in individual provider patient loads. Consultation volumes exhibit a similar response pattern (coefficient: 0.56, 95% CI $[0.14, 0.99]$, $p = 0.011$), as do physician services volumes measured in CHF billed (coefficient: 0.44, 95% CI $[0.03, 0.85]$, $p = 0.035$).

We find no statistically significant changes in consultations per patient ($p = 0.223$) or physician services per patient ($p = 0.848$), indicating that providers maintain their baseline service delivery patterns rather than intensifying care for existing patients. Similarly, mean patient age remains stable ($p = 0.451$), showing no systematic shift in age composition.

However, we observe a statistically significant reduction in the share of patients with an inpatient stay in the previous year (coefficient: -0.04, 95% CI $[-0.09, -0.01]$, $p = 0.007$) which is a key morbidity indicator. This compositional shift provides some evidence for hypothesis H4, suggesting that capacity-constrained providers may engage in selective patient acceptance, potentially prioritizing healthier cases while deferring or rejecting more complex patients

Table 3: Estimation Results Base Model: controlled for year and labor market greater region

	log(nr pat)	log(nr cons)	log(volume)	log(cons per pat)	log(vol per pat)	pat age	pat hosp
Panel A: Tertile 1 (Low Density)							
β_1, contemporary	0.405	0.562*	0.441*	0.157	0.036	0.451	-0.042
	(0.067)	(0.011)	(0.035)	(0.223)	(0.774)	(0.848)	(0.062)
β_2, leading	0.450*	0.064	0.036	-0.386	-0.414	-2.312	-0.050**
	(0.027)	(0.885)	(0.938)	(0.256)	(0.287)	(0.180)	(0.007)
Observations	2422	2422	2422	2422	2422	2422	2422
R-squared	0.075	0.059	0.057	0.049	0.076	0.103	0.056
Panel B: Tertile 2 (Medium Density)							
β_1, contemporary	0.028	0.234	0.151	0.206	0.123	1.363	-0.008
	(0.933)	(0.478)	(0.627)	(0.152)	(0.453)	(0.718)	(0.767)
β_2, leading	-0.267	-0.058	-0.109	0.209	0.158	3.643	0.030
	(0.466)	(0.874)	(0.749)	(0.130)	(0.382)	(0.310)	(0.169)
Observations	5382	5382	5382	5382	5382	5382	5382
R-squared	0.032	0.024	0.018	0.039	0.044	0.036	0.017
Panel C: Tertile 3 (High Density)							
β_1, contemporary	-0.266	-0.516**	-0.379*	-0.250*	-0.113	1.687	0.017
	(0.158)	(0.002)	(0.015)	(0.024)	(0.484)	(0.466)	(0.477)
β_2, leading	-0.068	-0.444	-0.308	-0.369***	-0.240	-1.193	-0.009
	(0.816)	(0.088)	(0.120)	(0.000)	(0.099)	(0.575)	(0.747)
Observations	10197	10194	10197	10194	10197	10197	10197
R-squared	0.053	0.087	0.027	0.062	0.039	0.038	0.018

p-values in parentheses, **standard errors**: clustered at LMR level, **significance levels**: *** p<0.001, ** p<0.01, * p<0.05

nr of pat: number of patients, **nr of cons**: number of consultations, **volume**: volume of physician services, **cons per pat**: consultations per patient, **vol per pat**: volume of physician services per patient, **pat hosp**: share of patients with an inpatient stay at t-1

requiring intensive post-hospitalization follow-up care. This mechanism raises concerns that the most vulnerable patient groups may be disproportionately affected by closures in undersupplied areas.

3.3.2 Medium Density Regions

Provider-level estimates for medium density LMRs yield null results across all outcome variables, paralleling the regional-level findings. The coefficients for patient numbers ($p = 0.933$), consultation volumes ($p = 0.478$), and physician services volumes ($p = 0.627$) are statistically indistinguishable from zero in both contemporaneous and leading periods. Similarly treatment intensity metrics like consultations per patient ($p = 0.152$) and physician services per patient ($p = 0.453$) show no significant deviations.

Patient composition measures remain stable, with no detectable changes in mean patient age ($p = 0.718$) or hospitalization rates ($p = 0.767$). These null findings indicate that in medium density regions, patient redistribution following closures occurs diffusely across a sufficient number of remaining providers such that marginal increases in individual workload remain statistically undetectable. The stability across all metrics confirms that medium density regions possess adequate spare capacity to absorb practice closures without imposing measurable burdens on individual remaining practitioners or altering their patient case mix.

3.3.3 High Density Regions

In high density LMRs, the provider-level analysis reveals statistically significant negative associations that mirror the regional-level patterns, reinforcing the interpretation of organizational transitions rather than competitive expansion. Consultation volumes among remaining providers decline by approximately half a percent per percentage point of closure intensity (coefficient: -0.52, 95% CI $[-0.84, -0.20]$, $p = 0.002$), while physician services volumes show a similar reduction (coefficient: -0.38, 95% CI $[-0.68, -0.08]$, $p = 0.015$). These volume contractions are accompanied by a significant decline in treatment intensity, with consultations per patient decreasing (coefficient: -0.25, 95% CI $[-0.46, -0.04]$, $p = 0.024$), though physician services per patient show no significant change ($p = 0.484$).

Patient numbers exhibit a negative point estimate that approaches conventional significance thresholds (coefficient: -0.27, $p = 0.158$), suggesting a tendency toward reduced practice size among remaining providers. Leading period effects for consultations (coefficient: -0.44, $p = 0.088$) and physician services (coefficient: -0.31, $p = 0.120$) reinforce this pattern, though with slightly weaker statistical support.

Regarding patient composition, we observe no substantial shifts in mean patient age ($p = 0.466$) or hospitalization rates ($p = 0.477$), indicating that the volume reductions are not accompanied by systematic changes in case mix complexity.

3.4 Provider Level Extended Regression Model

The baseline specification on the provider level yields relatively low R-squared values, ranging from 0.018 to 0.087 across outcomes and density tertiles (Table 3). To investigate the sensitivity of results to observed provider heterogeneity, we estimate an extended model including controls for physician sex, practice age, and FSO municipality classification (Table 10 in the Appendix). The inclusion of these practice characteristics substantially increases explanatory power, raising R-squared values to the range of 0.05 to 0.15 for most specifications.

Critically, the extended model reveals that in the low GP density tertile, all volume effects become statistically insignificant with the exception of the hospitalization rate coefficient, which remains significant and negative. The consultation volume estimate retains a positive coefficient of 0.40 but narrowly misses conventional significance thresholds ($p = 0.058$). This dynamic suggests that individual provider capacity to absorb displaced patients depends heavily on practice-specific characteristics—such as physician age, gender, and local setting—rather than being a purely mechanical function of regional density. In contrast, the negative associations in high density regions persist and remain statistically significant in the extended specification, reinforcing the interpretation that these effects reflect structural organizational dynamics rather than demand-side adjustments.

4 Discussion

This chapter interprets the empirical findings presented in the previous section in light of the conceptual framework and hypotheses developed in Section 2.2. It first discusses the evidence for each of the four main hypotheses (H1–H4) across GP density regimes. It then examines the structural dynamics of physician exits, addresses methodological considerations and limitations, and situates the results within the existing literature. Finally, it derives implications for health policy and outlines directions for future research.

4.1 Discussion of Main Hypotheses

4.1.1 H1: Closures reduce aggregate primary care volumes in undersupplied regions

H1 stated that in undersupplied markets, practice closures would lead to genuine reductions in aggregate primary care provision at the regional level. The regional estimates for low density LMRs show that closure intensity is associated with statistically significant reductions in both consultation volumes and total physician services, with elasticities close to unity. Taken together, these results are consistent with the interpretation that remaining providers are unable to fully offset the loss of capacity following closures in low density regions.

The provider-level estimates support this interpretation to the extent that remaining GPs in low density regions exhibit higher patient numbers and service volumes following closures. However, the magnitude of these increases falls short of a full one-to-one compensation of the lost supply. The extended provider-level model, which conditions on physician sex, practice age, and practice location, attenuates these volume effects and renders them statistically insignificant, while leaving the negative association with the share of recently hospitalized patients intact. This pattern, together with the robustness checks based on alternative inactivity thresholds reported in Appendix D, suggests that H1 is broadly supported in low density regions, although the exact dose–response relationship is sensitive to the way provider heterogeneity is modelled.

4.1.2 H2: Patients substitute lost local services by seeking care in neighbouring regions

H2 hypothesized that patients would compensate for local closures by seeking care in neighbouring LMRs, leading to changes in the domestic shares of consultations and physician services. Across all density tertiles, the regional regressions do not identify statistically significant associations between closure intensity and the domestic shares of consultations or volumes. This holds in the baseline specification and in the sensitivity analyses using alternative closure thresholds in Appendix D, where the timing of the response shifts between contemporary and leading terms but the absence of systematic changes in domestic shares persists.

Taken together, these findings provide no empirical support for H2. Within the limits imposed by annual data and the LMR definition, the results are more consistent with the interpretation that closure-induced reductions in local primary care supply in low density regions are not systematically offset by cross-regional patient flows.

4.1.3 H3: Remaining providers expand activity in oversupplied markets

H3 predicted that in oversupplied markets, remaining providers would expand service volumes to capture patients displaced by closures. Contrary to this expectation, both regional and provider-level estimates in medium and high density LMRs indicate no or negative associations between closure intensity and primary care volumes, as well as reductions in consultations per patient. These patterns are robust to the inclusion of practice-level controls in the extended provider specification and to changes in the closure definition documented in Appendix D. The results for medium density regions suggest that the system can fill the gap of exiting GPs but the required adjustment on the provider level might be too small to be statistically significant.

Several interpretations of the negative effects in high density regions are possible. Arguably the most plausible explanation in light of the data structure, is that the observed response reflects organisational restructuring rather than genuine capacity reductions: physicians leaving solo practices to join group practices or mixed entities may disappear from the analysed provider sample while continuing to deliver primary care elsewhere. Because the claims data do not

systematically identify primary care group practice entities, such transitions are not directly observable. Under this interpretation, the estimates for high density regions capture a contraction of measured solo-practice activity rather than a contraction of overall primary care supply. Consequently, H3 is not supported in its simple competitive form; instead, the evidence points to more complex organisational dynamics in saturated markets.

4.1.4 H4: Practice closures induce shifts in the patient collective of remaining providers

H4 stated that practice closures would lead to systematic shifts in the patient collective of remaining providers, particularly along morbidity dimensions. The results provide mixed evidence. In low density regions, the provider-level regressions show a statistically significant decline in the share of patients with an inpatient stay in the previous year, while average patient age remains unchanged. This pattern is robust to the inclusion of practice-level controls and to alternative inactivity thresholds, as shown in Appendix D. A plausible interpretation is that remaining providers, facing capacity constraints, selectively absorb patients with lower coordination needs and less complex care requirements, while more complex, recently hospitalised patients are less likely to be taken on.

In medium and high density regions, by contrast, there is no evidence of systematic shifts in patient age or hospitalization history associated with closure intensity. Overall, H4 thus receives partial support: compositional changes in the patient collective emerge in low density regions, whereas the patient mix of remaining providers appears more stable in medium and high density markets.

4.2 Interpretation of Heterogeneity by GP Density

Building on the hypothesis-based discussion, the results highlight that the supply-side response to practice closures is fundamentally shaped by regional provider density and the associated capacity regime.

4.2.1 Low Density Regions: Binding Capacity Constraints

In low density regions, the combination of large average patient lists, lower revenue and consultations per patient, and the regression-based evidence of reduced

regional volumes following closures points to a regime of binding capacity constraints. Remaining physicians appear to operate close to their feasible workload even before closures occur. In this setting, closures lead to a reduction in aggregate primary care provision, incomplete absorption of displaced demand by remaining providers, and changes in the composition of the patient collective along morbidity dimensions. The extended provider-level model indicates that the capacity to absorb additional patients is strongly conditioned by practice characteristics, underscoring that regional density captures only part of the relevant supply constraints.

4.2.2 Medium Density Regions: Market Flexibility

In medium density regions, the absence of statistically significant associations at both the regional and provider levels is consistent with a regime characterised by slack capacity. Closures do occur, but the redistribution of patients appears to be spread across a sufficiently large number of remaining providers such that marginal changes in workload and case mix at the individual level are not statistically detectable. Within the resolution of the data and model, medium density regions thus resemble a flexible equilibrium in which local shocks are absorbed without measurable frictions in primary care utilisation.

4.2.3 High Density Regions: The High-Density Puzzle

The negative volume and intensity effects observed in high density regions constitute a salient deviation from standard competitive predictions. Rather than expanding output to absorb freed-up demand, remaining providers in the analysed solo-practice sample exhibit contractions in both total and per-patient activity. In light of the descriptive evidence on practice forms and the limitations of the claims data in capturing group practices, a natural interpretation is that these patterns reflect structural reorganisation in saturated markets. As solo practitioners exit, either through retirement or through transitions into group settings, the measured solo-practice segment shrinks, while a growing share of primary care is likely provided by mixed or group entities outside the scope of the analytic unit. This interpretation is consistent with the robustness of the high-density findings across specifications, but it also underscores that the estimated effects for these regions should be interpreted as responses of the solo-practice segment

rather than of the entire primary care market.

4.3 Structural Dynamics of Physician Exits

Beyond the immediate market response, the descriptive analysis highlights three structural dynamics with long-term implications for the Swiss healthcare system.

4.3.1 Exogeneity of Closures

The age profile of closing physicians(averaging between 62 and 64 years) strongly supports the assumption that practice closures are driven by retirement decisions rather than local demand shocks. This "retirement-driven" nature reinforces the validity of the difference-in-differences design, as the timing of the treatment is likely exogenous to contemporaneous fluctuations in patient health or regional economic conditions. One option for further research could be to focus solely on retirement exits to strengthen the identification, as a generalization still would be valid.

4.3.2 The "Phasing Out" Phenomenon

The analysis of practice activity over time reveals a distinct "phasing out" behaviour. Closing physicians begin to reduce service intensity and patient volumes up to two years prior to official inactivity. While this gradual withdrawal allows for a smoother transition for some patients, it poses a potential risk to quality of care. If physicians reduce diagnostic intensity or follow-up frequency in their final years (as suggested by the divergence between patient counts and service volumes), patients may experience a subtle erosion of care quality long before the actual closure occurs. On the other side due to these patterns, it should be possible to predict practice closures with routinely collected data. Together with our industry partner tarifsuisse we implemented a logit and machine learning model to explore this possibility. Further information and preliminary results can be found in the Appendix E

4.3.3 Structural Change and Future Supply Security

The descriptive data documents a pronounced demographic divide. Low density regions are disproportionately served by male physicians, while high density regions attract a higher share of female practitioners. This structural imbalance suggests that the wave of retirements will hit rural areas hardest, precisely where capacity constraints are already binding. The feminisation of the workforce, often associated with a preference for part-time work and employment in group practices (Gisler et al., 2017), may further reduce the full-time equivalent capacity available to replace retiring solo practitioners in rural areas.

4.4 Methodological Considerations and Limitations

The empirical design is motivated by a causal interpretation of the estimated effects of practice closure intensity on regional and provider-level outcomes. Under the identification assumptions set out in Section 2.5.8, in particular conditional parallel trends and the predominantly retirement-driven timing of exits, the coefficients on closure intensity can be read as average marginal dose–response effects of closures on service volumes and patient composition. In this sense, the results for low-density regions are interpreted as evidence that practice closures causally reduce local primary care supply, while the estimates for medium- and high-density regions point to qualitatively different adjustment regimes.

At the same time, several methodological caveats qualify the strength of this causal interpretation. First, exogeneity of closure intensity is more demanding than exogeneity of timing and may be violated if regions with structurally weaker supply or different morbidity profiles both experience higher closure intensity and distinct underlying trends in utilization. Second, spillovers across LMRs, as discussed in the conceptual framework and in hypothesis H2, imply that the Stable Unit Treatment Value Assumption is only approximately satisfied. Although we control for inter-regional dynamics and find no effects, the estimated coefficients should still be viewed as net impacts within a region that may be partially offset by adjustments elsewhere. Third, the two-way fixed effects specification with continuous, repeated, and non-absorbing treatment, discussed in Section 2.5.9, may conflate contemporaneous and cumulative effects of closures, so that the coefficients capture an average response over the observed treatment histories rather than a simple one-time shock.

Beyond these identification issues, several data- and model-specific limitations merit attention. The stark negative effects on both volume and intensity in high-density regions warrant particularly cautious interpretation. One possibility is that remaining providers genuinely reduce activity following closures, reflecting competitive repositioning. However, as discussed in the regional analysis, a more plausible interpretation relates to measurement challenges. High-density regions are characterized by greater organisational complexity, with physicians transitioning between solo and group practice arrangements rather than truly exiting the market. When physicians exit solo practices to take up salaried employment in group practices at different locations, they disappear from the solo-practice sample while continuing to provide primary care services elsewhere. This selective exit of more active physicians from the measured solo provider population would mechanically reduce average volumes among remaining solo practitioners, even in the absence of any genuine reduction in overall primary care provision, while coinciding with intensified competition from newly established or expanding group practices.

Furthermore, the substantial difference in explanatory power (R^2) between the baseline and extended provider-level models suggests that individual practice heterogeneity plays a crucial role. In low-density regions, adding controls for physician characteristics and practice context renders all volume effects statistically in low density regions insignificant. This indicates that the capacity to absorb new patients is heavily dependent on individual factors—such as the age and gender of the remaining physician or the practice's local setting—rather than being a purely mechanical function of regional GP density. As a result, regional density tertiles capture only part of the relevant supply constraints, and residual heterogeneity at the provider level remains an important source of variation.

The robustness checks in Appendix D provide additional, although not definitive, support for the credibility of the empirical patterns in light of these limitations. Alternative inactivity thresholds yield qualitatively similar results across GP density tertiles: tightening the closure definition leaves both the sign and approximate magnitude of the main effects intact, whereas a more permissive threshold primarily shifts estimated responses from contemporaneous to leading terms without reversing their direction. Likewise, extended specifications that control for physician sex, practice age, and the FSO classification of the

practice location increase explanatory power and attenuate some volume effects in low-density regions, but they do not overturn the key result that responses differ systematically across local supply regimes. Taken together, these considerations suggest that the central heterogeneity patterns are not an artefact of arbitrary modelling choices, while the identification challenges imply that the estimates should be interpreted as plausibly causal but subject to residual bias.

4.5 Relation to the Literature

Our findings contribute a supply-side perspective to the existing patient-centered literature. Bischof and Kaiser (2021) documented that patients in Switzerland reduce GP visits and shift to specialists following closures. Our results show the mechanism behind this decline: in low density areas, it is not necessarily patient preference, but the binding capacity constraints of remaining GPs that force this reduction in primary care utilization. Similarly, our observation of "phasing out" behavior aligns with Zocher (2024), who documented anticipatory reductions in treatment intensity among exiting Austrian providers.

This thesis opens several avenues for future research. First, the counter-intuitive findings in high density regions highlight the need to better understand the dynamics of group practices. As the medical workforce shifts from solo practices to shared infrastructure models, future studies should employ data that allows for linking individual physicians to group entities to disentangle genuine market exits from organizational restructuring. Second the approach to think about initial market conditions should be applied in a different environment or be further finessed and combined with another data source (e.g. individual claims data) to show its validity. Lastly the main concern regarding validity of the results lies in the dynamics of the mixed group practices market. This is limitation is rooted in the process of data collection in swiss health insurance claims data but could be addressed with the approach used for calculating the cap of ambulatory physicians (*Höchstzahlen*) introduced by Jörg et al. (2022). This would allow to estimate the share of primary care services of mixed group practices and so gain a completer picture of the supply situation.

4.6 Policy Implications

The findings carry significant implications for health policy and workforce planning. First, the one-size-fits-all approach to regulation disregards the fundamental heterogeneity of the market. In low density regions, the primary challenge is absolute scarcity. Policies should focus on succession planning and incentives for practice takeovers to prevent the loss of capacity. The evidence that remaining doctors cannot absorb the surplus implies that relying on market self-regulation in rural areas is insufficient. Another way to tackle this problematic would be to financially incentivize GPs to serve under supplied regions, either on a system level (e.g. adjustment in tariffs) or regionally through policies of local administrations. Second, the potential "cream skimming" of healthy patients by remaining providers suggests a need for monitored care continuity, ensuring that multi-morbid patients are not left behind during practice transitions. Finally, the "phasing out" phenomenon suggests that monitoring systems could detect declining activity, allowing authorities to initiate succession support measures before the practice closes and continuity gaps emerge.

5 Conclusion

This thesis examines the supply-side effects of GP practice closures in Switzerland, contributing to a literature that has predominantly focused on patient-level responses. Using a difference-in-differences design with continuous treatment intensity on Swiss mandatory health insurance data (2018–2024), the analysis reveals pronounced regional heterogeneity in provider adjustments.

The findings confirm that practice closures have fundamentally different consequences across regional supply conditions. In low physician density regions operating near capacity, closures reduce aggregate primary care supply of consultations and billed volume by one percent per percentage point of closure intensity. Remaining physicians increase consultations by 0.5 percent but cannot fully compensate, leading to selective patient intake that disproportionately affects complex cases with prior hospitalizations. Medium density regions exhibit no measurable effects, indicating sufficient spare capacity for frictionless patient redistribution. High density regions show service volume declines likely reflecting organizational transitions rather than genuine care losses.

These results expose a critical policy gap: uniform approaches fail when regional supply conditions diverge sharply. In undersupplied areas, targeted succession planning and practice takeover incentives are essential to prevent systematic erosion of primary care access. Evidence of selective patient acceptance underscores the need for mechanisms ensuring care continuity for multimorbid patients during practice transitions. Coordinated workforce planning that accounts for regional heterogeneity is imperative to address the demographic challenge—every second Swiss physician is over 50 years old.

Future research should pursue three directions. First, refined data linking individual physicians to group practice entities would permit differentiation between organizational restructuring and genuine market exits in high-density regions. Second, newer econometric methods for staggered, continuous, and non-absorbing treatments could separate cumulative from contemporaneous effects. Third, the conceptual framework of regional capacity conditions developed here could be validated as an empirical mechanism for classifying under- versus oversupply in other healthcare systems.

This thesis demonstrates that practice closures are not homogeneous market shocks but trigger context-dependent adjustment processes whose consequences for care quality and access depend fundamentally on local capacity conditions. This insight should inform evidence-based, regionally differentiated health policy.

References

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

Bischof, T., & Kaiser, B. (2021). Who cares when you close down? The effects of primary care practice closures on patients. *Health Economics*, *30*(9), 2004–2025. https://doi.org/10.1002/hec.4287

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024). Difference-in-differences with a continuous treatment [NBER Working Paper No. 32117]. *Working Paper*.

De Pietro, C., Camenzind, P., Sturny, I., Crivelli, L., Edwards-Garavoglia, S., Spranger, A., Wittenbecher, F., & Quentin, W. (2015). Switzerland: Health system review. *Health Systems in Transition*, *17*(4), 1–288.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169

de Chaisemartin, C., & D'Haultfœuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, *26*(3), C1–C30. https://doi.org/10.1093/ectj/utac017

FMH. (2025). Fmh-ärztestatistik 2024 – tiefe grundversorgerdichte. *Schweizerische Ärztezeitung*, *106*(11-12), 30–36. Retrieved August 18, 2025, from https://www.fmh.ch/politik-medien/schweizerische-aerztezeitung/ausgabe-11-12/fmh-aerztestatistik.cfm

Gisler, L. B., Bachofner, M., Moser-Bucher, C. N., Scherz, N., & Streit, S. (2017). From practice employee to (co-)owner: Young GPs predict their future careers: A cross-sectional survey. *BMC Family Practice*, *18*(1), 12. https://doi.org/10.1186/s12875-017-0591-7

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Hjalmarsson, L., Kaiser, B., & Bischof, T. (2024). The impact of physician exits in primary care: A study of practice handovers. *Health Policy*, *135*, 104867. https://doi.org/10.1016/j.healthpol.2023.104867

Hostettler, S., & Kraft, E. (2018). FMH-Ärztestatistik 2017 – aktuelle Zahlen. *Schweizerische Ärztezeitung*, *99*(13-14), 408–413.

Jörg, R., Kaiser, B., Tuch, A., & Widmer, M. (2022). *Regionale versorgungsgrade pro fachgebiet als grundlage für die höchstzahlen in der ambulanten ärztlichen versorgung* (Obsan Bericht 05/2022). Schweizerisches Gesundheitsobservatorium (Obsan). Neuchâtel. https://www.obsan.admin.ch/sites/default/files/2022-11/Obsan_05_2022_BERICHT.pdf

Merçay, C. (2016). *Expérience de la population âgée de 18 ans et plus avec le système de santé – situation en Suisse et comparaison internationale* (Obsan Dossier No. 56). Swiss Health Observatory (OBSAN). Neuchâtel.

OECD & WHO. (2011). *OECD reviews of health systems: Switzerland 2011*. https://doi.org/10.1787/9789264120914-en

Senn, N., Ebert, S. T., & Cohidon, C. (2016). *La médecine de famille en Suisse: Analyse et perspectives sur la base des indicateurs du programme SPAM* (Obsan Dossier No. 55). Swiss Health Observatory (OBSAN). Neuchâtel.

Simonsen, M., Skipper, L., Skipper, N., & Thingholm, P. R. (2021). Discontinuity in care: Practice closures among primary care providers and patient health care utilization. *Journal of Health Economics*, *80*, 102551. https://doi.org/10.1016/j.jhealeco.2021.102551

Swiss Federal Statistical Office. (2021). Arbeitsmarktregionen 2020: Neue abgrenzung der funktionalen wirtschaftsräume der schweiz. Retrieved August 4, 2025, from https://www.bfs.admin.ch/bfs/en/home/statistics/regional-statistics/regional-portraits/functional-regions.html

Swiss Federal Statistical Office. (2025). Finanzierung des gesundheitswesens – 1995–2023. *Bundesamt für Statistik (BFS)*. Retrieved November 5, 2025, from https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/kosten-finanzierung.assetdetail.34788194.html

Zeller, A., & Giezendanner, S. (2020). Aktuelle daten zu arbeitszeiten, arbeitsgewohnheiten, motivation und arbeitspensen: Resultate der 4. workforce studie. *Primary and Hospital Care – Allgemeine Innere Medizin*, *20*(11), 325–328. https://doi.org/10.4414/phc-d.2020.10311

Zocher, K. (2024). Exiting primary care providers. *Health Economics*, *33*(5), 1033–1054. https://doi.org/10.1002/hec.4800

A Data Sources

A.1 National Provider Registry

The National Provider Registry (ZSR) is maintained by SASIS AG and serves as the authoritative register of licensed healthcare providers operating within the Swiss healthcare system. This registry contains comprehensive information on provider credentials, specializations, and institutional affiliations, enabling the identification and classification of service providers across the country. Additional information for personal Registry-IDs include date of birth and sex. The ZSR is not publicly available and access is granted conditional on approval of MHI insurers.

A.2 SASIS Data Pool

The SASIS data pool (DP) comprises detailed administrative claims data of services provided by healthcare providers in Switzerland. It constitutes a full census of inpatient and outpatient healthcare services billed over MHI, including comprehensive metrics at the provider level. Data include service volumes measured as billed claims, patient counts, case counts, and consultation volumes for both inpatient and outpatient settings. The patient population is stratified by key demographic characteristics including age, sex, and commune of permanent residence, enabling granular analysis of service utilization patterns across population subgroups per provider. The data pool is not publicly available and access is granted conditional on approval of MHI insurers.

A.3 FSO Labor Market Regions

Labor market regions (Arbeitsmarktregionen, LMR) are defined by the Swiss Federal Statistical Office (FSO) and represent functional spatial units based on commuting patterns and labor market integration. These regional classifications are employed to capture the geographic accessibility and demand dynamics of healthcare services, accounting for variations in supply and utilization across distinct labor markets within Switzerland.

A.4 FSO Permanent Resident Population

The permanent resident population data from the FSO comprises official population statistics for Switzerland derived from the statistical register (STATPOP). This data source encompasses all persons with permanent residence status in Switzerland, including Swiss nationals and foreign residents with long-term residence permits (duration exceeding twelve months). Population counts are disaggregated by standard demographic attributes including age, sex, and commune of permanent residence.

A.5 FSO Swiss Official Commune Register

The Swiss official commune register, maintained by the Federal Statistical Office, constitutes the authoritative nomenclature for all administrative municipalities and territorial divisions within Switzerland. This register provides the geographic and administrative framework for data aggregation and geographic stratification across all datasets employed in this analysis.

B Definitions

This appendix provides precise definitions of the key outcome variables employed in the empirical analysis. All definitions are derived from the SASIS data pool documentation (Richtlinien für die Datenlieferung an den SASIS-Datenpool, Version 10.6, valid from 01.01.2024), which governs the standardized reporting protocols for mandatory health insurance claims data in Switzerland.

B.1 Number of Patients

The number of patients represents the count of unique insured persons who received healthcare services from a specific provider (identified over the ZSR-ID) during a calendar year. Following the statistical case definition established by SASIS AG, a patient is classified as an *Erkrankte Person* if they generated at least one reimbursable claim that was processed and paid by the MHI within the reference period.

This definition captures only services covered under mandatory health insurance. Privately financed consultations, out-of-pocket payments not submitted for reimbursement, and services provided under supplementary insurance (VVG) are excluded from the count unless they are formally billed through the MHI system.

The number of patients serves as the primary measure of provider caseload and represents the extensive margin of healthcare utilization. It is less susceptible to physician-induced variation than consultation counts or billed service volumes, as it reflects discrete patient-provider relationships rather than the frequency or intensity of services rendered.

B.2 Number of Consultations

The number of consultations quantifies the frequency of distinct provider-patient interactions within the practice setting during the reference period. A consultation is defined as a temporally contiguous episode of medical advice, treatment, or examination (Sitzung) delivered by a healthcare provider (physician, physiotherapist, or other licensed practitioner) to one or more patients within the practice facility.

The consultation count captures the occurrence of provider-patient contacts within the practice, normally excluding home visits from this specific metric. For our analysis we add home visits to the consultations to measure all patient encounters. The counting rule stipulates that multiple services delivered during a single contiguous encounter are aggregated into one consultation event.

Consultation counts are derived from the presence of specific trigger positions in the TARMED fee schedule, which designate primary care encounters eligible for reimbursement under MHI. It is a key indicator of practice activity and workload, though it does not account for the complexity or duration of individual consultations.

B.3 Volume of Physician Services

The volume of physician services represents the aggregate monetary value of healthcare services billed and reimbursed through the mandatory health insurance system during the reference period. It is measured in Swiss Francs (CHF) and is part of the gross insurance benefit.

The gross insurance benefit encompasses all costs covered by MHI, it includes:

- Physician consultations and treatments reimbursed under the TARMED fee schedule which we refer to as volume of physician services
- Laboratory analyses, pharmaceutical dispensations, and ancillary services billed as part of physician-led care
- Medical aids (MiGeL) and materials provided in the outpatient setting

We focus on the volume of physician services and exclude other services as laboratory analyses and drug dispensations to closer capture the capacity of a physician. For example a physician may theoretically dispense almost an infinite amount of drugs in a year but he cannot bill more consultation minutes than he actually has working hours. This also better captures the primary care aspect of the service which relies heavily on time spent for consultations.

The volume of physician services serves as the primary measure of practice revenue and economic scale. When normalized by patient count or consultation count, it provides indicators of service intensity (e.g., revenue per patient or per consultation), which reflect the complexity, resource intensity, and reimbursement profile of the care delivered. This variable is central to understanding the financial sustainability of practices and the economic implications of practice closures for remaining providers.

B.4 Hospitalization Rate

The hospitalization rate represents the proportion of a provider's patient collective that experienced at least one inpatient stay in a hospital or nursing home during the preceding calendar year. It is constructed from the morbidity indicator *Spital-/Pflegeheimaufenthalt im Vorjahr*, which has been integrated into the SASIS data pool since January 1, 2012, and is employed in Switzerland's risk adjustment mechanism (Risikoausgleich) to account for variation in patient health status across insurers.

C Specification of extended Provider Level Model

$$
\begin{aligned}
Y_{irt} = \beta_0 &+ \beta_1 \text{ClosureRate}_{rt} + \beta_2 \text{ClosureRate}_{r,t+1} \\
&+ \text{sex}_i + \text{practice age}_i + \text{FSOclass}_i \\
&+ \delta_R + \tau_t + \varepsilon_{irt}
\end{aligned} \tag{3}
$$

The specification includes provider-level controls: sex and FSO classification are entered as categorical fixed effects, while practice age is specified as a continuous covariate. Regional effects (δ_R) and year fixed effects (τ_t) account for time-invariant heterogeneity across LMRs and secular trends, respectively.

D Sensitivity Analysis

The purpose of this section is to assess the robustness of the main results with respect to the definition of practice closures and the treatment of mixed group practices. First, alternative inactivity thresholds are considered to evaluate whether the empirical findings are sensitive to the operationalization of closures. Second, the importance of mixed group practices over time is examined, as these providers are excluded from the analysis and may therefore affect the interpretation of the results.

D.1 Alternative Thresholds for defining Practice Closures

The baseline definition classifies a GP practice as inactive once its annual activity falls below 50 patients and CHF 100,000 in billed physician services. To evaluate the sensitivity of the results to this choice, two alternative thresholds are employed: a higher threshold (100 patients and CHF 200,000) and a lower threshold (25 patients and CHF 50,000).

Table 4 documents how these alternative thresholds affect the number and timing of identified closures over the study period. The total number of closures between 2020 and 2024 varies only marginally across specifications. Instead, the thresholds mainly determine *when* a practice is classified as inactive, not *whether* it is eventually counted as a closure. This pattern is consistent with the phasing-out behaviour documented in Section 3.1, where closing practices

Table 4: Identified closures by year and different threshold levels

Threshold	2020	2021	2022	2023	2024	Total
High (200'000 CHF & 100 pat)	315	283	238	198	323	1357
Applied (100'000 CHF & 50 pat)	292	300	263	216	322	1393
Low (50'000 CHF & 25 pat)	237	311	288	238	329	1403

gradually reduce their billing activity in the years preceding complete inactivity. Consequently, the choice of threshold primarily shifts the assigned closure year within a narrow window rather than altering the set of exiting practices.

D.2 Estimation Results with alternative Thresholds for Closure

The robustness checks using alternative inactivity thresholds confirm that the main findings are not driven by the particular threshold chosen, but they do affect the magnitude and timing of the estimated effects.

Using the higher inactivity threshold yields estimates that are very similar to the main specification in both sign and magnitude across all GP density tertiles (Tables 5 and 6). The regional-level results continue to show that closures in low-density labour market regions are associated with a reduction in aggregate consultation volumes and billed physician services, while medium-density regions exhibit no statistically significant changes. In high-density regions, the negative associations with regional volumes persist. At the provider level, remaining GPs in low-density regions still expand their patient lists and service volumes in response to closures, whereas providers in high-density regions experience declines in consultations and billed volumes. Overall, tightening the inactivity criterion leaves the qualitative conclusions unchanged and only modestly affects point estimates.

Table 5: Results on the LMR level with higher threshold for inactivity

	log(volume)	log(nr of cons)	% domestic vol	% domestic cons
Panel A: Tertile 1 (Low Density)				
β_1, contemporary	-1.4851*	-1.4733*	6.5674	6.4554
	(0.012)	(0.012)	(0.382)	(0.386)
β_2, leading	-1.2285	-1.1825	2.4611	2.4920
	(0.091)	(0.094)	(0.746)	(0.737)
Observations	136	136	136	136
R-squared	0.625	0.620	0.430	0.443
Panel B: Tertile 2 (Medium Density)				
β_1, contemporary	-1.2710	-1.0404	9.5337	9.9688
	(0.344)	(0.435)	(0.208)	(0.184)
β_2, leading	-0.2315	-0.0196	-9.7811	-9.1535
	(0.853)	(0.987)	(0.465)	(0.509)
Observations	136	136	136	136
R-squared	0.705	0.711	0.513	0.508
Panel C: Tertile 3 (High Density)				
β_1, contemporary	-5.5662**	-4.7504**	-36.7910	-39.2941
	(0.001)	(0.002)	(0.242)	(0.209)
β_2, leading	-3.5858**	-3.2155**	-27.6302	-28.7008
	(0.008)	(0.004)	(0.339)	(0.321)
Observations	132	132	132	132
R-squared	0.777	0.757	0.032	0.031

Notes: p-values in parentheses. **Significance levels**: *** p<0.001, ** p<0.01, * p<0.05. **volume**: volume of physician services; **nr of cons**: number of consultations; **% domestic vol**: share of physician services volume in the LMR which is consumed by inhabitants of that LMR; **% domestic cons**: share of consultations in the LMR which is consumed by inhabitants of that LMR.

Table 6: Results on the provider level with higher threshold for inactivity

	log(nr pat)	log(nr cons)	log(volume)	log(cons per pat)	log(vol per pat)	pat age	pat hosp
Panel A: Tertile 1 (Low Density)							
β_1, contemporary	0.1846	0.3129	0.2251	0.1282	0.0405	1.8341	-0.0397
	(0.304)	(0.092)	(0.225)	(0.319)	(0.747)	(0.422)	(0.095)
β_2, leading	0.3669*	0.3778*	0.3674*	0.0109	0.0005	-1.0201	-0.0326
	(0.024)	(0.028)	(0.012)	(0.921)	(0.997)	(0.580)	(0.093)
Observations	2350	2350	2350	2350	2350	2350	2350
R-squared	0.080	0.065	0.062	0.055	0.085	0.115	0.069
Panel B: Tertile 2 (Medium Density)							
β_1, contemporary	0.1338	0.3511	0.3552	0.2174	0.2215	-3.5368	-0.0416
	(0.698)	(0.340)	(0.287)	(0.225)	(0.228)	(0.264)	(0.111)
β_2, leading	0.2566	0.3689	0.5003	0.1123	0.2437	-4.3310	-0.0266
	(0.525)	(0.389)	(0.221)	(0.472)	(0.265)	(0.255)	(0.481)
Observations	5529	5529	5529	5529	5529	5529	5529
R-squared	0.052	0.044	0.025	0.039	0.047	0.038	0.020
Panel C: Tertile 3 (High Density)							
β_1, contemporary	-0.3214*	-0.6013***	-0.3844*	-0.2799**	-0.0630	1.5976	0.0247
	(0.028)	(0.000)	(0.015)	(0.006)	(0.578)	(0.461)	(0.194)
β_2, leading	-0.2142	-0.5214**	-0.4213**	-0.3078***	-0.2071	-0.0872	0.0079
	(0.154)	(0.004)	(0.002)	(0.000)	(0.075)	(0.962)	(0.640)
Observations	9261	9260	9261	9260	9261	9261	9261
R-squared	0.075	0.117	0.042	0.070	0.050	0.046	0.025

p-values in parentheses, **standard errors**: clustered at LMR level, **significance levels**: *** p<0.001, ** p<0.01, * p<0.05

nr of pat: number of patients, **nr of cons**: number of consultations, **volume**: volume of physician services, **cons per pat**: consultations per patient, **vol per pat**: volume of physician services per patient, **pat hosp**: share of patients with an inpatient stay at t-1

Table 7: Results on the LMR level with lower threshold for inactivity

	log(volume)	log(nr of cons)	% domestic vol	% domestic cons
Panel A: Tertile 1 (Low Density)				
β_1, contemporary	-1.5803*	-1.5465*	14.3920	14.0388
	(0.027)	(0.025)	(0.189)	(0.198)
β_2, leading	-2.0309**	-1.9537**	1.1700	0.8813
	(0.007)	(0.009)	(0.878)	(0.905)
Observations	136	136	136	136
R-squared	0.617	0.616	0.297	0.314
Panel B: Tertile 2 (Medium Density)				
β_1, contemporary	-0.6019	-0.5838	-6.0047	-5.3243
	(0.663)	(0.659)	(0.634)	(0.672)
β_2, leading	-0.2670	-0.0933	0.8098	2.0847
	(0.852)	(0.946)	(0.953)	(0.879)
Observations	136	136	136	136
R-squared	0.721	0.724	0.503	0.500
Panel C: Tertile 3 (High Density)				
β_1, contemporary	-4.6210*	-4.0521*	-51.2057	-52.8208
	(0.021)	(0.042)	(0.133)	(0.119)
β_2, leading	-4.3302*	-3.8677*	-43.0699	-43.5031
	(0.030)	(0.033)	(0.178)	(0.176)
Observations	132	132	132	132
R-squared	0.766	0.752	0.046	0.045

Notes: p-values in parentheses. **Significance levels**: *** p<0.001, ** p<0.01, * p<0.05. **volume**: volume of physician services; **nr of cons**: number of consultations; **% domestic vol**: share of physician services volume in the LMR which is consumed by inhabitants of that LMR; **% domestic cons**: share of consultations in the LMR which is consumed by inhabitants of that LMR.

Table 8: Results on the provider level with lower threshold for inactivity

	log(nr pat)	log(nr cons)	log(volume)	log(cons per pat)	log(vol per pat)	pat age	pat hosp
Panel A: Tertile 1 (Low Density)							
β_1, contemporary	0.4762	1.0061**	0.8321*	0.5299**	0.3559	0.0976	-0.0358
	(0.115)	(0.007)	(0.033)	(0.008)	(0.080)	(0.981)	(0.269)
β_2, leading	0.3086	0.0091	-0.2554	-0.2998	-0.5640	-3.3406	-0.0470*
	(0.216)	(0.986)	(0.652)	(0.454)	(0.204)	(0.197)	(0.046)
Observations	2234	2233	2234	2233	2234	2234	2234
R-squared	0.074	0.059	0.049	0.042	0.055	0.118	0.066
Panel B: Tertile 2 (Medium Density)							
β_1, contemporary	0.1213	0.0845	-0.0983	-0.0540	-0.2196	2.3131	0.0040
	(0.704)	(0.799)	(0.767)	(0.782)	(0.301)	(0.559)	(0.928)
β_2, leading	-0.0457	-0.0908	-0.3436	-0.0567	-0.2979	1.1620	0.0059
	(0.903)	(0.803)	(0.369)	(0.729)	(0.151)	(0.748)	(0.844)
Observations	5908	5907	5908	5907	5908	5908	5908
R-squared	0.023	0.018	0.014	0.035	0.039	0.039	0.019
Panel C: Tertile 3 (High Density)							
β_1, contemporary	-0.2704	-0.4377*	-0.2982	-0.1619	-0.0278	2.9790	0.0241
	(0.128)	(0.019)	(0.087)	(0.116)	(0.837)	(0.332)	(0.451)
β_2, leading	-0.4235*	-0.7270***	-0.5665**	-0.2966**	-0.1431	1.4494	0.0143
	(0.022)	(0.000)	(0.002)	(0.007)	(0.291)	(0.584)	(0.603)
Observations	10678	10672	10678	10672	10678	10678	10678
R-squared	0.035	0.057	0.014	0.051	0.036	0.038	0.017

p-values in parentheses, **standard errors**: clustered at LMR level, **significance levels**: *** p<0.001, ** p<0.01, * p<0.05

nr of pat: number of patients, **nr of cons**: number of consultations, **volume**: volume of physician services, **cons per pat**: consultations per patient, **vol per pat**: volume of physician services per patient, **pat hosp**: share of patients with an inpatient stay at t-1

By contrast, the specification with the lower inactivity threshold amplifies some of the estimated effects and shifts part of the response from contemporaneous to leading terms (Tables 7 and 8). In particular, the regional coefficients on the lead treatment variable become statistically significant in low-density regions. This pattern is consistent with the notion that, under a more permissive threshold, practices are classified as inactive at an later stage of the phasing-out process. As a result, activity reductions and patient reallocation begin before the first year of recorded inactivity, which is then captured more strongly in the lead rather than the contemporaneous treatment term. While this affects the precise timing of the estimated response, the direction of effects by GP density tertile remains unchanged.

Taken together, the robustness checks indicate that the core empirical patterns are stable across alternative operationalizations of inactivity. The definition of the closure threshold mainly influences whether the supply shock is attributed to the year of first observed inactivity or to the preceding year, but does not overturn the main conclusions regarding heterogeneity across regional supply conditions.

D.3 Importance of mixed Group Practices over Time

A separate robustness concern arises from the increasing role of group practices, in particular mixed group entities that combine GPs and specialists under a single billing identifier. As discussed in Section 2.3, such providers are classified into a residual group practice category and are therefore excluded from the main analysis as they cannot be clearly attributed to primary or specialist care. Their evolution over time is documented in Figures 4 and 5.

Figure 4 shows that the number of mixed group practices, relative to all GP and group-practice providers, has increased steadily over the study period in all GP density tertiles. This growth is visible in low-, medium-, and high-density regions alike, indicating a structural shift away from traditional solo and small GP practices towards larger organizational units, often located in more urban or intermediary areas.

Figure 5 complements this picture on the revenue side of things. At the beginning of the observation period, mixed group practices bill roughly half the volume of pure GP practices in terms of physician services under mandatory

Figure 4: Relative importance of the number of mixed group practices compared to GPs

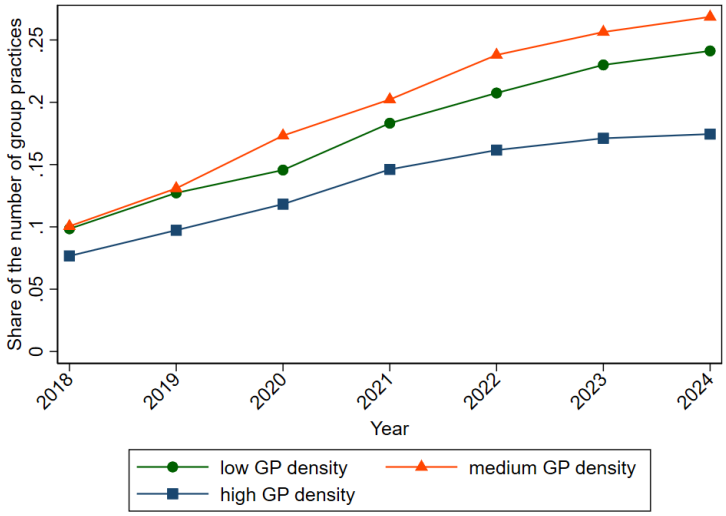
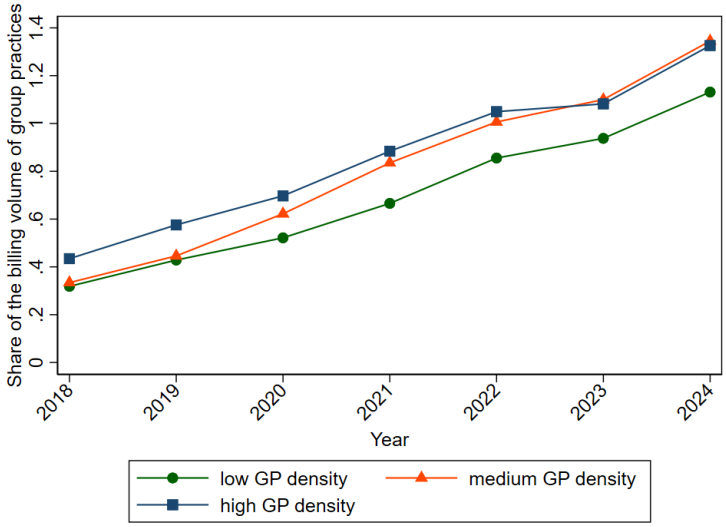


Figure 5: Relative importance of the billing volume of mixed group practices compared to GPs



health insurance. By 2024, their billed volume is approximately on par with that of GPs across all density regions. This convergence implies that mixed group practices have become as important as traditional GPs in the outpatient market.

The rising number and economic weight of mixed group practices therefore limits the external validity of the results, in particular in urban areas where mixed group practice models are most prevalent. The exclusion of these providers implies that the estimated effects pertain primarily to single-specialty GP practices and may not fully reflect the dynamics of the broader primary care landscape.

D.4 Discussion

The sensitivity analyses yield two main insights for the interpretation of the results.

First, the definition of the inactivity threshold for identifying practice closures does not appear to be a critical modelling choice. Alternative thresholds deliver qualitatively similar results, and the main conclusions regarding regional heterogeneity and provider responses are robust. The threshold primarily determines the exact timing of the recorded closure within the phasing-out trajectory, which is reflected in a reallocation of effects between contemporaneous and leading treatment coefficients. Substantively, the supply shock associated with a closure remains of similar magnitude, irrespective of the precise cut-off used to define inactivity.

Second, the analysis of group practices underlines that mixed group entities are becoming an increasingly important component of the outpatient sector. Their growing number and billing volume imply that excluding them from the treatment and outcome definitions may distort the measurement of both regional capacity and provider-level responses, especially in later years and in high-density markets. However, the extent and direction of this potential bias remain unclear, because the available data do not allow for a reliable decomposition of mixed group-practice activity into primary care and specialist services, nor for tracing how this composition varies over time or across regions.

Overall, the robustness checks suggest that the central findings of this thesis, namely the heterogeneous impact of practice closures across regional supply regimes and the constrained adjustment capacity in low-density areas, are not

driven by arbitrary modelling choices in the closure definition. At the same time, the increasing relevance of mixed group practices highlights an important structural limitation of the data and points to a key avenue for future research, requiring more granular information on the internal composition and functional role of group practices in primary care provision.

E Prediction of Practice Closures

E.1 Motivation and Objective

Early prediction of practice closures represents a valuable opportunity for health system planning and patient care continuity. Using claims data from billing systems, we can potentially identify practices at risk of closure in advance,ideally in the year before the event occurs. Due to the time lag in the data collection process this would allow us to at least identify in real time which practices are closing.

E.2 Methodological Approach

We employ two models that represent current standards in predictive classification: logistic regression and random forest algorithms. Both methods are applied to data aggregated at two temporal resolutions: annual and monthly. This dual-resolution strategy allows us to assess the trade-off between data granularity and predictive performance.

Our analysis targets two distinct prediction horizons: one year and two years in advance. For each scenario, we employ a balanced 50/50 split between training and validation data. The feature set includes trends in revenue and consultation volumes, with seasonal adjustment applied where appropriate (adjusted by AMR GR).

Model performance is evaluated using standard classification metrics: accuracy, precision, and AUC-ROC. These metrics together provide a comprehensive view of both overall performance and the model's ability to correctly identify at-risk practices.

E.3 Preliminary Findings

E.3.1 Annual Data: One-Year Prediction Horizon

Using logistic regression with annual data, we achieve approximately 50% sensitivity in predicting practice closures one year in advance. This result is encouraging as a starting point, substantially better than random assignment, yet it clearly indicates room for substantial improvement. The 50% detection rate suggests that roughly half of closures follow patterns observable in historical billing data, while the other half may stem from events not readily apparent in claims information.

E.3.2 Annual Data: Two-Year Prediction Horizon

Extending the prediction window to two years yields negligible predictive performance. Annual data simply lack sufficient information to forecast events so far into the future. Over a two-year period, numerous unmeasurable factors shift, including physician health status, family circumstances, financial conditions, and strategic decisions, making long-horizon prediction with aggregated annual data unfeasible.

This finding suggests that meaningful predictions require focus on the near term. The observable signals of impending closure appear concentrated in the months immediately preceding the event, not years earlier.

E.3.3 Monthly Data: Enhanced Temporal Resolution

Transitioning to monthly data represents a meaningful shift in analytical capacity. While monthly aggregation requires redefining the closure outcome with greater precision—identifying the specific month rather than just the year—it simultaneously provides substantially richer information about practice trajectories. We anticipate that monthly data will yield considerably improved predictions for several reasons:

- Monthly aggregation captures subtle, gradual changes in utilization patterns that are masked when data are collapsed into annual figures.
- The increased frequency of observations enables more refined detection of trend changes and inflection points.

- Finer temporal resolution facilitates feature engineering of moving averages, trend indicators, and momentum measures that characterize the pre-closure period.

We expect that monthly data will substantially enhance predictive performance, though implementation challenges remain around missing observations and seasonal adjustment procedures.

E.4 Understanding Heterogeneity in Closure Pathways

A critical insight from this analysis is that practice closures are not monolithic events. Different closures follow distinct pathways, and this heterogeneity directly shapes predictability. A substantial fraction of practice closures are preceded by observable, gradual reductions in activity. Over months or years, consultation volumes decline, revenues diminish, and overall practice size contracts. These patterns leave clear traces in billing data: declining utilization trends, shrinking patient populations, and reducing service provision. Such closures are fundamentally predictable. The signals exist in historical data. Appropriate statistical models can detect these patterns and forecast eventual cessation. This category likely encompasses many retirement-driven closures where physicians gradually wind down their practices over extended periods.

In contrast, many practice closures occur suddenly, with no advance warning discernible in claims data. A physician suffers acute illness or serious injury. A family emergency demands immediate departure. Unexpected personal circumstances force sudden retirement or relocation. These events generate little or no pre-closure signal in administrative data. Such closures are fundamentally unpredictable using retrospective claims data. No statistical model can forecast random health shocks or family emergencies. The absence of observable patterns reflects the genuine unpredictability of these events.

Among all closure types, retirement-driven practice closures present particularly strong opportunities for prediction. Several factors make retirements substantially more predictable than other closure scenarios: Retirement is concentrated among physicians in specific age ranges. Physician age is observable in claims data and should be a powerful predictor of retirement risk. Unlike sudden health shocks or emergencies, retirement is a planned decision. Physicians engage in deliberate planning processes—often spanning years—before

implementing retirement. This planning horizon may be reflected in gradual activity reductions: phased workload reduction, declining patient acquisition, or reduced service scope. Physicians may reduce hours, delegate certain services, or gradually transfer patients. These patterns are temporally extended and thus detectable through appropriate trend analysis. Taken together, these factors suggest that retirement-driven closures,which likely represent a substantial share of all practice cessations, should be highly predictable with well-designed models. A focused prediction system emphasizing demographic risk factors and activity trajectories could plausibly achieve much higher sensitivity specifically for retirement cases.

E.5 Limitations and Data Requirements

The current approach operates under several constraints that should be acknowledged. Claims data capture only utilization patterns and financial flows—not physician health status, subjective retirement intentions, practice finances, or succession planning efforts. These unmeasured factors likely influence closure timing substantially. Survey data or physician registry information could enhance predictive performance significantly but was not available for this analysis. Additionally, some physicians pursue phased retirement—gradually reducing hours while maintaining formal practice affiliation—creating ambiguity about the precise timing and nature of closure. Administrative definitions of closure may lag actual cessation of substantive clinical activity. Monthly data implementation introduces additional complexity: seasonal adjustment procedures must be carefully designed, missing observations require thoughtful handling, and the computational burden of ensemble methods increases substantially with temporal granularity.

E.6 Conclusion

This analysis demonstrates that practice closure prediction is feasible, particularly for closures preceded by observable gradual decline. With annual data, one-year-ahead predictions achieve approximately 50% sensitivity—a meaningful baseline that reflects the inherent split between predictable (gradual) and unpredictable (abrupt) closure pathways.

The transition to monthly data holds substantial promise for improving performance by capturing finer temporal dynamics. More importantly, this methodological framework enables targeted attention to high-opportunity prediction targets: retirement-driven closures that combine demographic identifiability, planned decision-making, and observable activity trajectories.

F Additional Tables and Figures

Here we present additional figures and tables supporting our methodological approach.

Table 9 characterizes closing GP practices by GP density tertiles over the period 2020–2023. The table documents the age profile of exiting physicians, practice age, and the distribution of closures across urban, intermediary, and rural municipalities, and is referenced in Section 3.1 to support the assumption that closures are predominantly retirement-driven and occur in all types of regions.

Figure 6 displays the spatial distribution of LMRs across the three GP density tertiles used in the heterogeneity analysis. As discussed in Section 2.5.7, this classification provides a relative measure of local supply conditions and underpins the interpretation of low-, medium-, and high-density regions as operating under distinct capacity regimes.

Figure 7 illustrates the extent to which LMRs function as effective healthcare catchment areas, by showing the share of outpatient services consumed by residents within their own region. This evidence, referenced in Section 2.5.2, supports the choice of LMRs as the relevant geographic unit for treatment assignment in the empirical analysis.

Table 10 reports provider-level estimation results from the extended specification that adds controls for physician sex, practice age, and FSO classification of the practice location. As discussed in Sections 3.3 and 4.1, these results show that accounting for observable provider heterogeneity increases explanatory power and leaves the main patterns in high-density regions intact, while attenuating some volume effects in low-density regions.

Table 9: Descriptive statistics of GP closures

Variable	low GP density	medium GP density	high GP density
Year			
2020	50 (28%)	99 (26%)	143 (28%)
2021	44 (25%)	120 (31%)	136 (27%)
2022	45 (25%)	92 (24%)	126 (25%)
2023	40 (22%)	75 (19%)	101 (20%)
Sex			
institution	6 (3%)	9 (2%)	7 (1%)
male	129 (72%)	279 (72%)	336 (66%)
female	44 (25%)	98 (25%)	163 (32%)
physician age	63.68 (8.50)	62.53 (9.54)	63.86 (10.56)
practice age	23.88 (12.44)	21.51 (12.62)	24.44 (12.64)
FSO classification of practice location			
urban	71 (40%)	229 (59%)	391 (77%)
intermediary	59 (33%)	99 (26%)	62 (12%)
rural	49 (27%)	58 (15%)	53 (10%)
N	179	386	506

[1] n (%); Mean (SD)

Figure 6: Distribution of labor market regions into GP density tertiles

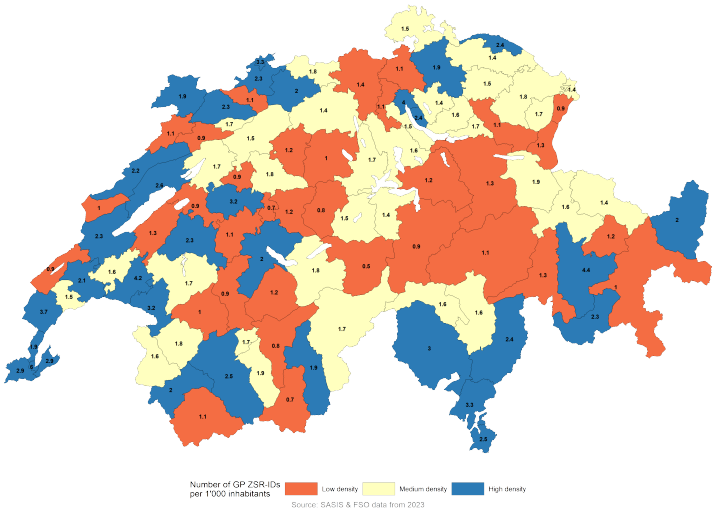


Figure 7: Labor market regions as a catchment area

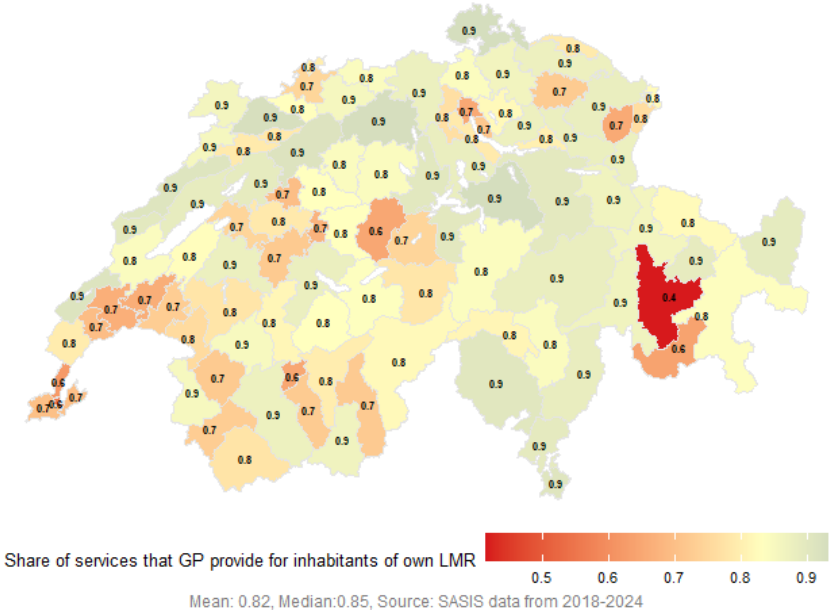


Table 10: Estimation results extended model: additionally controlled for sex, practice age and FSO classification

	log(nr pat)	log(nr cons)	log(volume)	log(cons per pat)	log(vol per pat)	pat age	pat hosp
Panel A: Tertile 1 (Low Density)							
β_1, contemporary	0.2997	0.4109	0.3093	0.1112	0.0096	-0.0821	-0.0489*
	(0.155)	(0.058)	(0.108)	(0.371)	(0.938)	(0.972)	(0.038)
β_2, leading	0.2869	-0.1565	-0.1742	-0.4434	-0.4611	-2.4866	-0.0581**
	(0.118)	(0.710)	(0.691)	(0.202)	(0.231)	(0.170)	(0.004)
Observations	2422	2422	2422	2422	2422	2422	2422
R-squared	0.134	0.126	0.141	0.069	0.094	0.147	0.072
Panel B: Tertile 2 (Medium Density)							
β_1, contemporary	-0.1848	-0.0636	-0.1958	0.1212	-0.0110	-0.3113	-0.0210
	(0.529)	(0.831)	(0.493)	(0.384)	(0.944)	(0.928)	(0.441)
β_2, leading	-0.5025	-0.3469	-0.4278	0.1556	0.0747	2.4815	0.0208
	(0.183)	(0.339)	(0.225)	(0.298)	(0.684)	(0.450)	(0.361)
Observations	5382	5382	5382	5382	5382	5382	5382
R-squared	0.128	0.109	0.120	0.047	0.056	0.071	0.025
Panel C: Tertile 3 (High Density)							
β_1, contemporary	-0.2150	-0.5012**	-0.3615	-0.2859**	-0.1466	0.8711	0.0159
	(0.116)	(0.004)	(0.062)	(0.002)	(0.259)	(0.722)	(0.520)
β_2, leading	-0.0655	-0.4699*	-0.3298*	-0.3966***	-0.2642*	-1.9545	-0.0110
	(0.734)	(0.024)	(0.036)	(0.000)	(0.016)	(0.357)	(0.678)
Observations	10197	10194	10197	10194	10197	10197	10197
R-squared	0.131	0.136	0.103	0.082	0.049	0.084	0.020

p-values in parentheses, **standard errors**: clustered at LMR level, **significance levels**: *** p<0.001, ** p<0.01, * p<0.05

nr of pat: number of patients, **nr of cons**: number of consultations, **volume**: volume of physician services,

cons per pat: consultations per patient, **vol per pat**: volume of physician services per patient