Incentives and interventions in primary care:

The effect on service provision, resource use and turnover

Ole Kristian Aars

Thesis submitted for the degree of Philosophiae Doctor

Department of Health Management and Health Economics

Faculty of Medicine

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-348-0351-2

Cover: UiO.
Print production: Graphic center, University of Oslo.

Summary

In many countries, general practitioners (GPs) play a fundamental role in primary healthcare services. Owing to their prominent role in healthcare, the increasing difficulty many Nordic countries experience in recruiting and retaining GPs is alarming. To meet increased patient demand, policymakers must understand how to use their limited resources to increase the supply of GP services. To help advance the evidence base, we investigate how incentives and interventions affects service provision, resource use and turnover amongst GPs.

In Paper 1, we develop a theoretical model to predict how different remuneration schemes and policy measures affect GP effort and the profession's overall attractiveness. The results show that none of the remuneration schemes increase both the profession's attractiveness and GPs' effort levels. We propose that improving working conditions is the only policy unequivocally positive for both these outcomes. In Paper 2, we empirically investigate the sources of earnings differentials and whether these translate into quitting decisions. Using registry data, we find that a 1% increase in income per consultation results in a one-percentage point reduction in the likelihood of quitting. We hypothesise that the sources of earnings differentials are higher effort and better system familiarity. In Article 3, we estimate the effect of an information campaign regarding the appropriate use of a double-consultation fee. We find that the campaign results in a 4–5 percentage point reduction in the use of such fees. In Article 4, we perform a field experiment using a nudging campaign to increase annual checkup attendance for people with type 2 diabetes. We find that sending an emails and SMS leads to a 4% increase in attendance.

Overall, our findings suggest that there is no silver bullet; financing systems that induce effort are unattractive to some doctors and more prone to opportunistic behaviour. The changing composition of GPs also reflects different preferences for organisational form, suggesting that a larger suite of contractual arrangements should be offered. Regardless, non-financial interventions will be necessary to ensure timely access to services and maintain collective trust in the system.

Norsk sammendrag

I mange land spiller fastleger en viktig rolle i helsetjenesten. De økende utfordringene med å rekruttere nye fastleger, og hindre at eksisterende slutter i jobben, er derfor urovekkende. For å møte den økende etterspørselen fra pasienter er det derfor viktig å forstå hvordan man kan øke tilbudet av fastlegetjenester med begrensede ressurser. For å øke forståelsen av tiltak, studerer vi hvordan insentiver og intervensjoner påvirker tilbudet av fastlegetjenester, ressursbruk og yrkestilfredshet.

I artikkel 1 utvikler vi en teoretisk modell for å vurdere hvordan ulike avlønningsordninger og intervensjoner påvirker fastleger tjenestetilbud, samt hvor attraktivt yrket vurderes. Resultatene viser at ingen avlønningsordninger gir positive utslag for både tjenestetilbudet og nytten til fastlegene. Bedre arbeidsvilkår er det eneste tiltaket som gir utelukkende positiv effekt på disse målene. I artikkel 2 undersøker vi kildene til ulik inntekt hos fastleger empirisk, og hvorvidt dette fører til at de slutter i jobben. Ved bruk av registerdata finner vi at en prosent økning i inntekt per konsultasjon fører til et prosent-poengs reduksjon i sannsynligheten for å slutte i jobben. En hypotese er at høyere innsats og bedre kunnskap om takst-systemet er en viktig driver for inntektsforskjellene. I artikkel 3 estimerer vi effekten av en informasjonskampanje vedrørende riktig bruk av en tidstakst. Vi finner at kampanjen resulterer i en 4-5 prosent-poengs reduksjon i bruk av taksten. I artikkel 4 gjør vi et felteksperiment for å studere effekten av en informasjonskampanje med formål om å øke bruk av årskontroll for personer med type-2 diabetes. Vi finner at informasjon via SMS og epost fører til en 4 prosent økning i oppmøte.

Våre funn tilsier at det ikke finnes et perfekt system for å organisere og avlønne fastleger; avlønningsordninger som gir insentiv om økt innsats, kan føre til at noen slutter, og andre handler opportunistisk. En endret sammensetning av fastleger reflekterer også ulike preferanser for organisasjonsform, noe som tilsier at ulike kontrakter burde tilbys. Uansett vil ikke-finansielle tiltak være nødvendig for å sikre rettidig tilgang til tjenester og opprettholde kollektiv tillit til systemet.

Acknowledgements

Working on this thesis has been a challenging and rewarding journey. Shortly after embarking on the work in January 2020, the COVID-19 pandemic hit. Although the lockdown situation was far from ideal, it presented some opportunities. I relocated to Tromsø during the first wave and spent time away from my PhD endeavour to help my previous employer, CEPI, work on the COVID-19 response. I returned to Oslo after the summer before moving back to Tromsø when the second wave hit in February 2021. This has been my permanent residence since, only interrupted by shorter periods in Oslo and an exchange to the University of Manchester. Finally, I was fortunate to teach a course in programming at the University of Tromsø in the autumn of 2022.

The pandemic and relocation were challenging to combine with advancing my research. However, with the help of numerous people, I am happy to be able to submit my thesis. First and foremost, I express my gratitude to Oddvar Kaarbøe, my main supervisor. I am extremely happy for the opportunity to work on a subject that has been an area of interest of mine for a long time. You have provided guidance and support when necessary and the freedom to pursue my interests. I also appreciate all the help I received from Tor Iversen, my second supervisor. I have always valued your advice and openness. The collaboration with my co-authors Igor Francetic, Geir Godager and Tron Anders Moger has also been important to me, and I've learned a lot from working with all of them. While cut short by my absence, I have also enjoyed the working life at the department of health and society. Daily virtual morning coffees with Adrian have helped keep my spirits high. I am especially grateful for the flexibility the department has shown me in relocating to Tromsø and pursuing other opportunities. Working from Tromsø would not have been possible without the generosity of the Business School at the University of Tromsø, which provided me with a desk to work from and colleagues to interact with. I convey my gratitude to Derek Clark for making this happen. I am also thankful to the Research Council of Norway, which provided the grant that allowed me to pursue this research.

Finally, I appreciate the support from my family, friends, and partner Oda, who have stood by me through my ups and downs.

List of papers

Paper 1:

Aars, O.K. and Kaarbøe, O., 2021. GP Recruitment and retention in the Nordic countries. *Nordic Journal of Health Economics*. *[Status: accepted]*

Paper 2:

Aars, O.K. and Francetic, I. 2023. "Just pay them more". Income generation and its effect on the retention of general practitioners in Norway. *Submitted to Health Economics [Status: submitted]*

Paper 3:

Aars, O.K., Godager, G., Kaarbøe, O.M. and Moger, T.A., 2022. Sending emails to reduce medical costs? The effect of feedback on general practitioners' claiming of fees. *Submitted to Journal of Experimental Economics and Behaviour [Status: revise & resubmit]*

Paper 4:

Aars, O.K. and Iversen, T., 2023. Does nudging increase the uptake of recommended services? *Submitted to International Journal of Health Economics and Management [status: under revision]*

Contents

1. Introduction

Primary healthcare has been considered the cornerstone of well-functioning healthcare systems since the Alma Ata Declaration in 1978 (World Health Organization (WHO), 1978). In many Western countries, the family doctor—or general practitioner (GP)—is the first point of contact for citizens' healthcare. Nordic countries, among others, are finding it increasingly difficult to retain and recruit GPs to maintain dependable access and quality care. This conflicts with the World Health Organization's (WHO) definition of the prerequisites for a well-functioning primary care system: the 'capacity to respond equitably and efficiently to the health needs of citizens…'(PAHO, 2023). Considering prevailing healthcare demands, we propose two primary strategies for augmenting general practice capacity: i) getting doctors to choose general practice as their career and ii) incentivising efficient supply services.

GPs are crucial in many health systems where patients rely on dependable access and quality care to lead healthy lives. GPs often have a gatekeeping role, with the discretion to grant sick leave, issue referrals to specialised treatment, and prescribe drugs and pharmaceuticals. They play a key role in preventive efforts through continuity of care, helping patients avoid severe health conditions that impede quality of life and are costly to the public. As such, the broader healthcare system and society rely heavily on their services.

The demand for GP services is significant, and it has only been exacerbated by ageing populations, widening responsibilities, and greater expectations regarding the range of services provided to patients. Therefore, governments face the challenge of designing GP systems that incentivise efficient supply of desired services without reducing the profession's attractiveness. However, these objectives may conflict, as policies that incentivise effort in the short term may also be those that drive people out of the profession in the long term. Moreover, in some instances, what 'desired services' comprise may differ in the eyes of GPs versus policymakers and may also depend on the GP system's design.

Economic incentives are often the starting point in designing a system to induce behaviours that align with desired goals, such as quantity, quality, or autonomy in providing services. The tasks GPs are responsible for may also affect which services are prioritised and whether these are deemed fulfilling for individual GPs. Adding to the challenge are the diverging preferences among GPs, who may assign different weights to the job's individual attributes.

One of the main characteristics of the market for health care is the existence of asymmetric information. Patients have less information than their GPs about optimal treatments, and payers may have limited

ability to affect the types of services GPs offer (Arrow, 1978). This may put patients at a disadvantage regarding access to care. In economics, this situation is often studied through agency theory, whereby a principal (the patient) is affected by the actions of an agent (the GP), who has more information than the principal. Payers can overcome challenges related to this inherent information asymmetry by incentivising GPs to act in their and/or the patients' interests (McGuire, 2000). This can be done through regulatory or financial mechanisms or professional norms. Regulatory mechanisms can incentivise these behaviours by either expanding or limiting the types of services GPs are responsible for, while financial mechanisms use payment schemes designed to align interests. Professional norms can motivate GPs to act in their and/or the patients' interests through medical school, specialisation, or conforming to a medical association's guidelines. Alternatively, payers can improve access to care by directly targeting patients through similar mechanisms, such as lowering co-payments and providing legal rights to access that can be tried judicially. Dealing with information asymmetry is a fundamental prerequisite for these mechanisms to be effective. While GPs can in theory provide complete information to patients, this may not be in their best interests because their objectives can conflict with those of patients (Mooney and Ryan, 1993). Therefore, one solution to the agency problem is to have a third party (i.e. the government) provide patients with information about optimal treatments, increasing patients' bargaining power relative to that of GPs. For example, public information campaigns about breast cancer can increase people's awareness about screening and other preventive measures GPs offer, increasing demand for their services.

To summarise, decisions by GPs to treat patients depend on their motivation to exert effort, how informed the patients are, and the GPs' ability to capitalise on their motivation in the system in question. The extent to which these elements affect GPs' utility and treatment decisions will depend on their individual preferences.

The standard economic framework for studying this type of situation is a utility maximisation problem. Here, a GP chooses an effort level that maximises their utility given their preferences for income and patient welfare. Building on Ellis and McGuire (1990), we specify GP i's utility function as

i) $U(e_i) = w(e_i) + \alpha_i \pi(e_i) - c(e_i)$,

where w is remuneration, and $\pi(e_i)$ is the likelihood of patients visiting GP *i*, both of which are a function of effort e_i. α_i is a measure of GP altruism—the value GP *i* places on treating patients. Finally, c represents the cost of effort. The model illustrates a GP's choice of effort as a function of remuneration, effort costs, and how highly they value patient welfare. A highly altruistic GP will need less monetary

inducement to provide a given level of effort. Subsequently, the GP can determine whether the resulting utility level exceeds what they can expect in other professions and decide whether to stay in the current profession.

Mathematical models, such as the one presented in Eq. (i), help provide high-level predictions and develop hypotheses for empirical testing. However, by design, they rely on strict assumptions and provide an incomplete picture of the issues at hand. We have identified four limitations of the standard model (i) that we believe make it imperfect for adequately reflecting the GP setting.

First, the model does not account for GPs' working conditions and the range of remuneration mechanisms available to the payer. GP systems are often characterised by fee for service (FFS), capitation, salary, or a combination of these. FFS and capitation remunerate the GP based on services provided and GP list length, respectively. The remuneration system in question will provide different incentives for effort levels (Gosden et al., 2000). Moreover, GPs are motivated by factors beyond income and patient welfare, such as career development and workplace flexibility (Lafortune, 2016).

Second, the supply of services is determined by individual GP effort and does not take into account the profession's overall attractiveness. This point relates to how a GP system can affect decisions regarding exit and entry into the profession, not merely how hard a GP works. For example, Holte et al. (2015) have shown that preferences for pecuniary and non-pecuniary incentives differ among doctors, illustrating that a remuneration system is not uniquely a matter of effort for existing GPs but relates to the profession's attractiveness.

Third, the model treats patients as passive consumers of health without the ability to affect the level of care received. Arguably, patients can affect the level of care they receive (Goodyear-Smith and Buetow, 2001), something that has been well established in the patient-empowerment literature (Samoocha et al., 2010). As such, a GP's treatment and effort decisions are not made in a vacuum but are influenced by the patient's ability to act as an agent for their health.

Fourth, the model assumes that a GP remuneration system can accurately reward effort. An underlying premise of this assumption is quantifiable and verifiable. One approach would be to pay an hourly wage, but this does not remove the possibility of shirking. Alternatively, using fees associated with specific services can better reflect effort but only under the assumption that the fees available accurately reflect the range of services offered and that the GP can use them correctly. Finally, as GPs are primarily private practitioners, it may be hard to guarantee that some do not game the system using upcoding practices.

The purpose of this thesis is to discuss these issues from different perspectives. First, we examine different mechanisms to increase the supply of services, regarding providing more *services* and attracting more *GPs*. Second, we discuss whether these mechanisms succeed in attaining the delivery of *desired* services from the payer's and patient's perspectives. Our overall assessment is that attaining these different objectives may be mutually exclusive. Ultimately, a GP system's design will depend on trade-offs made between patient access, efficiency, and the profession's overall attractiveness. If theoretical models fail to provide an accurate depiction of these trade-offs, other approaches can complement the general predictions by providing data on observed behaviour. Therefore, to portray these issues comprehensively, we apply different methodological approaches: theoretical models, observational data, and experimental designs. Although we limit our discussion to measures 'within' a given GP system, we acknowledge the existence of several exogenous push (i.e. medical education and tax incentives) and pull (i.e. delayed retirement incentives) factors that can increase the supply of services and the profession's overall attractiveness.

To study the supply of services, we extend model (i) to encompass working conditions and the range of remuneration mechanisms available to a payer. A revised model that includes outside options will better predict how a given GP system may work differently with respect to incentivising effort versus retaining staff. To illustrate the theoretical model's predictions, we empirically show if different GP characteristics are associated with effort and how income differentials can predict whether a given GP decides to leave the profession.

In studying services provision, we show how financial incentives may be unsuccessful in attaining the delivery of *desired* services and use experimental data to quantify the effects of interventions to overcome these problems. This is studied by looking at how government interventions can steer correct treatment decisions and kerb the incorrect use of public resources. We also study the effects of an intervention aimed at strengthening patient agency to improve access to recommended care.

The thesis is structured as follows. Chapter 2 provides a background on the organization of the GP system in Norway. We also discuss evolving recruitment difficulties and driving factors and contextualise these in a Nordic setting. Further, Chapter 3 introduces the methodology used in this thesis and how and why different methodological approaches can be more successful in answering distinct research questions. Subsequently, Chapters 4 and 5 represent the main body of the thesis, covering the supply of GP services and whether associated incentives are successful in responding to patient needs. Both chapters include discussions of the literature, followed by summaries of the papers. Chapter 6 provides a

discussion of the main results and contextualises the findings from the four papers. Finally, Chapter 7 highlights the main conclusions and present some policy recommendations.

2. Background

2.1 Organisation

The Norwegian healthcare system is primarily publicly financed (85%), with 15% out of pocket payment. Specialist care is managed by independent state owned regional health authorities (OECD, 2021). Provision of care is a mix of publicly owned service providers, private practitioners, and institutions that contract with health authorities and municipalities. The extent of private health insurance has been increasing in the last decade, but its overall share remains considerably small (Saunes et al., 2020).

Municipalities are responsible for providing primary healthcare services in Norway, including GP services. All citizens, permanent residents, and asylum seekers are entitled to enlist with a GP (Lovdata, 2021a), and most do. GPs can choose their list ceiling[1] but cannot select their patients (Lovdata, 2021b). Patients who are dissatisfied with their GP can freely change GPs up to twice a year (EY and Vista Analyse, 2019). The purpose of the list system is to ensure continuity of care by establishing relationships between GPs and patients. GPs act as gatekeepers for their patients, issuing referrals, sick leave certificates, and prescriptions for medications. As such, they act as the first point of contact for citizens' healthcare services. People can also visit an emergency room for acute care or private healthcare providers, but most contacts are handled by GPs.

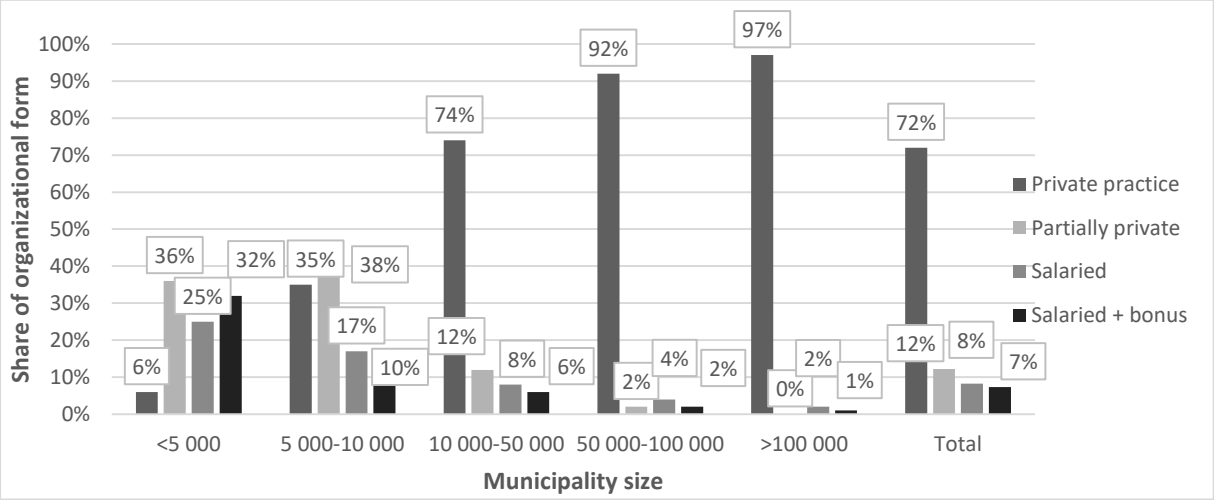
GPs are mostly private providers contracting with municipalities—around 85% according to official statistics (Gaardsrud, 2021). They are remunerated through a combination of capitation, FFS, and patient co-payments, which constitute approximately 30%, 60%, and 10% of income, respectively. Total gross earnings for GPs in this contractual form averaged NOK 1.9 million in 2020 (Statistics Norway, 2022). Fees covered under the FFS payment are set annually through negotiations between the Norwegian Medical Association, the Ministry of Health and Care Services, Regional Health Authorities, and the Norwegian Association of Local and Regional Authorities (Nilsen, 2022). Fees are administered by the Norwegian Health Economics Administration (HELFO), which also trains and audits GPs regarding the correct use of fees (Pedersen et al., 2022). Patient co-payments are capped at NOK 2921 (2022), after which the marginal treatment is free of charge (HELFO, 2022). Private practice GPs have overall responsibility for organising and financing their practices, either by themselves or jointly with other GPs

[1] Depending on contractual arrangements, the municipality may have direct or indirect co-decision over list length.

operating in group practices. This includes hiring staff and recruiting locums when practice owners are off work or on sick leave.

An increasing number of municipalities offer alternative remuneration schemes in response to recruitment difficulties or uncertain patient volume in sparsely populated areas. These schemes include i) salary, ii) salary + bonus, and iii) partially private practice. The latter refers to schemes where the municipalities keep (a share of) the capitation payment in return for covering fixed costs associated with a practice. These schemes are often offered when a small patient base does not allow running a financially viable private practice (Helseøkonomisk Analyse, 2021). However, more recently, larger municipalities are increasingly offering these contracts to attract GPs who value stability and flexibility (Deloitte, 2022). These alternative schemes are often costly for municipalities because the income GPs generate for the municipalities through capitation, FFS, and patient co-payments does not cover their costs fully (Pedersen et al., 2022). Table 1 shows the responses from a survey of Norwegian GPs about their organisational form. The figure shows a close correlation between municipality size and organisational form.

Figure 1: Organisational form of Norwegian GPs



Source: (Pedersen et al., 2022)

In addition to their daily work at a GP practice, most GPs are required to take shifts at emergency rooms. Nevertheless, many are exempt from this requirement due to age or because they work in larger municipalities with professionalised emergency rooms that have full-time doctors. In 2020, GPs undertook 56.7% of all consultations at emergency rooms (Sanvik et al., 2022). GPs in emergency rooms

(68.3 hours) work considerably more than those who are exempt (49.1 hours) (Keilegavlen et al., 2018). Table 1 presents selected statistics for GPs in Norway.

Table 1: Selected GP statistics

	2012 (2015)	2020
Average GP age	48.3	47.2
Share of GPs aged >55	35%	29%
Share of female GPs	38%	46%
List length	1 164	1 068
Share of open lists	40%	26%
Share of GPs salaried		15%
Share of specialists		63%
Consultations per patient	(2.66)	2.88

Source: (Gaardsrud, 2020)

The data show some notable trends: the age composition of GPs remains fairly stable, women constitute a larger share of the workforce, list lengths are decreasing, and consultations are increasing. This reflects a change in the composition of the GP workforce, as well as GPs adapting to heavier workloads through reduced list lengths. Together with an ageing patient population, these trends are likely due to GPs' changing responsibilities over time. In studying GPs' workloads, Morken et al. (2019) found that average working hours were 55.6 hours per week (49 hours, excluding out-of-hours primary healthcare). The high workload does not translate into more patient contact, which the authors estimated as around half of a GP's time. They attributed these developments to GPs' increased responsibilities. This aligns well with several reforms introduced in the last decade. A GP's role in preventive care has increased following the coordination reform, higher demands following reforms to ensure continuity of care for patients with drug and psychiatric issues, and increased administration following requirements to issue absentee sick leaves for high-school students (Trønderopprøret, 2018). While exemptions exist, GPs have also been required since 2017 to specialise in primary care (Norwegian Directorate of Health, 2019).

2.2 Recruitment and retention difficulties

The developments in the GP system have led to a shortage of GPs. Table 1 shows that the share of open lists is decreasing. This issue is not found only in smaller municipalities, as 65% of lists were closed in municipalities with 50 000+ inhabitants (The Norwegian Directorate of Health, 2022a). In two mappings regarding the recruitment situation, the number of municipalities stating that attracting new GPs had become more difficult increased from 11 in 2017 to 47 in 2019 (Norwegian Medical Association, 2019a). Kjosavik (2018) also reported insufficient recruiting of doctors to primary care.

Recruitment difficulties are not unique to Norway; most Nordic countries report similar challenges. These countries vary in their organisational form (Olsen et al., 2016), suggesting that the difficulties are not system-specific.

A survey conducted by the Finnish Medical Association shows that the average GP deficit (in health centres) was 7.5% in 2019, a deterioration from previous years (Finnish Medical Association, 2019). Other studies point to similar trends (Kuusio et al., 2010). Temp agencies are prevalent in Finland, where recruitment difficulties have also led municipalities to outsource primary care provision at the expense of continuity of care (Kokko, 2007).

Interviews with hiring managers in municipalities suggest that the difficulties are just as acute in Sweden (SOU, 2019). The relative high use of locum GPs in Sweden has been argued to be the result of recruitment difficulties. This is consistent with the higher use of GP locums in rural (46%) than urban (18%) areas (Vårdanalys, 2018).

In Denmark, several regions voiced recruitment difficulties in a 2017 report evaluating the GP scheme (Danish Ministry of Health, 2017). In the same report, the number of vacancies for specialist placement positions was highlighted as a proxy for recruitment difficulties. Although 2017 and 2019 data (Danish Health Authority, 2020, 2018) do not indicate a trend of increasing recruitment difficulties, they suggest that it is generally more difficult to recruit GPs than other specialist placements.

2.3 Factors affecting recruitment and retention

Similar to other authors (Bennett and Phillips, 2010; Young and Leese, 1999), we can dichotomise the recruitment issue into the demand and supply of GP services. Demand is a function of demographics, patient health, and the types of issues GPs solve; supply involves maintaining a stock of GPs that supply services. This stock and their service provision, in turn, depend on the profession's attractiveness and the resulting inflows and outflows.

For a given proportion of medical graduates choosing a GP career, increasing the pool (medical school intake) will increase the number of GPs. The trend in medical school intake has been positive for all Nordic countries and enhanced by an increase in the number of foreign-trained physicians. Workforce migration seems to play only a moderate role in all countries, likely due to language barriers (OECD, 2019). Excluding Finland, where the proportion has increased, all other Nordic countries have a lower share of GPs; on average, it is down from 19.6% in 2005 to 17.1% in 2018 (OECD, 2020a). This suggests that a given physician inflow has not translated into an equal flow of GPs.

An ageing GP population in several Nordic countries underscores the urgency of recruitment. In Sweden, close to 40% of specialists were 60 years or older in 2016, compared to only ~10% in 2002 (The National Board of Health and Welfare, 2020). In Denmark, the share of GPs approaching retirement age (60–69) has increased from ~10% to ~32% from 2000 to 2004 (Danish Ministry of Health, 2017), raising questions about the scheme's sustainability (Jessen, 2018). Furthermore, all countries have seen substantial changes in the gender balance, with an annual average increase in female GPs of ~1 percentage point in the last 15 years (OECD, 2020b).

We have seen that a GP shortage does not necessarily mean that we have fewer of them; it can also mean that we are asking them to do more than before. For example, more administration, coordination, and specialist procedures have been transferred to GPs in Sweden (Vårdanalys, 2017, 2018) and Norway (Gaski and Abelsen, 2015; Trønderopprøret, 2020) without alleviating the patient burden. Demographic changes have also led to a change in the burden of disease towards more lifestyle-oriented risk factors, increasing the demand for primary healthcare (Knudsen et al., 2019).

Together, these developments are cause for concern, especially since they translate into intentions to quit. In a Finnish study, more than half the respondents said they intended to quit their job (Kuusio et al., 2013). Similar, albeit smaller, results have been found in both Sweden (Eneroth et al., 2017) and Norway, where smaller municipalities especially experience significant turnover (Abelsen et al., 2015).

3. Methodological approach

The economics field has witnessed growth in the plurality of methods applied to answer research questions. Experimental economics, behavioural economics, and big data have added to the wealth of methodologies researchers apply to answer questions (Dow, 2007). In the broader economics literature, theory-based research dominated from the 1960s to 1990s, while employing empirical data from secondary sources has remained stable over time. Using data from surveys and registries has increased substantially since the turn of the millennium, largely due to its increased availability. Finally, experiments have become more prominent in the last decade or two (Hamermesh, 1960). These developments have also transcended sub-specialties of the field, including health economics (Fuchs, 2000).

Each of the methodologies used in health economics has its advantages, but they differ in their ability to answer specific research questions. Observational data can be used to make inferences about the effect of larger policies across numerous issues but seldom provide robust causal estimates regarding behavioural responses that experimental data can offer. Whereas theoretical frameworks often rest on strict assumptions that may not always be valid when applied in real-world settings, their strength lies in providing a framework for empirical research (Sloan and Hsieh, 2012, p. 202). As such, we argue that by applying different methodologies, we make a better case and provide a more complete picture of GP behaviour. In the next sub-sections, we discuss the strengths and limitations of these approaches and why their application is relevant to our setting.

3.1 Theory-based models

As previously discussed, the standard approach to modelling GP utility is making it a function of income, patient benefit, and leisure (Ellis and McGuire, 1990). The implication is that GPs face different trade-offs depending on their value judgements related to patient benefits or, put differently, how altruistic they are. Jack (2005) extended this model by relaxing the assumption of knowing a doctor's level of altruism. He showed that by paying for costs incurred, GPs reveal their preferences for payment schemes. Dionne and Contandriopoulos (1985) extended the basic physician utility model by including other issues such as prestige and ethical concerns. They also referred to the effect on long-term trade-offs, such as choosing to specialise in a certain field, but without identifying any equilibrium. Finally, Ellis (1998) depicted a model that included different disease states and severity among patients, showing that payment systems with full- or zero-cost coverage are sub-optimal.

Arguments have been made that standard neoclassical assumptions are too strict in the field of health economics, which is characterised by incomplete information, making utility maximisation challenging (Arrow, 1978). Moreover, its static approach is incomplete for providing a full picture of the issues at hand, given the dynamic nature of healthcare provision and transaction costs (Hodgson, 2008). Despite their limitations, theoretical model predictions have often been tried and found useful. Therefore, their strength lies in providing a framework for developing hypotheses for research questions. Importantly, this may be just as useful for new policies. While one can often draw upon previous empirical research, predictions are not always generalisable to other settings (Deaton, 2020a) and/or reflect policy changes that have yet to be tried elsewhere.

We argue that theoretical models provide a good starting point for describing GP behaviour. Importantly, the models discussed are primarily concerned with GP utility, treating participation in the profession as a given. In other words, optimisation is contingent on working as a GP without taking into account outside options. There is a wealth of empirical evidence on the characteristics associated with GP intentions to quit and GP effort responses to and preferences for different payment mechanisms. However, these questions have been tackled discretely or discussed under assumptions that are not valid considering the current recruitment difficulties. Incentivising GPs to do more might make more sense if there is a steady stream of new joiners. Given the competition between employers to attract new doctors, the question of supplying healthcare services is not only about inducing higher effort, but also ensuring there is someone in the GP office to induce. The purpose of our theoretical model in Paper 1 is to join these concepts.

3.2 Experiments

Randomised control trials (RCT), where individuals from a sample are randomly assigned to treatment and control groups, are often regarded as the gold standard for identifying the causal impacts of interventions (Glied et al., 2011). By keeping the treatment exogenous to other explanatory variables, RCTs are considered superior to observational studies for robustly quantifying the economically relevant magnitudes of policy effects (Newhouse, 1987). Lab experiments, in which subjects are given an abstract framing and an imposed set of rules (Harrison and List, 2004), have emerged as a new sub-discipline of RCTs. One of their benefits is exposing the underlying physiological mechanisms that drive results (Lunn and Choisdealbha, 2018), making it the preferred approach in the field of behavioural economics (Weber and Camerer, 2006).

Despite several other sub-classifications of experiments, in this context, we focus on *natural* field experiments. Harrison and List (2004) define the associated study environment as *'…one where the subjects naturally undertake these tasks and where the subjects do not know that they are in an experiment'.* From this definition, it follows that well-designed field experiments can control for bias arising from outside factors—after all, subjects are in their natural environment. In addition to being expensive, one of the more cited critiques follows from the definition: context comes at the expense of generalizability (Al-Ubaydli and List, 2015; List, 2011). Al-Ubaydli et al. (2021) argue that the difficulties associated with scaling can be overcome through several steps, including making studies policy relevant in the experiment's design phase. Finally, whereas RCT can give robust point estimates for means, they are less informative about the distribution of effects (Deaton, 2020b). Despite these drawbacks, the natural setting of field experiments provides researchers with unique opportunities to study economic theory in real-world settings (List, 2011).

One of the more influential experimental studies on remuneration schemes and insurance contracts in health economics is the RAND experiment, where a large number of people were randomised into different insurance schemes (Newhouse and Rand Corporation, 1993). Few experiments have come close to matching its scale or scope. While policymakers have the opportunity to pilot interventions (i.e. when changing remuneration schemes), willingness has often been lacking (Scott and Hall, 1995). One reason might be pushback when intervention groups receive perceivably more favourable or unfavourable conditions than control groups. For drugs and vaccines, this can also be related to the ethics of providing one group with inferior treatment (i.e. a placebo). In the case of Norway's remuneration system for GPs, it was introduced in five pilot municipalities before it was scaled and fully implemented in 2001 (Sosial- og helsedepartementet, 1997). While this gave compelling evidence for the policy change, the associated effects were static. Owing to incremental changes to capitation, fees, and overall GP responsibilities, findings from these reports would likely not fully hold after 25 years.

As for the two field experiments in this thesis (Papers 3 and 4), the critiques related to results generalizability should be motivation to conduct field experiments on GP and patient behaviours; causal links found in other study settings may be so context dependent that the results do not apply to the Norwegian setting. Moreover, we argue that our study settings are particularly policy relevant, as both studies could easily be scaled to other issues related to prevention and GP fee usage. Finally, we believe that our research questions are novel.

3.3 Observational studies

Where RCT's strength lies with regards to identifying causal relationships, they often fall short in feasibility; they can be costly, unethical, contextual, and narrowly oriented in outcomes. Therefore, observational studies have seen extensive application in the health economics field. They often rely on detailed registry data of large populations. Systematic reviews also show that they generally give predictions similar to those of RCTs (Anglemyer et al., 2014; Black, 1996; Ohn Oncato et al., 2000). Observational studies' usefulness and validity often depend on the quality of the data applied. Norway has one of the more detailed registries in the world, where personal identifier numbers can be used to link data sources, provide comprehensive controls, and analyse behaviour over time. Researchers can identify characteristics of individuals with different practice styles, but it is difficult to assess why they opted for a particular one. This relates to the issue of adverse selection, where individuals self-select into preferred payment schemes (Sloan and Hsieh, 2012, p 98). For example, observing that salaried GPs exert lower effort may be an artefact of slow-working individuals preferring that practice style rather than salary compensation leading them to be inefficient.

Methods to overcome self-selection include natural experiments and instrumental variable (IV) regression. Natural experiments require an exogenous shock to observe time trends before and after the change occurred; IV regressions do not have this requirement. In IV regression, a variable (instrument) is identified that affects the outcome variable but only through its effect on an endogenous variable. The approach mimics concepts of RCTs where a source of randomisation is identified in the instruments. As tests associated with IV regressions can never fully say whether an instrument is appropriate, the challenge lies in assessing its validity critically and logically (Baiocchi et al., 2014; Glied et al., 2011). Nevertheless, the approach has seen wide application in the health economics field, using genetic markers (Von Hinke et al., 2016), natural shocks (Harrison and List, 2004), and patient health status (Jacob et al., 2012) as instruments.

To our knowledge, few observational studies have explored quitting behaviour among GPs. While numerous studies have examined the effects of income on intentions to quit, these are survey-based and rely on cross-sectional data (i.e. Marchand and Peckham, 2017 and Owen et al., 2019). This also means that one cannot determine whether a GP saying they want to quit translates into actual quitting. In this context, survey data are based on *stated preferences*, while observational data are based on *revealed preferences*. It has been argued that survey data depict behaviour more accurately, although possibly more difficult to infer from statistical analysis (Mark and Swait, 2004). Using detailed Norwegian registry

panel data, we compile a unique dataset to compare practice characteristics of GPs leaving versus characteristics of those staying in practice. We estimate the effect of income on the decision to quit using an IV approach.

3.4 Data

In this thesis, we have used a theoretical model, observational data, and experiments. In the quantitative studies (Papers 2–4), we used Python 3.X for data handling and analysis and Stata 16/17 for econometrics.

In Norway, there are multiple health registries with substantial detail related to process of care and patient and doctor attributes. We rely primarily on two data registries particularly relevant for the GP setting: the registry for control and reimbursement of healthcare fees (KUHR) and the Norwegian GP registry (FLO). KUHR contains claims data for several healthcare professionals, including GPs. The claims are organised by date-stamped individual consultations and contain information about diagnosis, patient number, and fees used during the consultation (Norwegian Directorate of Health, 2023a). FLO contains annually updated information about GP and practice characteristics, including age, specialty, municipality code, list length, and patients enlisted with a given GP (Norwegian Directorate of Health, 2023b).

In Paper 2, we used registry data from KUHR and FLO. Paper 3 used monthly GP claims data from HELFO, sourced from KUHR. Paper 4 used data from a survey designed by the authors and collected by the Norwegian Diabetes Association.

3.5 Ethics

Research in health economics is subject to various ethical issues depending on the methodology employed. Although our research does not involve direct human subjects, it deals with human-related aspects and thus has the potential to positively or negatively affect their well-being. Therefore, a fundamental ethical principle is to ensure that we do not worsen outcomes for the individuals under study or provide inferior outcomes for those not included in the study. For all data-driven studies in this thesis, we obtained permission from the Norwegian Centre for Research Data (Norwegian: Norsk senter for forskningsdata, NSD).

The range of methodologies employed in this thesis necessitates different considerations. In the case of Paper 1, the methodology does not raise ethical concerns regarding research subjects because there are

none. The results and associated interpretations are inherently general, focusing on individual trade-offs rather than distributional aspects such as inter-GP resource allocation and patient selection. Paper 2 involves real subjects but relies on anonymized historical data. It should be noted that anonymized data does not guarantee that research subjects cannot be identified, particularly in our case where information about place, sex, and age is available. Hence, safeguarding the integrity of research subjects is paramount, as with other registry-based studies (Ludvigsson et al., 2015).

Papers 3 and 4 involve experiments aimed at altering individuals' behaviour. This methodology is prone to various pitfalls due to the potential positive and negative consequences for individuals in both the control and intervention groups. In 1978, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1978) published the Belmont Report, which offers ethical guidance for research involving human subjects. The key principles put forward in the report are respect for persons, beneficence, and justice. Respect for persons involves informed consent and the protection of autonomy, while beneficence entails the principle of 'do no harm,' and justice emphasizes non-exploitation in the selection of research subjects and ensuring that the benefits of research accrue to those participating in it (Ifcher and Zarghamee, 2016).

In our experiments, we did not seek informed consent from the research participants. It is important to note that this decision does not imply that it was unethical (Wilson, 2014). The primary reason for not seeking informed consent was to avoid contaminating the results by allowing research subjects to be aware of their participation in an experiment, which aligns with the previously mentioned definition by Harrison and List (2004). Consequently, we designed the studies in a manner that the data did not contain personal information enabling the identification of research subjects. However, we acknowledge that conducting research through governmental agencies and public awareness of the issue can lead to reduced trust in institutions (Hegtvedt, 2014).

Regarding the principle of 'do no harm,' we cannot rule out the possibility that some patients received inferior treatment as a result of the information campaign targeting general practitioners (GPs) in Paper 3. Although the content of the intervention did not imply that GPs should alter the treatments offered to patients, some may have done so out of fear of potential consequences even with well-documented use of a double consultation fee. Nonetheless, the intervention mirrored those previously initiated by HELFO and thus did not cause any additional harm triggered by the research. In Paper 4, the control group later received the same information as the intervention group, ensuring they were not 'harmed' by having inferior access to potentially health-improving information.

Finally, we assert that the research subjects were not exploited, and they were recipients of the benefits. In Paper 4, research subjects received compensation for participating in a survey, and this compensation was equally offered to both the intervention and control groups. As mentioned earlier, all participants also derived benefits from the study. In Paper 3, our sample consisted of individuals who were believed to have incorrectly utilized the studied fee. However, this sample was also relevant in terms of HELFO tailoring interventions independently of our study. Moreover, we contend that the GPs benefited from the study by ensuring trust in their long-term service provision. Several newspaper articles (Brembo and Andersen, 2022; Mo and Kongsrud, 2021) have previously revealed illicit use of remuneration schemes among doctors, and audits and feedback are considered effective tools for curbing such behaviour.

4. Mechanisms to increase the supply of services

Increasing access to or the supply of GP services can be interpreted as getting GPs to do more (the intensive margin) or attracting more doctors to the GP profession (the extensive margin). While some policies are fairly specific in targeting one or the other, policies that induce a certain type of (efficient) behaviour often equally affect the perceived attractiveness of the profession for some. GPs have heterogeneous preferences for practice characteristics, making it difficult for policies to monotonically improve utility for all.

Which policies constitute the most effective ones is also a function of the issue one is trying to solve—attracting more GPs or getting existing GPs to do more. For an infinite pool of GPs, it may be more important to ensure cost-effective delivery of care because the workplace is already deemed attractive. For many Nordic countries, the problem is largely about being able to meet patient demand. However strong incentives you provide, if no one wants to work as a GP, there is no effort to induce.

Notably, the problems that need solving are an artefact of the broader healthcare system's organisation. For instance, if a GP has high referral rates to outpatient services, patient care may be more costly and inefficient than treatment at a GP practice. This is one of the reasons the Norwegian government has had a long-stated objective for patients to be treated at the lowest service level of the health system: primary care (Norwegian government proposition, 1995). In studying theoretical predictions of different payment mechanisms, Allard et al. (2011) found capitation to be the most expensive in terms of more frequent patient referrals to specialist care. For services costly to the GP, FFS was expected to produce similar results. Thus, what constitutes the optimum will depend on trade-offs related to costs and quality of care.

4.1 Getting GPs to do more

Payment mechanisms in health care are often discussed in the context of agency theory. Here, the entity (principal) defines a contract with a service provider (agent) that aligns the agent's interests with those of the principal (Robinson, 2001). If we understand the payer's objectives as cost-control and healthy patients (Roberts et al., 2008), an underlying assumption is a desire to incentivise high effort levels – that is doing more, faster. The GP system's attributes complicate the matter, as reflected in the difficulty verifying an agent's effort and whether this, in turn, achieves good patient outcomes. The range of payment mechanisms has been found to respond to different objectives regarding efficiency and treatment selection.

A simple payment mechanism is a fixed salary, which provides predictability regarding cost control but uncertainty about a GP's effort. As the salary is fixed, the principal relies on the GP's intrinsic motivation to work hard and not shirk (Liu and Mills, 2007). As capitation rewards GPs who enlist many patients, GPs are rewarded for keeping patients happy and from switching (Gravelle, 1999). In FFS, GPs are paid retrospectively for the marginal treatment; thus, they are incentivised to provide a high level of effort (McGuire, 2000). This payment scheme is also argued to be more fair concerning 'equal pay for equal work' (Dugdale, 1997). This follows from GPs being paid for doing more, assuming that the size of fees capture the effort involved and that they are coded correctly.[2] Pay for performance (P4P) also pays GPs retrospectively but is based on outcomes instead of inputs (as in FFS). In theory, this means that the incentives of the GP and patient are well aligned (Green, 2014). FFS is equally expected to strengthen agency towards patients, as higher effort generally translates into improved care. Consequently, salary is expected to reduce a GP's agency towards a patient, as effort (and outcomes) only rely on the GP's intrinsic motivation. The negative effect on agency from the prospective nature of capitation is expected to be offset by the incentive of retaining and attracting new GPs.

There is a large literature of empirical studies testing the predictions of the different payment schemes. In studying Norwegian GPs, Brekke et al. (2020) found that FFS, compared to salary, leads to an increase in service provision for extensive and intensive margins; more consultations are undertaken, and more treatment is offered during consultations, including tests, prolonged visits, and referrals. Similar results were found in Canada, where physicians under FFS contracts provide considerably more patient visits than under other remuneration types (Devlin and Sarma, 2008). They also found evidence of self-selection into different contract types, suggesting that individual characteristics, not just incentives, are also at play. This is supported by Pedersen et al. (2012), who found that GPs working in solo practices have different preferences for organisational attributes and remuneration than those working in group practices. The effect of capitation is somewhat undetermined, as a GP's response is affected by the market conditions in the catchment area. More specifically, Iversen and Lurås (2000) found that GPs who experience a shortage of patients respond to the lost income from capitation by increasing the use of FFS. In a systematic review, Gosden et al. (1999) established that while the effect of salary is context-dependant, it generally leads to fewer services provided and better cost-control. Sørensen and Grytten (2003) find comparable results, estimating that GPs who move from a salary-based to an FFS contract increase services by 20–40%.

[2] We discuss this in more detail in section 5.

Some support has also been found for the existence of target income when moving from capitation to FFS—the effect of FFS in inducing effort is reduced after a certain income level is reached (Krasnik et al., 1990). One interpretation of target income is that it reflects the standard assumptions of the utility of income: positive but with diminishing marginal utility.[3] At high income levels, the marginal consumption provides limited extra utility. Therefore, high-earning GPs may decide to trade consumption for leisure, which is assumed to be a normal good.

In addition to income and leisure, altruism can be an important source of motivation for GPs (Chalkley and Malcomson, 1998) and therefore central to the design of remuneration systems (Delfgaauw, 2007). In a laboratory experiment, Godager and Wiesen (2013) showed that altruism is an important component of GP treatment decisions. The results also revealed heterogeneity in preferences, although most deem patient benefits to be as important, or more important than private earnings. This suggests that the effectiveness of different remuneration schemes is sensitive to the altruism of individual GPs. This becomes even more pertinent because financial rewards and altruism can be substitutes rather than complements. Several studies have established that introducing extrinsic rewards can crowd out the intrinsic motivation for a given task (Bénabou and Tirole, 2003; Deci, 1971). In healthcare, studies have shown this to be the case for retrospective reimbursement of services, such as FFS and P4P (Green, 2014).

4.2 Keeping GPs happy

The literature includes studies of GP effort levels in static environments, namely how practice styles and utility vary among existing GPs depending on the remuneration scheme in question. Put differently, participation in the GP profession is disregarded and treated as a given. However, with a shortage of GPs, policymakers increasingly find themselves competing against hospitals and other institutions for doctors. Therefore, a growing literature has been devoted to studying preferences for different practice styles, not merely the resulting effort levels.

There is a large survey-based literature on the source of GP satisfaction and motivation. While GPs often work as private practitioners with the discretion to choose recommended care, organisational constraints and strong incentives may be a source of discontent for some. Studies have shown that the lack of workplace autonomy and decision-making in patient care (Janus et al., 2007) and diversity of work and professional development (Van Ham et al., 2009) can reduce GP motivation. Flexible working

[3] In equation (i), this equates to $w' > 0$ and $w'' < 0$.

hours and a clear delineation of GP responsibilities have been found to be desirable practice characteristics (Simoens et al., 2002). Scott et al. (2006) found that these characteristics are not just associated with overall workplace satisfaction, but also translate into intention to quit.

In studying preferences over time, Holte et al. (2015a) found that an increasing number of GPs prefer salaried positions over private practice. Those preferring private practice name 'autonomy' and 'income' as the most important determinants, while those preferring salaried positions name 'better social security/benefits' and 'less administrative work'. Interestingly, however, the results are fairly consistent with regards to a majority preferring existing practice styles, pointing to a degree of path dependency in practice preferences. In surveying final year medical students in Norway, Abelsen and Olsen (2015) found that most respondents preferred contracts that are less activity-based compared to the current mixed capitation and FFS system. In the UK, exposure to P4P financing did not affect workplace satisfaction among existing GPs as long as income did not change (Allen et al., 2017). This suggests that preferences for different practice styles may be weighted differently for recruiting versus retaining GPs. There is also evidence of heterogeneous preferences as a function of GP age. In studying retention among GPs in England, Dale et al. (2015) found that improved working conditions, defined as a reduced workload and less administration, more patient time, greater clinical autonomy, and more flexible working hours, became relatively less important with age.

While the attributes of the system matter, not surprisingly, income also matters. In studying future doctors' specialty preferences, Sivey et al. (2012) found that wage expectations far exceeded potential general practice earnings. However, the authors also found that there were trade-offs with other attributes, suggesting that workplace opportunities, such as academic work or increased procedural work, could be just as desirable as $50 k in increased earnings. Accordingly, keeping GPs happy is not merely a budgeting exercise but a process of designing a system that enables professional development and autonomy in a flexible working environment. In studying practice style preferences, Holte et al. (2015a) further exposed that preferred practice styles were a trade-off between income and system attributes. For example, they found that most respondents in hired practices (where municipalities cover fixed costs and organise running the practice) would change to private practice if there was no change in the ability to control working hours, professional development, professional autonomy, and income.

4.3 Research questions

In this section of the thesis, we present two studies related to how financial and non-financial incentives relate to effort and the quitting decision. More specifically, we ask the following questions:

1. *How do pecuniary and non-pecuniary incentives affect the provision of care and attractiveness of the profession among GPs?*
2. *What are the determinants of earning differentials among GPs below retirement age, and do these translate into quitting behaviour?*

In the first paper, titled 'GP retention and recruitment in the Nordic countries', we expand the classical theoretical frameworks of Ellis and McGuire (1986) by including the outside option of quitting. In the second paper, titled "Just pay them more". Income generation and its effect on the retention of general practitioners in Norway, we use registry data to identify the characteristics of earning differentials and whether these, in turn, help explain quitting decisions.

As is clear from studies on intentions to quit and preferences for practice styles, GPs have different preferences for remuneration schemes and contracts beyond how they translate into effort. Put differently, while a GP may work harder and earn more under an FFS contract, they may still prefer a salaried contract. As such, policymakers must not only understand how incentives translate into higher service provision, but also whether the same incentives make the profession more attractive. This is especially the case in the presence of GP scarcity, where employers compete for labour. The first paper tackles this question by integrating the decision to provide effort into a framework with outside options. To this end, we draw on Brekke and Nyborg (2010), who study nurses' efforts and, subsequently, the employment decision. By incorporating unique features of the GP system, including the mixed payment scheme they face and the need to attract patients, we provide theoretical predictions of the aggregate—and not just individual—supply of GP services.

In the second paper, we join the concepts raised in Paper 1 to test some of its predictions empirically. More specifically, we employ a rich registry-based dataset to investigate which practice characteristics are associated with higher effort and, in turn, income. Subsequently, we apply an IV approach to quantify the effect of income on quitting behaviour. This provides new evidence of GPs' actual responses to earnings, not merely stated intentions. It also provides an indication of how an increase in income can prevent quitting. As such, policymakers may be better equipped to determine both the cost and effectiveness of different measures to reduce quitting.

Paper 1

Introduction

In this paper, we give a formal depiction of how different policy interventions for GP systems affect both effort and quitting. Our study's main motivation is the increasing attention paid to meeting the heightened demand for GP services while simultaneously making the profession attractive to existing and future GPs. In the literature, these two concepts have largely been studied discretely—how to incentivise higher effort versus making the profession attractive. The standard models are primarily concerned with effort and utility (i.e. Ellis and McGuire (1990), etc.), but few consider the choice to enter or leave a profession. González and Macho-Stadler (2013) looked at dual practices but not whether to quit or enter the profession. Moreover, joint hypotheses of the two are lacking.

Methods

Our point of departure is Brekke and Nyborg (2010), who investigate the decision to become a baker or a nurse and whether individuals will shirk in their jobs as nurses. In their theoretical model, the decision depends on an individual's level of altruism, salary, and the government's investment in technology. We adapt the model to account for different remuneration schemes and the characteristics unique to a GP setting. More specifically, we include the option to be remunerated through salary, capitation, or FFS. We also include an altruistic component but assume that GPs are only altruistic towards the patients they treat and not the list population as a whole. As such, GPs will have an incentive to attract patients by providing more effort. Finally, we include both practice quality and working conditions as two distinct effects. We argue that the former is observable to the patient and financed by the public, while the latter is not. Our modelling choices allow us to identify i) effort responses to different policies in the GP system and ii) whether utility increases relative to that in the non-GP sector, and thus the effect on retention and recruitment of new GPs.

Results

The effort decision

We find that altruistic GPs exert more effort because of the non-monetary reward they receive from treating patients. Better working conditions increase effort as the associated costs are reduced. The effect of improved practice quality is indeterminate; while it leads to more patients desiring to see their GP, the GP may respond by reducing or increasing effort depending on the substitution effect between consumption and effort. Higher salary or capitation leads to a lower effort because the valuation of the extra income earned through the FFS component is reduced. The effect of an increase in FFS is

indeterminate as it depends on whether the income effect of higher income dominates the substitution effect of leisure becoming more costly. The effect also depends on a GP's relative effort compared to that of the average GP.

The employment decision

In our model, increased remuneration has a positive effect on the decision to work as a GP. The exception is FFS, where the effect is indeterminate. As the GP system is tax-financed, GPs with low efforts relative to their peers will see a net reduction in disposable income from the increase in FFS. This ensures that new entrants will have high effort levels. Similar to the effort decision, the effect of improved practice quality is indeterminate. It depends on changes in a GP's disposable income, their valuation of it, and any altruistic gains from attracting more patients. Non-altruistic GPs will be worse off because their net income is reduced through higher taxes, resulting in exits from the profession. Finally, we find that improved working conditions make becoming a GP more attractive.

Conclusion

Our findings suggest that only the most motivated individuals choose to become GPs; they are encouraged by the altruistic benefit of treating patients. This also induces them to provide positive effort, which is reinforced if FFS is (part of) the remuneration scheme. An important reason for using FFS is that it induces effort for those who are not otherwise sufficiently motivated. While we see this prediction in our model for GPs with below-average efforts—which could be a policymaker's target group—the same group is altogether deterred from becoming GPs. The intuition behind this finding is that if you find yourself in a job you are not motivated to do, you will either quit or need strong inducements to work hard.

Unsurprisingly, we see that improved working conditions increase utility and thereby recruitment. This means that reducing administration (as seen in Norway) will most likely have a positive effects on attracting new GPs. This would also generally be the case for policies that increase remuneration. For negligible tax increases on investment in practice quality, we would see a positive effect on recruitment because it either helps new entrants attract patients or allows them to reduce their workloads.

Paper 2

Introduction

In most OECD countries, recruiting and retaining GPs are proving difficult. A shortage of GPs could affect the quality of care and prevent health systems from meeting the growing demand for primary care services. Remuneration policies are likely to contribute to these difficulties. However, little is known about observed GP quitting behaviours and decisions and how they relate to earning differentials among GPs. We consider the Norwegian setting, where most GPs are remunerated through a combination of capitation and FFS. Such activity-based financing can translate into earning differentials among GPs, either from working longer hours or from working more efficiently. In this study, we are concerned with the latter—income per consultation. We specifically consider two areas of interest: i) explanations for earnings differentials in income per consultation and ii) how practice characteristics and income predict quitting behaviour.

Methods

We leverage unique Norwegian registry data from 2010 to 2019, including ~100 million observations that cover consultation-level information of 5 million patients and 10 000 GPs. We identify quitting GPs as those that exit our dataset at some point during our study period. First, we explore sources for earnings differentials using linear regression models. Second, we explore the characteristics associated with quitters in our dataset, specifically the effect of income per consultation on the decision to quit. To overcome the endogeneity in the relationship between the quitting decision and income per consultation, we use an IV approach based on exogenous patients' case-mix. We argue that the GPs' inability to select patients makes the instrument appropriate, and we investigate the validity of our assumption through several tests. In robustness tests, we further stratify our sample by GP tenure and specialist status to account for systematic differences in practice styles acquired over time.

Results

We find evidence that heavier workloads, as measured by list length, overtime, and number of consultations, increase average income per consultation. As such, we find that the drivers behind higher aggregate income in a mixed FFS and capitation system also increase income for average income per consultation. In predicting the quitting decision, we find that quitting GPs tend to generate less income per consultation. Finally, longer consultations and being salaried are associated with a lower probability of quitting. Our results are generally robust across sub-samples of specialists versus non-specialists and tenured versus new GPs.

Conclusion

Our findings suggest that GPs who successfully align their practice style with the attributes of the remuneration system, rewarding GP output in terms of patient list and consultations, are more likely to stay in the profession. This may simply reflect heterogeneity in the preferences and financial optimisation abilities of GPs in Norway, suggesting that broadening the range of contractual arrangements may improve the profession's attractiveness and retention.

5. Delivering the desired services?

In the previous papers, we discussed incentives to increase the GP supply regarding both effort and the decision to work as a GP. This perspective focuses on incentivising effort and aligning GPs' objectives with those of the patient but not on whether the resulting use of resources is efficient. Notably, whether a given remuneration scheme enables an 'efficient' use of resources depends on its ability to produce a given output with the fewest number of inputs (technical efficiency) and whether scarce resources are allocated to maximise social welfare (allocative efficiency) (Sloan and Hsieh, 2012, p 19). For example, providing services with marginal patient benefits constitutes a technically inefficient use of resources, and prioritising patients who have little to gain that leads to foregoing treating patients in greater need represents allocative inefficiency.

In his theoretical model, Ellis (1998) found that prospective payment systems fare worst in terms of over-provision of services to low-severity patients (cream-skimming) and under-provision to high-severity patients (skimping). Cost-based reimbursement (such as FFS) leads to cream-skimming for all patient types. Therefore, Ellis argues that mixed payment schemes are preferable, as first-best solutions are not attainable. The underlying result suggests that case-mix cannot be fully compensated without unintended consequences in treatment selection, resulting in allocative inefficiency.

These predictions have been empirically confirmed. Eijkenaar et al. (2013) found evidence that unincentivised care deteriorates, and incentivised care is overprovided in P4P schemes. However, according to Shen et al.'s (2004) findings, this may not constitute a large welfare loss. In studying treatment decisions among physicians under FFS and capitation, they found no difference in the provision of treatments with 'undeniable' benefits. Treatments with small or questionable benefits were offered more extensively among physicians under FFS contracts.

Similarly, Petersen et al. (2006) found that although effort is generally increased, it may be only on paper. More specifically, they point to three studies showing that financial incentives led to an increase in documenting preventive efforts but not in the related service provision. Such unintended behaviour is often referred to as 'gaming the system' or 'upcoding' and is a clear example of technical inefficiency.

While remuneration schemes that adjust for case-mix can overcome some of these issues, case-selection is argued to be an inherent problem in all remuneration systems as care providers will always have more information than patients and payers. Gaming stems from these same information asymmetries but is likely more prevalent in performance-based systems where a clear incentive exists.

Despite the inherent potential in performance-based systems for unintended consequences in terms of inefficient resource use, policymakers have levers beyond fee-scheduling to limit their impact and steer GP behaviour towards delivering the desired services from the payer's perspective. In the next two sub-sections, we discuss issues related to gaming the system and skimping on care and present mitigating tools available to policymakers.

5.1 Curbing excessive service provision

While guidelines for medical practice and ethical conduct exist, there is in practice little direct oversight of GPs' daily work. Working as private practitioners, GPs in most countries are entrusted to self-report their activities, which, in the case of FFS, is the basis for reimbursement. This creates an incentive to game the system. This may involve the grey area of 'upcoding' patient health states from 'mild' to 'severe' to unlock higher fees or the more fraudulent behaviour of asking for reimbursement for treatment that never took place.

Upcoding has been uncovered in hospital settings, where fixed reimbursements are provided based on diagnosis related groups (the 'DRG' system) (Silverman and Skinner, 2004). Mitigating interventions to introduce patient risk adjustment were also found to invoke gaming in terms of actively enrolling 'overpriced' patients (Brown et al., 2014). In a primary care setting, Gravelle et al. (2010) studied changes to practice style following the 2004 introduction of a P4P scheme for British GPs. The scheme rewarded GPs for the proportion of patients treated for 11 disease domains, where patients deemed not to benefit from the treatment were recorded as exceptions. The authors found that exception rates increased significantly following the programme's introduction, suggesting that GPs were motivated by gaming the system to increase incentive payoffs. Brunt (2011) found evidence that 15% of expenditures for general office visits for outpatient care were due to upcoding treatment intensity. Through an audit of Norwegian GPs in 2020, HELFO uncovered incorrect coding practices, resulting in nine doctors being reported to the police and eight having their medical licenses revoked (Norwegian Directorate of Health, 2020).

However, for many payers, it may generally be difficult to prove that upcoding or gaming practices are intentional and constitute fraudulent behaviour (Bauder et al., 2017). For example, upcoding may not be intentional but a consequence of bounded rationality. Bounded rationality is the cognitive inability or unwillingness to assess all possible options in decision-making, resulting in 'satisfactory' rather than perfect outcomes (Cawley and Ruhm, 2011). Many FFS systems are fairly granular by design to account for the wide number of potential services. In Norway, there are more than 200 different codes associated with the FFS system for GPs (Norwegian Medical Association, 2019b). While this level of specificity can succeed in providing an accurate reflection of service provision, it relies heavily on the GPs' ability to use it properly.

While incorrect coding or excessive treatment may not harm a patient, it can result in inefficient resource use. Thus, several tools have been used to affect GPs' behaviour. These range from supportive efforts that tackle bounded rationality through information sharing to disciplinary corrections that expose incorrect practices and provide punitive measures.

Supportive efforts are often part of general service agreements between a payer and a provider of care. In Norway, this includes optional and mandatory training courses in fee usage, as well as informational material regarding updates to existing fees (The Norwegian Directorate of Health, 2009). If there is suspicion of extensive misuse of fees, the healthcare administrator may also initiate informational campaigns or more targeted audit and feedback measures. Ivers et al. (2012) defined audit and feedback as a process where 'an individual's professional practice or performance is measured and then compared to professional standards or targets'. In their literature review, they estimated a 4.3 percentage point increase in compliance with desired practice. The relatively modest effect size may be due to the lack of consequences for individual GPs. Introducing peer scrutiny or the potential for financial retribution has been found to amplify the effects.

In a laboratory experiment, Godager et al. (2016) studied the effect of providing performance feedback. In their study, participants were told in advance that relative performance information would be shared with their peers. In response to the intervention, the authors found that individuals in the exposure group increased the proportion of correct diagnoses. In a field experiment in China, Cheo et al. (2020) studied the effect of a mystery shopper scheme. They found that announcing a forthcoming performance audit on prescription use resulted in fewer prescriptions. In other words, being reminded about the possibility of an audit can be just as effective as the audit itself.

Research question

In this section, we want to study the effect of informing GPs about the appropriate use of a double-consultation fee, which is used for long consultations. More specifically, we ask the following:

- *Does informing GPs about the correct use of a double-consultation fee reduce how frequently it is used?*

Several studies have explored the effect of messages related to adherence to clinical guidelines (i.e. Ivers et al., 2012). Our study differs in that we do not refer to a change in practice; we only point to situations where consultations should be coded as double consultations. Accordingly, we try to elicit a change in incorrect coding. Although similar experiments have been used to measure the effect on

misstating tax returns (i.e. Bott et al., 2019), our study represents a novel approach in the primary care setting.

Paper 3

Introduction

Audits and feedback are used as strategies to guide healthcare professionals' practices towards certain targets. The Norwegian Health Economics Administration (HELFO) reimburses GPs for the FFS component of their remuneration and provides oversight and training to ensure that fees are used according to the rules and regulations. HELFO has previously uncovered incorrect use—and sometimes misuse—in fee practices among GPs. The fee used for double consultations (DCF) has been a particular concern, as its frequent use makes it vulnerable for incorrect application. It can be used for consultations with a duration of more than 20 minutes and stands at €21. To study the effects of an informational campaign targeting use of DCF, we partnered with HELFO to conduct a nationwide field experiment to study the behavioural responses of providing GPs with feedback on their fee claims.

Methods

The intervention was targeted at the 691 GPs who most frequently claimed DCF fees. We randomised GPs into a control group and either one of two intervention groups, 'mild' and 'strong'. Both intervention groups received an email from HELFO. The email contained information about appropriate use of the DCF fee and how much that a GP had used it in comparison to their peers. While the email body received by the two intervention groups was identical, the email headings differed. The 'mild' group received *'Information about DCF',* and the 'strong' group received *Regarding your use of DCF'.* To study the intervention's effects, we used OLS regression on registry data with monthly observations for every GP in the study from prior to and after the intervention. We used dummy-variables for post-intervention months to study the duration of any effects.

Results

The intervention caused a 4–5 percentage point reduction in use of the double-consultation fee in the first four months after the emails. The effects persisted over time, with significant negative effects found in months 5–9 and 10–14 after the intervention. There was no statistical difference in the effects between 'mild' and 'strong'. We also found a borderline significant effect for the control group in the first four months following the email. We attribute this to information sharing about the letter in closed Facebook groups. We further estimate that the intervention led to a reduction in the Norwegian government's yearly healthcare spending of approximately €877 000 (or €1 270 per GP).

Conclusion

Our results show that there are substantial and durable behavioural responses from low-cost interventions via email. The effects can have a significant financial impact with the potential for additional cost-savings if scaled to the entire GP population. The intervention email only advocated using the fee correctly and made no reference to changing practice styles. However, GPs may have responded by reducing the number of long consultations rather than undertaking the same consultations and coding them correctly. As we cannot identify the mechanisms in our results, it is not possible to establish whether the resulting fee usage constituted a net welfare gain.

5.2 Responding to patient needs

GP practices are characterised by a hectic work environment with time pressure, high levels of responsibility, and competing demands (Odebiyi et al., 2022; Trønderopprøret, 2018). This requires that GPs work effectively but also prioritise according to the more urgent needs. Whereas GPs are given the discretion and confidence to make these prioritisations according to ethics and treatment guidelines, the fact that certain tasks and performance are measured and financially incentivised may affect treatment decisions. For example, GPs under time pressure or in disagreement about the relevance of certain treatment regimens have been found to deprioritise patient groups (Arreskov et al., 2019; Sandelowsky et al., 2016).

While it can be effective to outsource clinical decision-making to a GP regarding information asymmetry, society's value judgements may differ from those of GPs in terms of who gets treated (first). Patients with lower incomes have been found to receive shorter consultations and fewer medical tests, underpinning a socioeconomic gradient in GP treatment decisions (Brekke et al., 2018). In other cases, personal convictions may lead to a refusal to provide care. In a study of Norwegian GPs, seven doctors who refused to refer patients for abortions stated that religious beliefs were their reason for defying the government mandate to do so (Nordberg et al., 2014).

Extending these findings to theoretical model (i), we may say that the altruistic component of a given GP is sensitive to the patient population being treated. In correcting for imbalance in access to care, policymakers can try to affect the inputs in a GPs choice architecture. If a GP's prioritisation is based on incomplete information, providing the missing information can make a GP more amenable to acting in patient interests. Studies on mandating access through provisions in medical guidelines can be effective, but disseminating updated information and regulations can also be difficult (Grimshaw et al., 2004). In some cases, certain GPs may even find new treatment guidelines undesirable, as has been the case with access to abortion (Magelssen, 2018).

In cases where mandates are unsuccessful, another option is to increase the payoff associated with effort towards certain patient groups. Scott et al. (2009) evaluated an incentive programme for GPs to improve treatment of chronic care patients. Following the programme's introduction, the likelihood of ordering HbA1c tests for diabetic patients increased by 20%. The programme's significant effect may be explained by the relatively sizeable incentive. GPs were compensated A$40 for follow-up of diabetes

patients, and income from the programme has constituted around 10% of GP earnings. The mixed results in other studies have been attributed to different sized bonus payments (Petersen et al., 2006).

As discussed, failure to treat patients may be grounded in supply-side issues related to trade-offs at the GP level. However, it may be equally explained by imperfect information or behavioural issues on the patient side. A large literature has been devoted to study the effect of reminders for scheduled appointments. In a literature review Hasvold and Wootton (2011) found that attendance at hospital appointments could be improved by 30% using SMS reminders and 39% using automated telephone calls. In these studies, an implicit assumption is that patients are already aware of and presumably have agreed to an appointment. This may explain the relatively large effects.

In other circumstances, a person may have weighed the costs and benefits of asking for a GP appointment based on incomplete information. Often, there is no set interval for GP checkups for patients not in a risk group; however, GPs and patients themselves may be unaware that they are in a risk group. For example, it is recommended that men in Norway have PSA testing if they have two or more relatives diagnosed with prostate cancer before the age of 60 (The Norwegian Directorate of Health, 2010). This recommendation may be unknown to patients who could ask for one, or a GP may not be informed about the patient's hereditary predisposition and fail to offer one. In response to these issues, it is not uncommon for governments to have information campaigns; examples include awareness of skin cancer (Iannacone and Green, 2014) and smoking cessation (The Norwegian Directorate of Health, 2022b). In studying the success factors of health campaigns, Randolph and Viswanath (2004) highlight the need for targeted messaging with actionable recommendations.

While the remaining campaigns mentioned succeeded in identifying a target audience of forgetful patients, they failed at recruiting previously unreached patients. Whereas broader information campaigns may succeed in the latter, their broad catchment group may limit their usefulness in targeting smaller patient groups.

Research question

In this section we want to explore the effect of information campaigns to increase the uptake of recommended services for people with type 2 diabetes (T2D). Specifically, we ask:

- Does sending email and SMS-based information to people with T2D about the benefits of annual checkups increase attendance?

Although a specific fee has been introduced in Norway for annual checkups, patient attendance remains low. An explanation may be that GPs are not sufficiently incentivised to invite their patients or patients are not sufficiently informed about the benefits. We therefore want to study the effect of a 'nudging' intervention, where members of the Diabetes Association of Norway (DAN) are given information about the benefits of checkups and urged to have one.

Unlike the previous studies discussed, we target a well-defined sample with a design that is easily scaled to larger patient populations.

Paper 4

Introduction

The purpose of this paper was to study the effect of an empowerment campaign targeting people in Norway with type-2 diabetes (T2D). In Norway, it is recommended that people with T2D have annual checkups with their GP. Coupled with proper self-management of the disease, GPs play an important role in screening for ailments caused by the disease and making plans for preventive measures (BMJ Best Practice, 2022). Despite being part of the GP guidelines and the introduction of a specific fee, many GPs fail to invite their patients with T2D to annual checkups. We partnered with DAN to study whether an informational campaign targeting their members would increase attendance to annual checkups.

Methods

With DAN's help, we identified 12 484 members who stated they had T2D and had provided contact information in the form of a telephone number and/or an email address. We randomised two separate groups into control and intervention groups: group 1 had telephone and email information and group 2 had only telephone information. The intervention consisted of an informational email or SMS (depending on whether the individual was in Group 1 or 2), followed by an SMS reminder a week later. For Group 1, the first email contained the heading 'Have you gone for your annual diabetes checkup?' The short text contained information about T2D, the benefits of having an annual checkup, and a link to DAN webpages that included an information video. Group 2 received an abbreviated version of the email as an SMS. To study the intervention's effects, we sent identical questionnaires to each of the four sub-groups (by group and intervention/control) four weeks after the first intervention. A reminder regarding the questionnaire was sent one week later, mimicking the timelines of the intervention. The response rates for groups 1 and 2 were 31% and 22%, respectively. The response rates for group 1 increased by ~300% following a reminder delivered by SMS.

Results

To estimate the effects of the intervention, we first pooled the responses from the two sub-groups and used personal characteristics to correct for imbalance in the data. Using a probit model, we found the intervention's overall effect to be 3.5%, significantly different from that of the control group. In sub-group analyses, we found a significant effect in group 1 (4.2%) but not in group 2 (3.7%). Given the strong predictive power of having previously gone to a checkup, we also investigated the effect of the intervention contingent on having gone to a checkup before. We could not establish whether the effects differed depending on previous checkups.

Conclusion

Our results suggest that information campaigns can effectively increase attendance to annual checkups. Given the relatively small sample size, we could not establish whether SMS was superior to email. Considering the increase in take-up rates following the SMS reminder for Group 1, we are inclined to interpret SMS as more favourable than email. Acknowledging that DAN members are expected to be well-informed about T2D from the outset, it is hard to extrapolate our findings to a larger population. If anything, we believe our findings reflect a lower bound.

6. Discussion

This thesis investigated policy measures and interventions to increase the supply of GP services and ways to ensure that these services match the needs of patients and the intentions of payers. In Paper 1, we presented a theoretical model to illustrate how policy changes in the GP system are predicted to affect effort levels and the profession's attractiveness. We showed that some payment mechanisms and policies cause diverging results regarding these two outcomes. In Paper 2, we empirically tested some of the predictions made in Paper 1. We first showed which GP characteristics are associated with higher income (and effort levels) before using an IV approach to estimate the causal effect of lower income on quitting. In Papers 3 and 4, we discussed alternative policy interventions for steering GP behaviour to ensure access for specific patient groups. In Paper 3. we presented the results of an information campaign targeting GPs with high use of double-consultation fees, showing a reduction in the use of such fees following the intervention. We further quantified the money saved from the reduction and discussed the implications regarding the social welfare effect. In Paper 4, we estimated the effect of a patient-empowerment nudging campaign targeted at people with T2D. The study showed that email- and SMS-based interventions can be effective for increasing attendance to annual checkups.

Asymmetric information is a central component of how healthcare and GP systems are organised. We have shown how information asymmetry gives rise to agency problems related to ensuring sufficient and efficient provision of services that in turn match patients' needs. In response, payers can introduce incentives and organisational models that align GPs' incentives with their own. To avoid unintended consequences, rewards must reflect the patient case-mix a GP faces. Moreover, GPs must be provided with support and be under some degree of oversight to ensure that fees are used in line with their intended purpose and that weaker patient groups are not neglected. Nevertheless, the degree of such measures must be weighed against GPs' desires for autonomy and flexibility. Ultimately, we find that the objectives of efficiency, the profession's attractiveness, and access to services may be competitive, meaning trade-offs have to be made.

Paper 1

This paper builds on previous theoretical depictions of GP behaviour but adds critical components of GP motivation such as working conditions and practice quality. In addition to predicting optimal effort levels, it shows how different policy measures affect decisions to enter or stay in the GP profession.

In the paper, we extend the literature by explicitly depicting the most applied GP remuneration schemes. Previous studies have either focused on general remuneration schemes for healthcare services (i.e. Ellis and McGuire, 1990) or on private insurance markets (i.e. McGuire, 2000). Our predictions support previous findings but provide a more encompassing model for policymakers who are interested in increasing the number of GPs, not just how hard they work.

We find that remuneration often has opposing effects on effort and the profession's attractiveness. Higher capitation leads to reduced effort due to a decrease in the relative importance of income from FFS; GPs opt for more leisure instead. However, the same policy change is expected to increase the profession's attractiveness, making it easier to recruit and retain staff. This is consistent with a recent policy proposition from the Norwegian Government that aims to increase capitation for the first 1 000 patients on a GP's list to facilitate a smaller burden of work without loss of income (The Norwegian Directorate of Health, 2022c). Regardless of the number of patients on a GPs list, the policy would represent a net income gain. As per the policy aims, our model would predict lower effort but increased retention and recruitment.

Another prediction of our model is that the effect of FFS on effort and utility is a function of a GP's effort level (and thereby that GP's earnings) relative to that of their peers. Low performers will exert more effort as a result of an FFS increase but experience reduced utility. The implication is that an FFS increase can drive out those in need of motivation while reducing effort levels among those that stay. This effect is offset by ensuring that new entrants would be motivated to work hard.

How practice quality affects effort and utility is indeterminate in our model. One of the main drivers is the extent to which GPs would like to see more patients. If a GP is not incentivised financially and/or has little altruistic motivation, improved practice quality will lead to reduced effort since they no longer have to work as hard to attract new patients. A similar intuition follows for GP utility; those who benefit from increased patient contact are better off. An example of this finding relates to the introduction of e-consultation services, which have seen large growth in Norway (Wynn, 2020). As e-consultations reduce transaction costs for patients, demand increases, leading some GPs to reduce effort levels.

This result can also be generalised to demand for GP services; other things being equal, higher demand reduces the importance of attracting new patients. In Norway's case, the GP system was introduced to induce competition between GPs to attract patients. In recent years, however, GP scarcity has removed the competitive element between GPs; skimping on effort will probably not reduce the likelihood of

attracting patients as long as there are patient waiting lists. Conversely, any unmet demand may be sub-optimal for some GPs, as their inability to care for all patients drives them to work harder than desired. In fact, the only policy measure that improves both effort *and* utility is improving working conditions. The theoretical explanation is that it does not affect the likelihood of attracting patients, which will—in most cases—have opposing effects on effort and utility. The effect only works through a GPs' cost of effort. The practical and somewhat more complicated interpretation is that working conditions will in reality be a function of the remuneration scheme that a GP operates under. Nonetheless, policy levers, such as reducing administration or instituting predictable working hours, may work independently of payment schemes.

Paper 2

The second paper in the thesis investigated the determinants of earnings and estimated the effect of income on quitting. Using rich registries of GP characteristics and practice styles over a 10-year period, we estimated the determinants of income differentials and the effect of income on quitting.

The analysis of income determinants shows that our theoretical model's main predictions in Paper 1 empirically hold. As we look at income per consultation and not aggregate income, the results represent a higher effort level while at work and not whether a GP works longer hours. Nevertheless, we find that GPs with a higher volume of services also earn more per consultation. More specifically, those with more consultations per listed patient, more overtime, and a longer list length earn more per consultation. We cannot say whether this means that those exposed to higher demand respond by working more efficiently or that it reflects a sample of GPs who are more financially motivated and adapt their practice styles based on the intensive and extensive margins. The list length's positive effect on income suggests that the FFS elements offset any incentives to skimp on service provision. Finally, survey responses suggest that most GPs in Norway desire to reduce list length but choose not to do so to avoid income loss and/or fear that patients will lose access to GP services (Pedersen et al., 2022).

We also find that older, tenured, and specialist GPs earn more than their counterparts. This is unsurprising, given that older GPs have had more time to acquire job-specific knowledge regarding treatment decisions and fee use. Unsurprisingly, salaried GPs earn less, likely because of the lack of incentive for effort. A younger patient pool is also associated with lower earnings, something we attribute to the reduced treatment intensity needed for that group compared to older patient groups.

As for quitting behaviour, our IV analysis suggests that lower income leads to increased probability of quitting. The effect size is estimated to be a 10 percentage points drop in the probability of quitting for a 10 percent increase in average income per consultation. Moreover, we find that a smaller share of long consultations and working as a salaried GP are associated with lower likelihood of quitting. This is consistent with our theoretical model's predictions, where salaried GPs see a monotonic increase in utility, while the effect of FFS is inconclusive. As for a high proportion of long consultations, it may be an artefact of GPs who prefer more patient contact being punished financially in an FFS system for foregoing additional consultations with other patients.

One of our study's main limitations is that we cannot determine the underlying mechanisms behind our results, a general weakness of observational studies (see section 3). For example, unobserved heterogeneity among individuals who choose a salaried contract may drive the differences in incomes and thus effort levels rather than the incentives associated with the different remuneration schemes. Moreover, while the IV should, in theory, isolate the effect of income on quitting, this approach has weaknesses. First, an instrument's validity will always contain a degree of judgement. We have argued why a GP's inability to affect their patient pool—at least for the diagnoses studied—makes the applied diagnosis bundles exogenous and thereby relevant for our case. Moreover, we argue that the natures of the diagnoses are unlikely to affect the outcome variable (quitting). Second, although we found that income affects quitting behaviour, we cannot determine the underlying mechanism. Based on our theoretical model in Paper 1, we think that an outside option would generate higher utility based on the optimal effort levels in a GP setting. Whether this effort level is driven by lower altruism or different (preferences for) working conditions or leisure remains unknown.

Despite these limitations, we believe the paper constitutes a significant contribution to the literature. To our knowledge, no other studies have been able to quantify the effect of income on quitting behaviour. Moreover, by studying income *per consultation*, we could verify that FFS is a driver for effort levels on the intensive margin: put differently, working faster and not just more.

Paper 3

The third paper investigated the effect of an information campaign targeted at GPs with high use of double-consultation fees (DCF). DCF can be used if a consultation has a duration of 20 min or more. We found that following the receipt of an email with information about the GP's relative use of DCF, we

observed a reduction of ~4–5 percentage points. The effect was long-lasting and did not depend on whether the message addressed the GP specifically or was generally worded.

Despite numerous studies on the effect of audits and feedback, they have largely been focused on high-intensity interventions (such as telephone or in-person visits) or altering clinical practice. Our study represents a slightly different approach—we do not ask GPs to change how they treat patients, just ensure that consultations are coded correctly. By partnering with HELFO, we also ensure that the study objects are in their usual environment and unaware of the study, representing a 'natural field experiment'.

Our experiment points to an important limitation in an FFS system—fair remuneration is based on the assumption that the range of fees i) accurately reflects the range of treatments performed by GPs and ii) that they are used correctly. For example, Chan et al. (2019) showed that hospital reimbursement schedules for surgeries are based on incorrect estimates of the associated time and effort involved. Accordingly, they conclude that some surgeries are being overcompensated, while others are being undercompensated. In the US, billing complexity has been found to explain up to 18% of claims challenged under Medicare, suggesting incorrect use (Gottlieb et al., 2018). Heath (2020) argued that it is difficult to uncover whether the use of certain fees is justifiable because medical professionals often write clinical records themselves. Accordingly, an auditor in most cases can only see whether fees are consistent regarding the *stated* procedures. The same problem exists in our study; we could not determine the underlying drivers because we cannot observe GPs in their office.

Heath (2020) proposes a typology to distinguish between types of unintended fee use: 'gamesmanship', 'fraud', and 'honest fraud'. In our case, gamesmanship refers to the GP keeping a patient at their office for longer than required to break the 20-minute mark. While consistent with the rules, it defeats the fee's purpose and may not add any value to patient welfare. 'Fraud' involves stating that a 21-minute consultation took place when the GP knew it lasted only 19 minutes, violating the rules. Finally, 'honest fraud' occurs if a GP feels poorly compensated for a complicated consultation and adds the DCF fee as a form of entitlement to offset hard work that, in the eyes of the GP, has not been fully compensated.

Although this study does not make judgements regarding the gravity of the distinct types of behaviour, it notes that any such behaviour is unlikely to improve patient welfare. In addition to the mechanisms discussed, an alternative representation of the observed effects in our experiment is that the DCF is used as a substitute for other fees. Faux et al. (2021) showed that doctors often lack billing literacy and

argued for stronger educational support. Therefore, for our setting, any incorrect use of DCF may be due to the inability to identify more case-specific fees under time constraints.

Finally, despite the letter referring to coding practices and not treatment, some GPs may have responded by reducing the time spent with patients due to fear of audits. Ultimately, the effect on net welfare will depend on which of the mechanisms are (most) at work. Regardless, this points to one of the agency problems of an FFS system: it relies on accurate reporting, and the size of the fees reflects the effort associated with different treatments.

Paper 4

In the last paper, we studied the effect of an information campaign to increase attendance to recommended annual checkups by people with type 2 diabetes. The results suggest that email and SMS can be effective in increasing attendance to recommended care. To interpret the results, it is important to understand why people with T2D do not attend checkups in the first place. We argue that it is due to supply side or demand side issues—that is GPs failing to invite patients for checkups, or patients not desiring them.

On the supply side, GPs may be unaware of guidelines or unwilling to follow them. While guidelines for T2D have been in place for a long time, information overload can make it difficult for GPs to stay abreast of optimal care (Klerings et al., 2015). In Paper 3, we found that information campaigns can be effective for disseminating information and altering GP behaviour, but GPs may still choose to abstain. One plausible reason is that GPs may not want to create extra demand if they are already experiencing high workloads. While incentive schemes can be effective for inducing improved diabetes treatment in other settings (Lai and Hou, 2013), the checkup-specific fee introduced in Norway has not been as successful. One possible reason is that the fee does not sufficiently compensate for the level of treatment. Put differently, the FFS is unsuccessful in appropriately adjusting for patient case-mix, leading GPs to skimp on service provision. To this effect, Denmark introduced a new capitation scheme in 2018 where additional funding was provided based on the number of people on a GP's list that have diabetes. The scheme resulted in providing more care for 'low-needs patients', while no difference was found for 'high-needs patients' (Pulleyblank et al., 2020). Norway has recently introduced capitation adjustments based on age, sex, location, socioeconomics, and expected use of GP services (HELFO, 2023).

On the demand side, co-payments can in theory act as a deterrent for patients. In practice however, around a fourth of Norwegian citizens have reached the co-payment ceiling (Norwegian Parliament,

2020), suggesting that chronic patients do not regard costs as a restriction. Therefore, we hypothesise that non-attendance to checkups is the result of being unaware of the recommendation or deeming it unnecessary; thus, a natural solution is to provide information. Depending on the reasons for not attending, information may comprise a reminder or a convincing statement that challenges prior reasons for non-attendance. This is also the purpose of our study: to inform and convince people with T2D to attend annual checkups. Information can be given in different forms; in-person education programmes (Loveman et al., 2008) and web-based interventions (Samoocha et al., 2010) have been found to be effective in diabetes self-care. Telephone and SMS reminders have also been shown to improve attendance at hospital appointments (Hasvold and Wootton, 2011). The limitation of these approaches however, is that they are concerned with people who have already decided to seek care. Our study targets patients who have not been reached or may need convincing.

The estimated result of a 3.5% increase in attendance implies that more than 400 additional people would attend annual checkups if the entire sample of 12 484 had received the intervention. As noted in section 3, extrapolating these results to the entire T2D population is difficult. First, our sample from the Diabetes Association of Norway is arguably more informed about the importance of screening and annual checkups than the entire T2D population. Second, a substantial proportion of our sample had previously gone to a checkup, suggesting that the intervention may have led them to move up a checkup that the GP would have scheduled anyway. Therefore, it is difficult to say whether the estimate represents a higher or lower bound if the entire T2D population in Norway was targeted. Despite these limitations, our study shows that nudging campaigns to increase awareness of optimal care can be effective in strengthening patients' agency towards their health. Given the small costs involved with such a campaign, it is also likely a highly cost-effective initiative. Finally, we argue that despite our methodology's limitations regarding extrapolating results, it would be difficult to perform a population-wide experiment because of ethical and financial concerns. Using an observational design on non-randomised data following a larger campaign would also have made it difficult to draw causal links from the intervention to the observed behaviour.

7. Conclusion and policy recommendations

In this thesis, we have discussed how different GP remuneration systems can achieve an increased supply of services and how policymakers can achieve the desired level of service provision by going beyond financial incentives. To provide a comprehensive assessment of GP and patient responses to different policy measures, we applied several research methodologies, including a theoretical model, observational data, and experimental methods. In this research, we found that incentives to increase the supply of services among existing GPs may drive some GPs out of the profession and deter others from entering it. Moreover, we exposed some of the unintended consequences of an FFS system regarding billing and patient selection and measured the effect of interventions to offset these.

We hypothesise that the asymmetric information inherent in GP systems makes it difficult for GPs and the public to fully agree on which and how many patients get treated using what resources. Different preferences and motivations can explain why some GPs do not offer annual checkups unless prompted to by empowered patients or why some use fees excessively without government intervention, while others do not. In our theoretical model, individual altruism was an important predictor of effort and utility. However, failing to provide people with T2D annual checkups does not necessarily mean that a GP is not acting altruistically. Conversely, a GP may decide that the opportunity cost of an annual checkup is delaying care for other patients with more urgent needs. As such, the GP would be acting altruistically, just not in the views of society, who, in turn, might be casting judgement based on imperfect information. Nonetheless, GPs differ in their medical judgements, which are at times based on personal views (i.e. abortion). As such, third parties who provide patients with information may serve as tools to enable access to care with a given GP—treatment that would have been offered unprompted by others. In other circumstances, access to GP services is not necessarily a consequence of a GP's prioritisation but that of patients failing to recognise that they should ask for it.

Just as some patients do, GPs themselves may lack information. For example, a given GP may not have long-term population health information and the need for preventive treatment front-of-mind when no effect is seen instantly. The same goes for public resources, which are scarce and should be used efficiently. As such, there is a need for societal control to ensure the equitable and sustainable use of resources. The organisation of many systems where GPs work as private practitioners stresses the point. Whereas doctors working at hospitals or as bureaucrats have organisational controls with learning and development, feedback, and wage-setting in-house, this is outsourced for GPs.

In many respects, the FFS system inhibits many features related to oversight, as pay is linked to performance. As performance can be subjective and differ depending on the particular GP, there is arguably an equitable element in having all GPs measured against the same metrics. We have shown that an FFS system goes farther in aligning GPs' incentives with those of payers than other payment schemes but not without unintended consequences. For example, Paper 2 revealed that the FFS system generates significant differences in income per consultation, leading some GPs to quit. While we cannot say whether this is due to patient complexity, we hypothesise that some GPs may feel unfairly treated. Accordingly, governmental oversight can be important for ensuring both patient access and GP motivation. The tricky part is avoiding undue interference with GPs' autonomy, a feature valued by many.

We argue that our findings expose an inherent limitation of GP systems in general: there is no such thing as a perfect system, albeit there may be one for some objectives. Returning to our understanding of increasing capacity, i) getting doctors to choose general practice as their career and ii) incentivising efficient supply services, we argue that trade-offs must be made, especially when matching GP services with patient needs. What motivates doctors to choose a GP career may sometimes lead to sub-optimal effort levels, poorer patient access, or inefficient use of resources. Preferences are also heterogeneous, forcing policymakers to decide who and which objectives are more important. One option, therefore, is to provide GPs with a suite of different contracts, enabling them to self-select the one that maximizes their utility.

In our view, the reduced access to GPs over time regarding list capacity suggests that the current mixed remuneration system has been unsuccessful in meeting demand. Accordingly, we argue that policymakers should focus on changes that lead to happier rather than more efficient GPs. An efficient GP who ends up quitting creates transaction costs, in addition to defeating the system's purpose of continuity of care. Finally, it is important to acknowledge that the demand for GP services is not only driven by patients' health, but the organisational needs as defined by the government. In some instances, GPs have a monopoly in issuing sick leaves, referrals, and treatments. Relieving GPs of obligations that are of little societal benefit or that can be performed by others can make life at the GP office more meaningful. Coupling this with changes that also make it more bearable can improve working conditions. Ultimately, this is the only policy that will unequivocally improve GP effort *and* welfare.

References

Abelsen, B., Gaski, M., Brandstorp, H., 2015. Duration of general practitioner contracts. Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin 135, 2045–2049. https://doi.org/https://doi.org/10.4045/tidsskr.15.0003

Abelsen, B., Olsen, J.A., 2015. Young doctors' preferences for payment systems: the influence of gender and personality traits. Human Resources for Health 69. https://doi.org/https://doi.org/10.1186/s12960-015-0060-0

Al-Ubaydli, O., Lee, M.S., List, J.A., Mackevicius, C.L., Suskind, D., 2021. How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling. Behavioural Public Policy 5, 2–49. https://doi.org/10.1017/bpp.2020.17

Al-Ubaydli, O., List, J., 2015. On the Generalizability of Experimental Results in Economics: With A Response To Camerer, in: Handbook of Experimental Economic Methodology. https://doi.org/10.3386/w19666

Allard, M., Jelovac, I., Léger, P.T., 2011. Treatment and referral decisions under different physician payment mechanisms. Journal of Health Economics 30, 880–893. https://doi.org/10.1016/j.jhealeco.2011.05.016

Allen, T., Whittaker, W., Sutton, M., 2017. Does the proportion of pay linked to performance affect the job satisfaction of general practitioners? Social Science and Medicine 173, 9–17. https://doi.org/10.1016/J.SOCSCIMED.2016.11.028

Anglemyer, A., Horvath, H.T., Bero, L., 2014. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database of Systematic Reviews 2014. https://doi.org/10.1002/14651858.MR000034.PUB2/MEDIA/CDSR/MR000034/IMAGE_N/NMR000034-CMP-001-04.PNG

Arreskov, A.B., Graungaard, A.H., Kristensen, M.T., Søndergaard, J., Davidsen, A.S., 2019. General practitioners' perspectives on chronic care consultations for patients with a history of cancer: A qualitative interview study. BMC Family Practice 20, 1–11. https://doi.org/10.1186/S12875-019-1009-5/FIGURES/1

Arrow, K.J., 1978. Uncertainty and the welfare economics of medical care, in: Uncertainty in Economics. Elsevier, pp. 345–375. https://doi.org/https://doi.org/10.1016/B978-0-12-214850-7.50028-0

Baiocchi, M., Cheng, J., Small, D.S., 2014. Instrumental variable methods for causal inference. Statistics in medicine 33, 2297–2340. https://doi.org/10.1002/SIM.6128

Bauder, R., Khoshgoftaar, T.M., Seliya, N., 2017. A survey on the state of healthcare upcoding fraud analysis and detection. Health Services and Outcomes Research Methodology 17, 31–55. https://doi.org/10.1007/S10742-016-0154-8/TABLES/2

Bénabou, R., Tirole, J., 2003. Intrinsic and Extrinsic Motivation. The Review of Economic Studies 70, 489–520. https://doi.org/10.1111/1467-937X.00253

Bennett, K.L., Phillips, J.P., 2010. Finding, recruiting, and sustaining the future primary care physician workforce: a new theoretical model of specialty choice process. Academic Medicine 85, S81–S88.

https://doi.org/https://doi.org/10.1097/ACM.0b013e3181ed4bae

Black, N., 1996. Why we need observational studies to evaluate the effectiveness of health care. BMJ 312, 1215–1218. https://doi.org/10.1136/BMJ.312.7040.1215

BMJ Best Practice, 2022. Type 2 diabetes mellitus in adults - Symptoms, diagnosis and treatment [WWW Document]. BMJ Best Practice US. URL https://bestpractice.bmj.com/topics/en-us/24 (accessed 2.13.23).

Bott, K.M., Cappelen, A.W., Sørensen, E.O., Tungodden, B., 2019. You've Got Mail: A Randomized Field Experiment on Tax Evasion. Management Science 66, 2801–2819. https://doi.org/10.1287/MNSC.2019.3390

Brekke, K.A., Nyborg, K., 2010. Selfish bakers, caring nurses? A model of work motivation. Journal of Economic Behavior & Organization 75, 377–394. https://doi.org/https://doi.org/10.1016/j.jebo.2010.05.008

Brekke, K.R., Holmås, T.H., Monstad, K., Straume, O.R., 2020. How does the type of remuneration affect physician behavior? American Journal of Health Economics 6, 104–138. https://doi.org/10.1086/706624

Brekke, K.R., Holmås, T.H., Monstad, K., Straume, O.R., 2018. Socio-economic status and physicians' treatment decisions. Health Economics 27, e77–e89. https://doi.org/10.1002/HEC.3621

Brembo, F., Andersen, B., 2022. Står tiltalt for 33 000 tilfeller av bedrageri: – Alvorlig tiltale. NRK.

Brown, J., Duggan, M., Kuziemko, I., Woolston, W., 2014. How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program. American Economic Review 104, 3335–64. https://doi.org/10.1257/AER.104.10.3335

Brunt, C.S., 2011. CPT fee differentials and visit upcoding under Medicare Part B. Health Economics 20, 831–841. https://doi.org/10.1002/HEC.1649

Cawley, J., Ruhm, C.J., 2011. The economics of risky health behaviors, in: Handbook of Health Economics. Elsevier, pp. 95–199. https://doi.org/https://doi.org/10.1016/B978-0-444-53592-4.00003-7

Chalkley, M., Malcomson, J.M., 1998. Contracting for health services when patient demand does not reflect quality. Journal of Health Economics 17, 1–19. https://doi.org/https://doi.org/10.1016/S0167-6296(97)00019-2

Chan, D.C., Huynh, J., Studdert, D.M., 2019. Accuracy of Valuations of Surgical Procedures in the Medicare Fee Schedule. New England Journal of Medicine 380, 1546–1554. https://doi.org/10.1056/NEJMSA1807379/SUPPL_FILE/NEJMSA1807379_DISCLOSURES.PDF

Cheo, R., Ge, G., Godager, G., Liu, R., Wang, J., Wang, Q., 2020. The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment. Health Economics Review 2020 10:1 10, 1–19. https://doi.org/10.1186/S13561-020-00290-Z

Dale, J., Potter, R., Owen, K., Parsons, N., Realpe, A., Leach, J., 2015. Retaining the general practitioner workforce in England: what matters to GPs? A cross-sectional study. BMC family practice 16, 140. https://doi.org/https://doi.org/10.1186/s12875-015-0363-1

Danish Health Authority, 2020. Opslåede og besatte hoveduddannelsesforløb 2019. København.

Danish Health Authority, 2018. Opslåede og besatte hoveduddannelsesforløb i 2017. København.

Danish Ministry of Health, 2017. Lægedekning i hele Danmark.

Deaton, A., 2020a. Introduction: Randomization in the Tropics Revisited, a Theme and Eleven Variations, Randomized Control Trials in the Field of Development. Oxford University Press. https://doi.org/10.1093/oso/9780198865360.003.0002

Deaton, A., 2020b. Instruments, Randomization, and Learning about Development. Journal of Economic Literature 48, 424–455. https://doi.org/https://doi.org/10.1257/jel.48.2.424

Deci, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology 18, 105. https://doi.org/https://doi.org/10.1037/h0030644

Delfgaauw, J., 2007. Dedicated doctors: public and private provision of health care with altruistic physicians. https://doi.org/https://doi.org/10.2139/ssrn.958693

Deloitte, 2022. Kapasitet og dekning i fastlegeordningen.

Devlin, R.A., Sarma, S., 2008. Do physician remuneration schemes matter? The case of Canadian family physicians. Journal of Health Economics 27, 1168–1181. https://doi.org/10.1016/J.JHEALECO.2008.05.006

Dionne, G., Contandriopoulos, A.-P., 1985. Doctors and their workshops: a review article. Journal of Health Economics 4, 21–33. https://doi.org/https://doi.org/10.1016/0167-6296(85)90021-9

Dow, S.C., 2007. Variety of methodological approach in economics. Journal of Economic Surveys 21, 210–230. https://doi.org/10.1111/j.1467-6419.2007.00510.x

Dugdale, P., 1997. The better practice program funding formula: A neo-liberal approach to funding general practice. Australian Journal of Public Administration 56, 65–76. https://doi.org/10.1111/J.1467-8500.1997.TB01266.X

Eijkenaar, F., Emmert, M., Scheppach, M., Schöffski, O., 2013. Effects of pay for performance in health care: A systematic review of systematic reviews. Health Policy 110, 115–130. https://doi.org/10.1016/J.HEALTHPOL.2013.01.008

Ellis, R.P., 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. Journal of Health Economics 17, 537–555. https://doi.org/10.1016/S0167-6296(97)00042-8

Ellis, R.P., McGuire, T.G., 1990. Optimal payment systems for health services. Journal of health economics 9, 375–396. https://doi.org/https://doi.org/10.1016/0167-6296(90)90001-J

Ellis, R.P., McGuire, T.G., 1986. Provider behavior under prospective reimbursement: Cost sharing and supply. Journal of health economics 5, 129–151. https://doi.org/https://doi.org/10.1016/0167-6296(86)90002-0

Eneroth, M., Gustafsson Sendén, M., Schenck Gustafsson, K., Wall, M., Fridner, A., 2017. Threats or violence from patients was associated with turnover intention among foreign-born GPs–a comparison of four workplace factors associated with attitudes of wanting to quit one's job as a GP. Scandinavian Journal of Primary Health Care 35, 208–213. https://doi.org/https://doi.org/10.1080/02813432.2017.1333319

EY and Vista Analyse, 2019. Evaluering av fastlegeordningen.

Faux, M., Adams, J., Wardle, J., 2021. Educational needs of medical practitioners about medical billing: a scoping review of the literature. Human resources for health 19, 84. https://doi.org/10.1186/S12960-021-00631-X/TABLES/1

Finnish Medical Association, 2019. Terveyskeskusten lääkäritilanne 2019 [WWW Document]. URL https://www.laakariliitto.fi/site/assets/files/5223/x_terveyskeskusten_laakaritilanne_2019_final.pdf

Fuchs, V.R., 2000. The future of health economics 1. Journal of Health Economics 19, 141–157. https://doi.org/https://doi.org/10.1016/S0167-6296(99)00033-8

Gaardsrud, P.Ø., 2021. Styringsdata for fastlegeordningen, 4. kvartal 2020. Norwegian Directorate of Health, Oslo.

Gaardsrud, P.Ø., 2020. Styringsdata for fastlegeordningen, 4 kvartal 2019 [GP management data, 4th quarter 2019]. Norwegian Directorate of Health, Oslo.

Gaski, M., Abelsen, B., 2015. Samhandlingsreformen og kommunens ansvar for integrerte legetjenester. Fagbladet Samfunn og Økonomi 52–66.

Glied, S., Smith, P.C., Jones, A.M., Rice, N., 2011. Econometric evaluation of health policies: Oxford Handbook of Health Economics, in: Oxford Handbook of Health Economics. Oxford University Press, pp. 890–923. https://doi.org/https://doi.org/10.1093/oxfordhb/9780199238828.001.0001

Godager, G., Hennig-Schmidt, H., Iversen, T., 2016. Does performance disclosure influence physicians' medical decisions? An experimental study. Journal of Economic Behavior and Organization 131, 36–46. https://doi.org/10.1016/J.JEBO.2015.10.005

Godager, G., Wiesen, D., 2013. Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. Journal of health economics 32, 1105–1116. https://doi.org/https://doi.org/10.1016/j.jhealeco.2013.08.008

González, P., Macho-Stadler, I., 2013. A theoretical approach to dual practice regulations in the health sector. Journal of Health Economics 32, 66–87. https://doi.org/10.1016/J.JHEALECO.2012.08.005

Goodyear-Smith, F., Buetow, S., 2001. Power issues in the doctor-patient relationship. Health Care Analysis 9, 449–462. https://doi.org/10.1023/A:1013812802937/METRICS

Gosden, T., Forland, F., Kristiansen, I., Sutton, M., Leese, B., Giuffrida, A., Sergison, M., Pedersen, L., 2000. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians. Cochrane Database of Systematic Reviews. https://doi.org/https://doi.org/10.1002/14651858.CD002215

Gosden, T., Pedersen, L., Torgerson, D., 1999. How should we pay doctors? A systematic review of salary payments and their effect on doctor behaviour. QJM: An International Journal of Medicine 92, 47–55. https://doi.org/10.1093/QJMED/92.1.47

Gottlieb, J.D., Shapiro, A.H., Dunn, A., 2018. The complexity of billing and paying for physician care. Health Affairs 37, 619–626. https://doi.org/10.1377/HLTHAFF.2017.1325/ASSET/IMAGES/LARGE/FIGUREEX4.JPEG

Gravelle, H., 1999. Capitation contracts: access and quality. Journal of Health Economics 18, 315–340.

https://doi.org/10.1016/S0167-6296(98)00046-0

Gravelle, H., Sutton, M., Ma, A., 2010. Doctor behaviour under a pay for performance contract: Treating, cheating and case finding? The Economic Journal 120, 129–156. https://doi.org/10.1111/j

Green, E.P., 2014. Payment systems in the healthcare industry: An experimental study of physician incentives. Journal of Economic Behavior & Organization 106, 367–378. https://doi.org/10.1016/j.jebo.2014.05.009

Grimshaw, J., Thomas, R., MacLennan, G., Fraser, C., Ramsay, C.R., Vale, L., Whitty, P., Eccles, M.P., Matowe, L., Shirran, L., Wensing, M.J., 2004. Effectiveness and efficiency of guideline dissemination and implementation strategies. https://doi.org/https://doi.org/10.3310/hta8060

Hamermesh, D.S., 1960. Six Decades of Top Economics Publishing: Who and How? Journal of Economic Literature 162–172. https://doi.org/10.1257/jel.51.1.162

Harrison, G.W., List, J.A., 2004. Field Experiments. Journal of Economic Literature XLII, 1009–1055. https://doi.org/https://doi.org/10.1257/0022051043004577

Hasvold, P.E., Wootton, R., 2011. Use of telephone and SMS reminders to improve attendance at hospital appointments: A systematic review. Journal of telemedicine and telecare 17, 358–364. https://doi.org/https://doi.org/10.1258/jtt.2011.110707

Heath, J., 2020. Ethical Issues in Physician Billing Under Fee-For-Service Plans. The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine 45, 86–104. https://doi.org/10.1093/JMP/JHZ029

Hegtvedt, K.A., 2014. Ethics and experiments. Laboratory experiments in the social sciences 23–51.

Helfo, 2023. Pasienttilpasset basistilskudd - Helfo - for helseaktører [WWW Document]. URL https://www.helfo.no/fastlegeordninga/pasienttilpasset-basistilskudd (accessed 3.29.23).

HELFO, 2022. Egenandelstak for helsetjenester.

Helseøkonomisk Analyse, 2021. Finansiering av fastlegeordningen.

Hodgson, G.M., 2008. An institutional and evolutionary perspective on health economics. Cambridge Journal of Economics. https://doi.org/10.1093/cje/bem033

Holte, J.H., Abelsen, B., Halvorsen, P.A., Olsen, J.A., 2015a. General practitioners' altered preferences for private practice vs. salaried positions: a consequence of proposed policy regulations? BMC Health Services Research 15, 1–10. https://doi.org/https://doi.org/10.1186/s12913-015-0777-4

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

Iannacone, M.R., Green, A.C., 2014. Towards skin cancer prevention and early detection: evolution of skin cancer awareness campaigns in Australia. Melanoma Management 1, 75–84. https://doi.org/10.2217/MMT.14.6

Ifcher, J., Zarghamee, H., 2016. Ethics and experimental economics. Practicing Professional Ethics in Economics and Public Policy 195–205.

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J.M., Odgaard-Jensen, J., French, S.D., O'Brien, M.A.,

Johansen, M., Grimshaw, J., Oxman, A.D., 2012. Audit and feedback: Effects on professional practice and healthcare outcomes. Cochrane Database of Systematic Reviews 2012. https://doi.org/10.1002/14651858.CD000259.pub3

Iversen, T., Lurås, H., 2000. Economic motives and professional norms: the case of general medical practice. Journal of Economic Behavior & Organization 43, 447–2681. https://doi.org/https://doi.org/10.1016/S0167-2681(00)00130-X

Jack, W., 2005. Purchasing health care services from providers with unknown altruism. Journal of health economics 24, 73–93. https://doi.org/https://doi.org/10.1016/j.jhealeco.2004.06.001

Jacob, B.J., Moineddin, R., Sutradhar, R., Baxter, N.N., Urbach, D.R., 2012. Effect of colonoscopy on colorectal cancer incidence and mortality: an instrumental variable analysis. Gastrointestinal Endoscopy 76, 355-364.e1. https://doi.org/10.1016/J.GIE.2012.03.247

Janus, K., Amelung, V.E., Gaitanides, M., Schwartz, F.W., 2007. German physicians "on strike"—Shedding light on the roots of physician dissatisfaction. Health Policy 82, 357–365. https://doi.org/10.1016/J.HEALTHPOL.2006.11.003

Jessen, B., 2018. Over 100.000 danskere står nu uden den klassiske familielæge - og tallet ser ud til at stige. Berlingske.

Keilegavlen, I., Og, R., Eikeland, O.J., 2018. Fastlegers tidsbruk.

Klerings, I., Weinhandl, A.S., Thaler, K.J., 2015. Information overload in healthcare: too much of a good thing? Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 109, 285–290. https://doi.org/10.1016/J.ZEFQ.2015.06.005

Knudsen, A.K., Allebeck, P., Tollånes, M.C., Skogen, J.C., Iburg, K.M., McGrath, J.J., Juel, K., Agardh, E.E., Ärnlöv, J., Bjørge, T., 2019. Life expectancy and disease burden in the Nordic countries: results from the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. The Lancet Public Health 4, e658–e669. https://doi.org/https://doi.org/10.1016/S2468-2667(19)30224-5

Kokko, S., 2007. Towards fragmentation of general practice and primary healthcare in Finland? Scandinavian Journal of Primary Health Care 25, 131–132. https://doi.org/https://doi.org/10.1080/02813430701576482

Krasnik, A., Groenewegen, P.P., Pedersen, P.A., V Scholten, P., Mooney, G., Gottschau, A., Flierman, H.A., Damsgaard, M.T., 1990. Changing remuneration systems: effects on activity in general practice. British Medical Journal 300, 1698–1701. https://doi.org/10.1136/BMJ.300.6741.1698

Kuusio, H., Heponiemi, T., Sinervo, T., Elovainio, M., 2010. Organizational commitment among general practitioners: a cross-sectional study of the role of psychosocial factors. Scandinavian journal of primary health care 28, 108–114. https://doi.org/https://doi.org/10.3109/02813431003779647

Kuusio, H., Heponiemi, T., Vänskä, J., Aalto, A.-M., Ruskoaho, J., Elovainio, M., 2013. Psychosocial stress factors and intention to leave job: differences between foreign-born and Finnish-born general practitioners. Scandinavian Journal of Public Health 41, 405–411. https://doi.org/https://doi.org/10.1177/1403494813477248

Lafortune, G., 2016. Analytical framework of health labour markets, in: OECD (Ed.), Health Workforce Policies in OECD Countries: Right Jobs, Right Skills, Right Places. OECD Publishing, Paris.

Lai, C.L., Hou, Y.H., 2013. The association of clinical guideline adherence and pay-for-performance among patients with diabetes. Journal of the Chinese Medical Association 76, 102–107. https://doi.org/10.1016/J.JCMA.2012.06.024

List, J.A., 2011. Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. Journal of Economic Perspectives 25, 3–16. https://doi.org/10.1257/jep.25.3.3

Liu, X., Mills, A., 2007. Agency Theory and Its Applications in Health Care, in: Public Ends, Private Means: Strategic Purchasing of Health Services. The World Bank, pp. 151–178.

Lovdata, 2021a. Forskrift om pasient- og brukerrettigheter i fastlegeordningen - - Lovdata [WWW Document]. URL https://lovdata.no/dokument/SF/forskrift/2012-08-29-843/§2#§2 (accessed 1.4.23).

Lovdata, 2021b. Forskrift om fastlegeordning i kommunene - Kapittel 6. Fastlegens liste - Lovdata [WWW Document]. URL https://lovdata.no/dokument/SF/forskrift/2012-08-29-842/KAPITTEL_6#§35 (accessed 1.4.23).

Loveman, E., Frampton, G.K., Clegg, A.J., 2008. The clinical effectiveness of diabetes education models for Type 2 diabetes: A systematic review. Health technology assessment 12, 1–136.

Ludvigsson, J.F., Håberg, S.E., Knudsen, G.P., Lafolie, P., Zoega, H., Sarkkola, C., von Kraemer, S., Weiderpass, E., Nørgaard, M., 2015. Ethical aspects of registry-based research in the Nordic countries. Clinical epidemiology 491–508.

Lunn, P.D., Choisdealbha, Á.N., 2018. The case for laboratory experiments in behavioural public policy. Behavioural Public Policy 2, 22–40. https://doi.org/10.1017/BPP.2016.6

Magelssen, M., 2018. Gjeninnfør reservasjonsadgang for fastleger. Tidsskrift for Den norske legeforening 138. https://doi.org/10.4045/TIDSSKR.18.0821

Marchand, C., Peckham, S., 2017. Addressing the crisis of GP recruitment and retention: a systematic review. Br itish Journal of General Practice 67, e227–e237. https://doi.org/https://doi.org/10.3399/bjgp17X689929

Mark, T.L., Swait, J., 2004. Using stated preference and revealed preference modeling to evaluate prescribing decisions. Health Economics 13, 563–573. https://doi.org/10.1002/HEC.845

McGuire, T.G., 2000. Physician agency, in: Handbook of Health Economics. Elsevier, pp. 461–536. https://doi.org/https://doi.org/10.1016/S1574-0064(00)80168-7

Mo, A.-K., Kongsrud, H.T., 2021. Leger jukser til seg millioner, men få blir straffet. NRK.

Mooney, G., Ryan, M., 1993. Agency in health care: getting beyond first principles. Journal of health economics 12, 125–135. https://doi.org/https://doi.org/10.1016/0167-6296(93)90023-8

Morken, T., Rebnord, I.K., Maartmann-Moe, K., Hunskaar, S., 2019. Workload in Norwegian general practice 2018–an observational study. BMC health services research 19, 434. https://doi.org/https://doi.org/10.1186/s12913-019-4283-

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. Department of Health, Education, and Welfare, National Commission for the˜….

Newhouse, J.P., 1987. Health Economics and Econometrics. The American Economic Review 77, 269–274.

Newhouse, J.P., Rand Corporation, 1993. Free for All?: Lessons from the Rand Health Insurance Experiment. Harvard University Press.

Nilsen, T.M., 2022. Enighet om Normaltariffen. Legeforeningen.

Nordberg, E.M.K., Skirbekk, H., Magelssen, M., 2014. Conscientious objection to referrals for abortion: pragmatic solution or threat to women's rights? BMC medical ethics 15. https://doi.org/10.1186/1472-6939-15-15

Norwegian Directorate of Health, 2023a. KUHR-databasen [WWW Document]. URL https://www.helsedirektoratet.no/tema/statistikk-registre-og-rapporter/helsedata-og-helseregistre/kuhr (accessed 5.3.23).

Norwegian Directorate of Health, 2023b. FLO (fastlegeordningen) [WWW Document]. URL https://www.helsedirektoratet.no/tema/finansiering/andre-finansieringsordninger/flo-fastlegedata (accessed 5.3.23).

Norwegian Directorate of Health, 2020. Årsrapport 2019 [Annual report 2019].

Norwegian Directorate of Health, 2019. Kompetansekrav for leger i kommunene - Helsedirektoratet [WWW Document]. URL https://www.helsedirektoratet.no/nyheter/krav-til-spesialisering-for-leger-i-kommunene (accessed 3.17.22).

Norwegian government proposition, 1995. Innstilling fra sosialkomiteen om samarbeid og styring - mål og virkemidler for en bedre helsetjeneste.

Norwegian Medical Association, 2019a. Fastlegesituasjonen forverret, viser ny kartlegging [WWW Document]. URL https://www.legeforeningen.no/nyheter/2019/fastlegesituasjonen-forverret-viser-ny-kartlegging/

Norwegian Medical Association, 2019b. Normaltariff for fastleger og legevakt 2019-2020 Den norske legeforening [WWW Document]. URL https://normaltariffen.legeforeningen.no/asset/pdf/Fastlegetariffen-2019-2020.pdf (accessed 4.7.22).

Norwegian Parliment, 2020. Innstilling fra helse- og omsorgskomiteen om Endringer i folketrygdloven (sammenslåing av egenandelstak 1 og 2).

Odebiyi, C.;, Walker, B.;, Gibson, B.;, Sutton, J.;, Spooner, M.;, Checkland, S., 2022. Eleventh National GP Worklife Survey 2021 Item Type Article.

OECD, 2021. State of Health in the EU: Norway.

OECD, 2020a. Medical graduates [WWW Document]. URL https://data.oecd.org/healthres/medical-graduates.htm

OECD, 2020b. Health Care Resources [WWW Document]. URL https://stats.oecd.org/Index.aspx?DataSetCode=HEALTH_REAC

OECD, 2019. Recent Trends in International Migration of Doctors, Nurses and Medical Students. OECD Publishing, Paris.

Ohn Oncato, J.C., Irav Hah, N.S., Alph H Orwitz, R.I., 2000. Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. New England Journal of Medicine 342, 1887–1892. https://doi.org/10.1056/NEJM200006223422507

Olsen, K.R., Anell, A., Häkkinen, U., Iversen, T., Ólafsdóttir, T., Sutton, M., 2016. General practice in the Nordic countries. Nordic Journal of Health Economics 4, 56–67. https://doi.org/https://doi.org/10.5617/njhe.2801

Owen, K., Hopkins, T., Shortland, T., Dale, J., 2019. GP retention in the UK: a worsening crisis. Findings from a cross-sectional survey. BMJ Open 9, e026048. https://doi.org/10.1136/BMJOPEN-2018-026048

PAHO, 2023. Primary Health Care [WWW Document]. URL https://www.paho.org/en/topics/primary-health-care (accessed 5.3.23).

Pedersen, K., Godager, G., Rognlien, H.D., Tyrihjell, J.B., Værnø, S.G., Iversen, T., Holte, J.H., Abelsen, B., Pahle, A., Augestad, L.A., others, 2022. Evaluering av handlingsplan for allmennlegetjenesten 2020--2024: Evalueringsrapport I.

Pedersen, L.B., Kjær, T., Kragstrup, J., Gyrd-Hansen, D., 2012. General practitioners' preferences for the organisation of primary care: a discrete choice experiment. Health policy 106, 246–256. https://doi.org/https://doi.org/10.1016/j.healthpol.2012.03.006

Petersen, L.A., Woodard, L.C.D., Urech, T., Daw, C., Sookanan, S., 2006. Does pay-for-performance improve the quality of health care? Annals of Internal Medicine 145, 265–272. https://doi.org/10.7326/0003-4819-145-4-200608150-00006/SUPPL_FILE/PETERSEN_AT1_145-4-265-DC1-1.PDF

Pulleyblank, R., Skovsgaard, C.V., Olsen, K.R., 2020. PA3 New Primary Care Incentive Scheme for Type 2 Diabetes in Denmark. Value in Health 23, S408–S409. https://doi.org/10.1016/j.jval.2020.08.068

Randolph, W., Viswanath, K., 2004. Lessons learned from public health mass media campaigns: Marketing Health in a Crowded Media World. Annual. Review of Public Health 25, 419–456. https://doi.org/10.1146/annurev.publhealth.25.101802.123046

Roberts, M.J., Hsiao, W., Berman, P., Reich, M.R., 2008. Getting health reform right: a guide to improving performance and equity. New York. https://doi.org/https://doi.org/10.1093/acprof:oso/9780195371505.001.0001

Robinson, J.C., 2001. Theory and practice in the design of physician payment incentives. The Milbank Quarterly 79, 149–177. https://doi.org/https://doi.org/10.1111/1468-0009.00202

Samoocha, D., Bruinvels, D.J., Elbers, N.A., Anema, J.R., van der Beek, A.J., 2010. Effectiveness of web-based interventions on patient empowerment: A systematic review and meta-analysis. Journal of medical Internet research 12, e23. https://doi.org/https://doi.org/10.2196/jmir.1286

Sandelowsky, H., Hylander, I., Krakau, I., Modin, S., Ställberg, B., Nager, A., 2016. Time pressured deprioritization of COPD in primary care: a qualitative study. Scandinavian Journal of Primary Health Care 34, 55–65. https://doi.org/10.3109/02813432.2015.1132892

Sanvik, H., Hundskår, S., Blinkenberg, J., 2022. Statistics from out-of-hours primary health care 2021. Bergen.

Saunes, I.S., Karanikolos, M., Sagan, A., 2020. Health Systems in Transition Norway Health system review. Health system in transition. Norwegian Institute of Public Health 22.

Scott, A., Gravelle, H., Simoens, S., Bojke, C., Sibbald, B., 2006. Job Satisfaction and Quitting Intentions: A Structural Model of British General Practitioners. British Journal of Industrial Relations 44, 519–540. https://doi.org/10.1111/J.1467-8543.2006.00511.X

Scott, A., Hall, J., 1995. Evaluating the effects of GP remuneration: problems and prospects. Health policy 31, 183–195. https://doi.org/10.1016/0168-8510(94)00697-D

Scott, A., Schurer, S., Jensen, P.H., Sivey, P., 2009. The effect of an incentive program on quality of care in diabetes management. Health Economics 18, 1091–1108. https://doi.org/10.1002/HEC.1536

Shen, J., Andersen, R., Brook, R., Kominski, G., Albert, P.S., Wenger, N., 2004. The Effects of Payment Method on Clinical Decision-Making: Physician Responses to Clinical Scenarios. Care 42, 297–302. https://doi.org/10.1097/01.mlr.0000114918.50088.lc

Silverman, E., Skinner, J., 2004. Medicare upcoding and hospital ownership. Journal of Health Economics 23, 369–389. https://doi.org/10.1016/J.JHEALECO.2003.09.007

Simoens, S., Scott, A., Sibbald, B., 2002. Job satisfaction, work-related stress and intentions to quit of Scottish GPs. Scottish Medical Journal 47, 80–86. https://doi.org/10.1177/003693300204700403

Sivey, P., Scott, A., Witt, J., Joyce, C., Humphreys, J., 2012. Junior doctors' preferences for specialty choice. Journal of Health Economics 31, 813–823. https://doi.org/https://doi.org/10.1016/j.jhealeco.2012.07.001

Sloan, F.A., Hsieh, C.-R., 2012. Health Economics. MIT Press.

Sørensen, R.J., Grytten, J., 2003. Service production and contract choice in primary physician services. Health Policy 66, 73–93. https://doi.org/10.1016/S0168-8510(03)00026-5

Sosial- og helsedepartementet, 1997. St.meld. nr. 23 (1996-97), 030005-040006. regjeringen.no.

SOU, 2019. Digifysiskt vårdval: Tillgänglig primärvård baserad på behov och kontinuitet.

Statistics Norway, 2022. Fastlegers inntekter og kostnader 2020.

The National Board of Health and Welfare, 2020. Statistikdatabas för hälso- och sjukvårdspersonal [WWW Document]. URL https://sdb.socialstyrelsen.se/if_per/val.aspx

The Norwegian Directorate of Health, 2022a. Oppfølging, utredning og organisering av diabetesomsorgen - Helsedirektoratet [WWW Document]. URL https://www.helsedirektoratet.no/retningslinjer/diabetes/oppfolging-utredning-og-organisering-av-diabetesomsorgen#arskontroll-og-bruk-av-noklus-diabetesskjema (accessed 8.31.22).

The Norwegian Directorate of Health, 2022b. Stopptober og andre tobakkskampanjer. - Helsedirektoratet [WWW Document]. URL https://www.helsedirektoratet.no/tema/tobakk-royk-og-snus/kampanjer-og-informasjonsmateriell (accessed 3.2.23).

The Norwegian Directorate of Health, 2022c. Finansieringsordninger for stabilitet og rekruttering - Helsedirektoratet [WWW Document]. URL https://www.helsedirektoratet.no/rapporter/handlingsplan-for-allmennlegetjenesten-arsrapport-2021/en-attraktiv-og-trygg-karrierevei-malomrade-1/finansieringsordninger-for-stabilitet-og-

rekruttering (accessed 3.27.23).

The Norwegian Directorate of Health, 2010. PSA - Prostataspesifikt Antigen [WWW Document]. URL https://nhi.no/imagevault/publishedmedia/figfrr6pwz41bn3owdxu/11182-2-psa-prostataspesifikt-antigen.pdf (accessed 3.31.23).

The Norwegian Directorate of Health, 2009. Vedlegg 4a til § 5-4 - Kollektivavtalen - Helsedirektoratet [WWW Document]. URL https://www.helsedirektoratet.no/rundskriv/kapittel-5-stonad-ved-helsetjenester/vedlegg-4a-til--5-4-kollektivavtalen (accessed 3.2.23).

Trønderopprøret, 2020. Fastlegeordningen 2.0 [WWW Document]. URL https://www.flo20.no/

Trønderopprøret, 2018. Trønderopprørets fastlegeundersøkelse og helsemedarbeiderundersøkelse.

Van Ham, I., Verhoeven, A.H., Groenier, K.H., Groothoff, J.W., De Haan, J., 2009. Job satisfaction among general practitioners: A systematic literature review. The European Journal of General Practice 12, 174–180. https://doi.org/10.1080/13814780600994376

Vårdanalys, 2017. En primär angelägenhet.

Vårdanalys 2018:5, 2018. Allmän tillgång?

Von Hinke, S., Davey Smith, G., Lawlor, D.A., Propper, C., Windmeijer, F., 2016. Genetic markers as instrumental variables. Journal of Health Economics 45, 131–148. https://doi.org/10.1016/J.JHEALECO.2015.10.007

Weber, R.A., Camerer, C.F., 2006. “Behavioral experiments” in economics. Experimental Economics 9, 187–192. https://doi.org/10.1007/s10683-006-9121-5

Wilson, B.J., 2014. The meaning of deceive in experimental economic science, in: The Oxford Handbook of Professional Economic Ethics. Oxford University Press New York, NY.

World Health Organization, 1978. Declaration of Alma Ata.

Wynn, R., 2020. E-ealth in Norway Before and During the Initial Phase of the Covid-19 Pandemic. The Importance of Health Informatics in Public Health during a Pandemic 272.

Young, R., Leese, B., 1999. Recruitment and retention of general practitioners in the UK: what are the problems and solutions? British Journal of General Practice 49, 829–833.

Papers

I

GP Recruitment and retention in the Nordic countries

OLE KRISTIAN AARS[1*]
ODDVAR KAARBØE[1,2]

[1] Department of Health Management and Health Economics, Faculty of Medicine, University of Oslo, Oslo, Norway
[2] Department of Global Public Health and Primary Care and Department of Economics, University of Bergen, Norway

Abstract: Demographic changes and decentralization of health care provision have led to a higher demand for General Practitioners' (GP) services. As a result, many countries, including the Nordics, report that recruiting and retaining GPs is increasingly difficult. Coupled with younger GPs increasingly valuing work/life balance, countries are ever more concerned about ensuring a sustainable GP supply going forward. In the Nordics, several policies have been implemented to this effect. The purpose of this article is to develop a theoretical framework for informing such policy choices. Our focus is on remuneration schemes, GPs' working conditions and practice quality as levers to incentivize effort and to attract GPs. We show that policies that have a positive effect on recruiting and retaining GPs can have a negative effect on the effort GPs exert. Since reduced effort might have a negative effect on the services patients receive, the total effects of the policies are uncertain. We further show that the dominating effect is sensitive to the extent that GPs are altruistic and care for patients' benefit of treatment, providing important insights for policy makers who want to increase GP supply.

JEL classification: C02, I11, J33, J20
Keywords: payment systems, general practitioners, recruitment, retention

1 Introduction

Demographic changes and decentralization of health care provision have led to a higher demand for General Practitioners (GP) services in many countries, including the Nordics. At the same time, with larger groups of older GPs about to reach retirement, and younger GPs having preferences for more leisure, a potential result is reduced supply (Sivey et al., 2012). This is partly driven by an increasing share of female physicians, who generally want to work less. The increased demand for their services, however, has increased working hours, exacerbating characteristics of the job they find undesirable (Dale et al., 2015). As a result, many GPs are considering leaving practice pre-retirement (Kuusio et al., 2013; Eneroth et al., 2017).

* Correspondence to: Ole Kristian Aars, Department of Health Management and Health Economics, Faculty of Medicine, University of Oslo, Oslo, Norway. Email: olekaar@medisin.uio.no.
Published: Online December 2022. dx.doi.org/10.5617/njhe.8560

Many countries report that recruiting and retaining GPs is increasingly difficult[2] and reforms and policies have been implemented to tackle the challenges. Some of these have been related to how GPs are remunerated; some have focused on the well-being of the GPs or their working conditions and some have been investments in public goods that improve GPs service delivery. An overview with examples of reforms is provided in the Appendix.

We investigate the effects of these policies in a theoretical model for provision of GP health services depicted by two choices; i) whether to work as a GP and ii) how much effort to exert into patient care. We then show how these two trade-offs are influenced by policy-instruments related to remuneration, working conditions and practice quality[3]. That is, we focus on policies relevant for those who already have a relevant education and can work as a GP. We further show how these trade-offs are sensitive to a person's level of altruism, i.e., the caring of patients' benefit of treatment (when working as a GP).

One important insight from our model is that policies that have a positive effect on recruiting GPs can have a negative effect on the effort GPs exert, making the total effects of the policies uncertain. We further show that the dominating effect is sensitive to preferences and characteristics of the GPs, providing important insights for policy makers who want to increase GP supply.

1.1 The Brekke and Nyborg-model

Our model builds on Brekke and Nyborg (2010) (hereinafter "BN"). BN presents a model of work motivation where individuals have heterogeneous preferences for a self-image of being important to others. Specifically, they investigate how an individual's self-image interacts with the way the individual is paid to determine her choice of occupation and effort. If pay is determined by individual productivity, the individuals (called bakers) produce a private good and exert an effort level that correspond to the standard homo economicus model. That is, the preference for being important to others has no behavioral consequence for those who on the margin are compensated by their marginal social value. However, receiving a fixed wage (financed through taxes) individuals (called nurses) exert higher effort in the production of health care than the standard economic model predicts. The mechanism behind is that nurses are motivated by doing good for others, at increasing rates of higher self-image.

Turning to the choice of occupation, BN predicts that individuals with intermediate preferences for being important will seek employment as bakers. Individuals with the highest and lowest work motivation become nurses, whose effort cannot be verified and whose remuneration is a fixed wage. I.e., the nursing sector also attracts poorly motivated workers that shirk. The intuition is that a job switch to the nursing sector might raise or lower net social welfare. It depends on the balance between the positive welfare effect related to increased production in the nursing sector, and the negative welfare effect of increased taxes to finance the new nurse's wage. A baker with low self-image does not care much if her job switch lowers others' welfare, while bakers with intermediate self-image feel bad about moving to the nursing sector and by exerting relatively low effort. These workers choose to continue as bakers. Finally, those with strong social preferences would exert a high effort level after the job switch so that the total welfare effect is positive. Those people would change their occupation and become nurses.

---

[2] See e.g., Denmark, (Danish Ministry of Health, 2017); Finland, (Keskimäki et al., 2019); Iceland, (Haraldsdóttir, 2010), Norway, (KS, 2019); Sweden, (Vårdanalys, 2018:5)

[3] We have chosen these instruments on the basis of the identified reforms in Table 1, as well as categories discussed in the literature (i.e., Lafortune, 2016)

An interesting policy implication is that wages in the nursing sector should be kept low to keep shirkers out. However, this will lead to a too small nursing sector (relative to first best). BN proposes another policy to attract highly motivated nurses: Investing in capital equipment (a public good) that increases nurses' efficiency in helping others. Examples of capital equipment are better diagnostic equipment in hospitals and library resources in universities and schools.

1.2 Adapting the BN-model

We argue that alterations should be made to make the BN-model more relevant for GPs. Below, we outline our modelling choices, highlight differences compared to BN, explain why we believe our model is a better fit to the GP sector, and show how our results differ.

One alteration is related to the remuneration system. While nurses, and many other health care workers, receive a fixed wage, GPs are typically remunerated with a combination of fee for service (FFS), capitation, and fixed salary. GPs are thus able to affect their own income by exerting effort, either directly if paid by FFS or indirectly if paid by capitation.

The motivation for a patient to seek treatment from a GP is different from that of a nurse. Specifically, where the patient typically cannot choose her nurse, this is to a large extent the case for a GP. Aboulghate et al. (2012) has shown that most patients prefer to see a particular GP, as motivated by continuity of care and overall satisfaction of the service that is provided. The act of seeking care is driven by the expected treatment from that individual GP, suggesting that increased effort can increase the likelihood of a patient seeking treatment.

An implication of patients' active choice of GPs is that GPs also increase their likelihood of providing good care by improvements in the quality of their practice, as this would be observable for potential patients. Nurses however, are mostly employed at hospitals, elderly care, or community health centers – places where care is sought on the basis of that institution serving a catchment area, or that a third-party gatekeeper (usually a GP) has referred one there. We argue that there is an important distinction to be made with regards to practice quality (which is observable to the patient) and working conditions (which is not).

Lastly, we believe that the inclusion of an altruism parameter better reflects the choices GPs make than self-image. In this sense we are closer to Arrow (1963) who coined the importance of the physician's other-regarding motive to care for a patient when describing physician behavior. We argue that the notion of self-image is not unique to the healthcare profession, but rather a subjective measure of how a person values their contribution to society. Akerlof and Kranton (2005) argues that "identity" – as their use for the term self-image – reflects the level which a person identifies with attributes of his/her work, irrespective of societal impact. Similarly, Tirole (2002) discusses self-image in the context of self-esteem, suggesting that the notion relates to one owns value-judgement. Altruism is widely understood as feeling good about doing good for *others*, and in the context of healthcare, the weight a physician attaches to a patient's health benefit. Altruism is a more precise representation of a GPs motivation, both since medicine is a profession of "doing good for others", but also since the definition provides a direct linkage between the act of providing care to a patient and feeling good about it. We believe that having a dedicated and continuous responsibility for patients on your list, suggests that a GP cares more about the wellbeing of her patients rather than her net contribution to social welfare. This is also supported by Allan et al. (2007) showing that GPs have little awareness of the economic costs of the care they provide, suggesting that any marginal decreases in others utility (e.g., from increased taxes), should not be taken into account. Rachlin and Jones (2008) have shown that altruism is a decreasing function of social distance – i.e., that you care more about people that are close than distant to you.

This suggest that a GP has a higher degree of altruism for the patient in her office, vs. the one on her list. We have operationalized this by assuming that the GP only cares for the patients she sees. This modelling choice is in line with recent papers that consider GP altruism in the cases where the GP faces a list system, see e.g., (Godager et al., 2015; Brekke et al., 2017; 2020).

Our modelling choices have implications for how our results relate to BN. While BN finds that the nursing sector consists of individuals with the lowest and highest work motivation, we find that only individuals with a certain level of work motivation will become GPs. We believe that this property of our model is capturing an important element of the GP sector, as GPs are typically not being thought of as having low motivation for their patients. Second, our model raises questions about the proposition of overinvesting in capital equipment (practice quality). We find that this policy might be counterproductive, i.e., raising the quality of the practice may actually make public sector work less attractive. This result holds even in the case where better practice quality raises the (marginal) effect of effort on attracting patients. Finally, by using the most common remunerations that GPs receive, we identify how a change in the FFS component has heterogeneous effects on attracting GPs due to the indeterminate effect a change in the FFS has on the GPs net income. This differs from BN where all nurses are "penalized" equally from a tax increase because their wage is fixed.

2 A model for GPs and consultants

We consider an economy with $i = 1,2, \dots, I$ individuals. There are two types of jobs in the economy. Either individual i works as a consultant producing a consumption good b or she is a GP funded by the public and producing health. We assume that there are no transaction costs of switching professions, implying that all individuals are qualified to work as GPs. Hence, a consultant is a person who is qualified to work as a GP but has chosen another career and is paid according to the value of her services in her new profession. Examples of such careers include working as health bureaucrats, doctors in private practice, academics, or hospital doctors. The main difference is that consultants are paid according to their marginal productivities[4]; GPs are not. They are paid by the government. All governmental expenses are financed through taxation[5].

Let $d = 0,1,2, \dots, I$ be the number of GPs. There is a list system in the public sector. Everyone is signed up with a doctor, and each doctor has I/d patients on her list[6]. Every individual i has preferences of the following type

$$U_i = u(b_i) - c(e_i, \beta) + \alpha_i \pi(e_i, \theta) \frac{I}{d} \tag{1}$$

where $u(.)$ is a strictly concave and increasing function of i's consumption of good $b_i \geq 0$, $e_i \geq 0$ is effort, and $\beta \geq 0$ is a parameter measuring a GP's working conditions, that are unobservable to the patient. An increase in the parameter reflects more unfavorable working conditions. Examples include the level of administrative and other non-patient related tasks and workplace flexibility. A GPs cost of effort is $c(e_i, \beta)$, where $c(0, \beta) = 0$. $c_e > 0$, $c_{ee} > 0$, $c_\beta > 0$, $c_{\beta\beta} > 0$ denote the first-order and the second-order derivatives respectively. We assume that $c_{\beta e} > 0$, meaning that the cost of effort increases when working conditions deteriorate. This parameter is analogous to working conditions in BN, but is included as part of the cost function to reflect that any improvements in

[4] We discuss this assumption in the discussion section.
[5] To keep the model simple we will assume that consultants are paid by private funds. I.e., the government does not raise taxes to finance their salary.
[6] We discuss this assumption in the discussion section.

working conditions increases utility irrespective of the GP treating any patients. Since consultants are paid according to their marginal product, working conditions are normalized to 0, giving us the cost function $c(e, 0)$.

We assume that (some) individuals are altruistic and care about patients' benefit of treatment when working as a GP. Specifically let $\alpha_i \in [0,1)$ measure the individuals' altruism. An individual with α=0 does not have altruistic preferences, but all individuals with $\alpha > 0$ care about their patients'. We further assume that the GPs can treat the patients that visit them, i.e., improve their health. With this assumption it follows that a GP's altruistic component is increasing in the number of patients who visit her. Furthermore, we assume that a GP can affect the likelihood of patients visiting them by her choice of effort.

Let $\pi(e_i, \theta) \in [0, 1)$ measure the likelihood of patients visiting a GP with effort e and quality of the practice θ. The quality of the practice is related to investments by the government and is a public good that improves treatment and/or patient experience of a visit. Let $\pi_\theta > 0$, $\pi_e > 0$, $\pi_{\theta e} \geq 0$, $\pi_{\theta\theta} \leq 0$ and $\pi_{ee} \leq 0$. Hence, higher efforts and better practice quality attract more patients, but (possible) at decreasing rates. If $\pi_{\theta e} > 0$ then effort and practice quality are complements; better practice quality increases the marginal effect of effort on attracting patients. We also assume that effort is non-verifiable[7], but both effort and the quality of a practice are observable (for the patients).

If an individual works as a consultant, she produces a private good b, in a perfectly competitive market. Working as a GP, she is paid by the public through a combination of a fixed salary $\varphi \geq 0$, is a fixed capitation $\delta \geq 0$ per individual on the GPs list and a FFS component $\gamma \geq 0$). The budget constraints are given by

$$b_i = w - t, \quad \text{where} \quad \begin{array}{ll} w = \varphi + \frac{I}{d}(\delta + \pi(e_i, \theta)\gamma) & \text{for GPs}^8 \\ w = f(e) & \text{consultants} \end{array}$$

where $f(e_i)$ is the market value of the consultant's production (b is the numeraire), $f_e > 0$, $f_{ee} < 0$ and $f(0) = 0$. Remuneration and practice quality are financed through a lump-sum tax t. All individuals pay the same tax, i.e., $t = (\overline{w}d + K(\theta))/I$, where $\overline{w} = \varphi + (I/d)\,(\delta + \bar{\pi}(e)\gamma)$ is the average remuneration of the GPs, $\overline{\pi}(e) = \sum_i \pi(e_i)/d$ is the average share of patients visiting a doctor and $K(\theta)$, with $K'(\theta) > 0$, is the cost of providing GPs with a practice quality. The lump-sum tax is thus given by $t = (\varphi d + K(\theta))/I + \delta + \bar{\pi}(e)\gamma$. It then follows that a GP's consumption of the private good, b_i, is given by her disposal income:

$$b_i = w_i - t = \frac{\varphi(I-d) - K(\theta)}{I} + \frac{(I-d)\delta}{d} + \frac{\gamma\left((I-1)(\pi(e_i)) - \sum_{-i}\pi(e_{-i})\right)}{d} \tag{2}$$

An individual makes two choices: i) For a given type of employment, individual i maximizes her utility with respect to effort $e_i \geq 0$ and ii) chooses whether she produces the private good b or works as a GP. The choice is based on the maximal utility obtained in the two jobs. For both decisions the individual makes, we assume she takes everybody else's choices as given. Hence, our equilibrium concept is a Nash equilibrium, see e.g., Gibbons (1992).

[7] If effort was verifiable it could be contracted upon and remunerated according to the GP's marginal product.
[8] We simplify notation and write $\pi(e_i)$ instead of $\pi(e_i, \theta)$.

2.1 The effort decision

We assume that the optimal effort of a consultant producing the private good is sufficiently high to cover the lump-sum tax, that is $f(e_i^{P*}) > (\varphi d + K(\theta))/I + \delta + \bar{\pi}(e)\gamma$; where e_i^{P*} denotes i's optimal effort choice. Since she does not treat any patients, her optimal effort choice is the solution to the following maximization problem: $U_i^P = \max\{u(f(e_i) - t) - c(e_i)\}$. The first order condition is:[9] $u_b f_e - c_e = 0$.

We now consider a GP's effort decision. From (1) and (2), we see that a GP's maximal utility is given by

$$U_i^D = \max_{e_i}\left\{u\left(\frac{\varphi(I-d)-K(\theta)}{I} + \frac{(I-d)\delta}{d} + \frac{\gamma\left((I-1)\left(\pi(e_i)\right) - \sum_{-i}\pi(e_{-i})\right)}{d}\right) - c(e_i,\beta) + \alpha_i\frac{I}{d}\pi(e_i,\theta)\right\} \tag{3}$$

Differentiation with respect to e_i, yields the following first-order condition:

$$u_b(b_i)\frac{\gamma(I-1)}{d}\pi_e + \frac{\alpha_i I}{d}\pi_e - c_e = 0 \tag{4}$$

The first part captures the monetary gain of exerting effort. It depends on the marginal utility of the private good, and the increased FFS, net of the increased tax payment. The second part is capturing the altruistic benefit of exerting effort when working as a GP. The third part is the cost of effort. The second-order condition for a maximum is given by

$$u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)^2 + u_b(b_i)\frac{\gamma(I-1)}{d}\pi_{ee} + \frac{\alpha_i I}{d}\pi_{ee} - c_{ee} < 0, \tag{5}$$

which is satisfied since $u_{bb} < 0, \pi_{ee} \leq 0$ and $c_{ee} > 0$.

Assuming an internal solution, i.e., $e_i^{D*} > 0$, the first-order condition defines a GP's optimal effort e_i^{D*} as a function of the parameters, i.e., $e_i^{D*} = e_i^{D*}(\gamma, \delta, \varphi, \theta, \beta, \alpha_i)$. By differentiating the first-order condition with respect to e_i^{D*}, α_i, θ and β we get the following comparative static results, (the denominator is negative due to the second-order condition):

$$\frac{de_i^{D*}}{d\alpha_i} = \frac{-\frac{I}{d}\pi_e}{u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)^2 + u_b(b_i)\frac{\gamma(I-1)}{d}\pi_{ee} + \frac{\alpha_i I}{d}\pi_{ee} - c_{ee}} > 0 \tag{6}$$

$$\frac{de_i^{D*}}{d\theta} = -\frac{u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)\left(\frac{\gamma(I-1)}{d}\pi_\theta - \frac{K'(\theta)}{I}\right) + \left(u_b(b_i)\frac{\gamma(I-1)}{d} + \frac{\alpha_i I}{d}\right)\pi_{e\theta}}{u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)^2 + u_b(b_i)\frac{\gamma(I-1)}{d}\pi_{ee} + \frac{\alpha_i I}{d}\pi_{ee} - c_{ee}} \gtreqless 0 \tag{7}$$

[9] The second-order condition for a maximum is satisfied since $u_{bb} < 0, f_{ee} < 0$ *and* $c_{ee} > 0$.

$$\frac{de_i^{D*}}{d\beta} = \frac{c_{e\beta}}{u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)^2 + u_b(b_i)\frac{\gamma(I-1)}{d}\pi_{ee} + \frac{\alpha_i I}{d}\pi_{ee} - c_{ee}} < 0 \tag{8}$$

First, from (6), more altruistic GPs exert higher effort because of the non-monetary reward of treating patients. Second, from (7), the effect of better practice quality on a GP's optimal effort is indeterminate. It depends on the curvature of the utility function ($u_{bb} < 0$) and on the relationship between effort and practice quality ($\pi_{e\theta}$). Suppose $\pi_{e\theta} = 0$; i.e., a change in the practice quality does not affect how a change in effort affect patients' choice of seeing their GP or not, and that the marginal cost of changing the practice quality is small, i.e., $K'(\theta)$ small, then the optimal effort is *decreasing* in practice quality. The mechanism behind this result is that when the practice quality is improved, more patients would like to visit their GP for a given level of the GP's effort. Since effort is costly, the GP respond by reducing her effort (the substitution effect). At the same time, the extra tax burden on the GP is small so that the negative income effect of higher taxes does not dominate the substitution effect. On the other hand, if the marginal utility of the private good is approximately constant ($u_{bb} \approx 0$), the complementarity between effort and practice quality ensures that optimal effort is increasing with better practice quality. Third, from (8), the effect of better working conditions on effort is positive since better working conditions (a reduction in β) reduces the cost of effort.

Regarding the effects of changes in the remuneration parameters $\varphi,\ \gamma,$ and δ we obtain:

$$\frac{de_i^{D*}}{d\varphi} = \frac{-u_{bb}(b_i)\frac{(I-1)(I-d)}{dI}\gamma\pi_e}{u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)^2 + u_b(b_i)\frac{\gamma(I-1)}{d}\pi_{ee} + \frac{\alpha_i I}{d}\pi_{ee} - c_{ee}} \leq 0 \tag{9}$$

$$\frac{de_i^{D*}}{d\delta} = \frac{-u_{bb}(b_i)\frac{(I-1)}{d}\frac{(I-d)}{d}\gamma\pi_e}{u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)^2 + u_b(b_i)\frac{\gamma(I-1)}{d}\pi_{ee} + \frac{\alpha_i I}{d}\pi_{ee} - c_{ee}} \leq 0 \tag{10}$$

$$\frac{de_i^{D*}}{d\gamma} = -\frac{\frac{(I-1)}{d}\pi_e\left[u_{bb}(b_i)\gamma\left(\frac{I}{d}\pi(e_i^{*D}) - \bar{\pi}(e)\right) + u_b(b_i)\right]}{u_{bb}(b_i)\left(\frac{\gamma(I-1)}{d}\pi_e\right)^2 + u_b(b_i)\frac{\gamma(I-1)}{d}\pi_{ee} + \frac{\alpha_i I}{d}\pi_{ee} - c_{ee}} \gtreqless 0 \tag{11}$$

From (9) and (10) we see that a higher salary or an increased capitation reduce a GP's optimal effort level. This is because the valuation of the extra income earned through the FFS component is reduced[10]. From (11), we see that the effect of increased FFS is indeterminate. To understand why, notice that a higher FFS increases income which is a positive effect. However, since the marginal utility of the private good is decreasing ($u_{bb}(b_i) < 0$) and the cost of effort is strictly increasing in effort, a GP might respond by reducing her effort (substitution effect). This is the case for example when the effect on the marginal utility is strong ($u_{bb}(b_i)$ "large"). On the other hand, if the effect on the marginal utility is weak ($u_{bb}(b_i)$ "small"), the GP will respond to increases in FFS by raising her effort. The effect also depends on how much effort the GP exerts relative to the

[10] Because overall income has increased and $u_{bb}(b_i) < 0$.

average effort level of GPs, i.e., on $\pi(e_i^{*D}) - \bar{\pi}(e)$. GPs that exert a high (low) level of effort relatively to others are ceteris paribus more likely to reduce (increase) their effort level. An implication is that one should expect heterogeneous responses in effort following a change in the FFS. Finally, the size of the FFS affects the effort response. Specifically, the higher the FFS the more likely it is that the effort response is negative if the GP's effort was more than average to start with.

2.2 The employment decision

When considering which job to take an individual is comparing the maximal utility she can get. Let $\Delta U(\alpha_i) = U_i^D - U_i^P$ denote the value function of the individual's maximization problem, i.e., the difference in the maximal utilities for individual *i* being a GP or a consultant. From this, we get:

$$\Delta U(\alpha_i) = u\left(\varphi + \frac{I}{d}(\delta + \pi(e_i^{*D})\gamma) - t\right) - u(f(e_i^{P*}) - t) - c(e_i^{D*}) - c(e_i^{P*}) + \alpha_i \frac{I}{d}\pi(e_i^{D*}) \tag{12}$$

From the envelope theorem it follows that

$$\frac{\partial \Delta U(\alpha_i)}{\partial \alpha_i} = \frac{I}{d}\pi(e_i^{D*}) \geq 0, \text{ with strict inequality for } e_i^{D*} > 0. \tag{13}$$

That is, the difference in the maximal utilities is weakly increasing in α_i. Notice that the difference in maximal utilities depends on the number of GPs. Specifically, when the number of GPs increases the differences in the maximal utilities decreases, but the expression is always non-negative.[11] Taken together, these observations imply that if an individual with altruism $\tilde{\alpha}$ choose to become a doctor, then all individuals *i* with $\alpha_i \in [\tilde{\alpha}, 1)$ will also choose to become GPs.
The value function $\Delta U(\alpha_i)$ depends on the tax level, which again depends on the number of GPs. The (Nash) equilibrium in the economy is thus a tax level and a corresponding altruism $\in [0,1)$ such that all individuals with altruism $\alpha_i \in [0, \alpha^*)$ become private consultants and all individuals with $\alpha_i \in (\alpha^*, 1)$ are GPs.

We now derive conditions to ensure that both types of employment are active in equilibrium, and that at least the most altruistic GP exert a positive effort level.[12]

Consider first the case where no individuals work as GPs. In this scenario, each consultant's production just covers her own consumption. I.e., $b_i = f(e_i^{P*})$.

Suppose a random consultant is considering becoming a GP, and that she chooses an effort level such that the cost of effort is the same irrespectively of whether she works as a GP or not, i.e., $c(e, \beta) = c(e_i^{P*}, 0)$. Since she is the only GP, the tax level is $\varphi/I + \delta + \pi(e_i^{P*})\gamma$, and she receives wages $w = I(\varphi + \delta + \pi(e_i^{*P})\gamma)$. Hence, a sufficient condition for the consultant to switch employment is $w - t = (I-1)(\varphi/I + \delta + \pi(e_i^{P*})\gamma) > f(e_i^{P*})$. That is, her net salary is higher than the value of her production.

[11] The intuition behind this result is that while consultants' utility decreases due to an increased tax burden, a GP's disposal income decreases both because additional GPs raise the tax burden, and because the GP's remuneration decreases (if the GP is not paid only a fixed salary). If the GP is remunerated with a fixed salary, the result goes through since more GPs decreases a GP's altruistic utility.

[12] Since the cost of practice quality is independent of the number of GPs, we ignore this cost when deriving the conditions.

Only in the non-generic case will it be optimal for the GP to exert an effort level that results in the same effort cost as when she produces the consumption good b. In the general case, she chooses either zero effort, or the effort level that solves the GP's first-order condition (given in (4)). The next equation gives the condition for the GP to exert zero effort.

$$u((I-1)(\varphi+\delta+\pi(0)\gamma)-u((I-1)(\varphi+\delta+\pi(e_i^{*D})\gamma)-c(0)+c(e_i^{*D})+\alpha_i I(\pi(0)-\pi(e_i^{*D})>0 \qquad (14)$$

Suppose there is no FFS ($\gamma = 0$). Then the two first parts cancel each other out, and we are left with $c(e_i^{*D}) + \alpha_i I(\pi(0) - \pi(e_i^{*D}) > 0$. Obviously, this is positive for small α. Hence, individuals with little concerns for patients' benefit, who considers becoming a GP in a system with no FFS, will shirk (no effort).

We do however believe that (at least) the most altruistic GP finds it optimal to provide a positive effort level independent of the remuneration system she faces. That is, she will choose a positive effort level also when the remuneration system does not contain an FFS. We thus impose the condition that the benefits related to altruism outweighs the cost of effort for the most altruistic GP, i.e., the GP with $\alpha \cong 1$. $c(e_i^{*D}) \leq \left(\pi(e_i^{*D}) - \pi(0)\right) I$.

Consider now the case where everybody works as GPs. In this case, I=d; and a GP's net income is given by $w_i - t_i = \gamma\left(\pi(e_i) - \bar{\pi}(e)\right)$. Obviously, all GPs that see less patients than the average, will receive a negative disposal income, i.e., $\pi(e_i) - \bar{\pi}(e) < 0$. Since a GP's effort is (weakly) increasing in α, the GP with the lowest altruism exerts the least effort. Hence, this cannot be an equilibrium for this GP with no altruism since we have assumed that a private consultant's optimal effort is sufficiently high to cover the lump-sum tax, i.e., $f(e_i^{*P}) > \varphi + \delta + \gamma\bar{\pi}(e)$.

The following proposition sums up our results.

Proposition 1.

Suppose $(I-1)\left(\varphi/I + \delta + \gamma\pi(e_i^{*P})\right) > f(e_i^{*P}) > \varphi + \delta + \gamma\bar{\pi}(e)$, *and* $c(e_i^{*D}) \leq \left(\pi(e_i^{*D}) - \pi(0)\right) I$. *Then*

1. *Both consultants and GPs are active in the corresponding Nash equilibrium.*
2. *Since* $\partial\Delta U(\alpha_i)/\partial\alpha_i = (I/d)\,\pi(e_i^{*D}) \geq 0$, *and the difference in maximal utilities is decreasing in the number of GPs, the Nash equilibrium is unique, and there exist an* $\hat{\alpha} \in [0,1)$ *such that* $\Delta U(\alpha_i) < 0\ (>0)$ *for* $\alpha_i \in [0, \hat{\alpha})\ (\hat{\alpha}, 1)$.
3. *The most altruistic GP exerts a positive effort level.*

We now investigate how the proposed governmental policies affect the attractiveness of being a GP. The criterion of becoming a GP is that the value function of individual's maximization problem, i.e., the difference in the maximal utilities (12) is positive. Here, $e_i^{D*} = e_i^{D*}(\gamma, \delta, \varphi, \theta, \beta, \alpha_i)$ is the utility maximizing effort for a GP with altruism α_i implicitly defined for a given set of policy parameters by the first-order condition (4). Differentiation of $\Delta U(\alpha_i)$ (equation (12)) with respect to the policy parameters, where $\partial e_i^{*D}/\partial\vartheta$, $\vartheta = (\gamma, \delta, \varphi, \theta, \beta)$ is implicitly given by (7)-(11), gives the following proposition, assuming an interior solution.

Proposition 2.

1. $\partial\Delta U/\partial\beta = -C_\beta(e_i^{*D},\beta) < 0$ *and* $\partial\Delta U/\partial\theta = u_b(\gamma((I-1)/d)\,\pi_\theta(e_i^{*D},\theta) - K'(\theta)/I) + (\alpha_i I/d)\,\pi_\theta(e_i^{*D},\theta)$ is indeterminate. *Becoming a GP is strictly more attractive when the working conditions* (β) *improves, but the effect of the quality of the practice* (θ) *is indeterminate.*
2. $\partial\Delta U/\partial\varphi = u_b(\cdot)((I-d)/I) > 0, \partial\Delta U/\partial\delta = u_b(\cdot)((I-1)/I) > 0$ *and* $\partial\Delta U/\partial\gamma = u_b((I/d)\,\pi(e_i) - \pi(\bar{e}))$ is indeterminate. *An increase in the GP's salary or the capitation raises the attractiveness of becoming a GP, while an increase in the FFS has an indeterminate effect.*

Proof:
We only provide the proof of the last statement of ii). The other statements can be proved similarly. By differentiating $\Delta U(\alpha_i)$ wrt. γ we obtain:

$$\frac{\partial\Delta U(\alpha_i)}{\partial\gamma} = u_b\left(\frac{I}{d}\pi(e_i) - \bar{\pi}(e)\right) + \left(u_b(b_i)\frac{\gamma(I-1)}{d}\pi_e + \frac{\alpha_i I}{d}\pi_e - c_e\right)\frac{\partial e_i^{*D}}{\partial\gamma}$$

Notice that the last parenthesis is zero as it corresponds to the first-order condition. Hence, $\partial\Delta U(\alpha_i)/\partial\gamma = u_b((I/d)\,\pi(e_i) - \bar{\pi}(e)) > 0\ (<0)$ depending on the sign of the first parenthesis.

From Proposition 2 it follows that most of the policy parameters have the expected effect, i.e., a more generous remuneration and improvements in the working conditions make it more attractive to work as a GP. Moreover, the positive relationship between becoming a GP and altruism ensures that increases in remuneration attracts the most motivated individuals to enter the profession first. However, when it comes to the FFS, the effect can still be indeterminate as it depends on how high effort a GP exerts relative to the average effort level of GPs. If a GP exerts a low level of effort relative to the other GPs, few patients visit her. If this were the case, an increase in the FFS might result in a *decrease* in the GP's disposal income as the (negative) effect of the tax increase outweighs the increase in gross income. Because FFS is contingent on effort in our model, this secures that new recruitments provide high levels of effort. Interestingly, this is the opposite of what the model predicted for choice of effort; individuals who continue working as GPs are induced to work harder – if not, they lose money.

Similarly, the effect of improved practice quality is indeterminate. It depends on changes in a GPs disposable income, their valuation of it, in addition to any altruistic gains of attracting more patients. Specifically, we see that when the GP's remuneration does not contain FFS elements, then only highly altruistic individuals may find it more attractive to become GPs when the practice quality is improved. The non-altruistic GPs will be worse off since their net income is reduced through higher taxes, resulting in exit from the profession.

3 Discussion

Many countries have challenges with sustaining an adequate level of GP service provision and have implemented several policies to that end. The predictions of our model show that while all policies discussed in this paper can have positive effects, ill-defined objectives can have unintended consequences for the sustainability of GP schemes.

Our findings suggest that only the most motivated individuals choose to become GPs; they are encouraged by the altruistic benefit from treating patients. This also induces them to provide

positive effort, which is reinforced if FFS is (part of) the remuneration scheme. Not surprisingly, we see that improved working conditions increases utility, and thereby recruitment. This means that reduction in administration (as seen in Norway), better coordination of services (as seen in Sweden) or removal of undesirable responsibilities through task shifting (as seen in Finland and Sweden) most likely will have positive effects on attracting new GPs. This would also generally be the case for policies that increase remuneration.

Depending on the preferences of the individual, remuneration schemes can have an effect beyond income. For example, the uncertain nature of activity-based remuneration schemes can be a negative attribute for some – especially younger GPs (Abelsen and Olsen, 2012). This suggests that income guarantees positively affect working conditions. On the other hand, a heavy focus on extrinsic rewards (like FFS) can have the opposite effect among some people, crowding out intrinsic motivation (Deci, 1971). However, there are also studies showing that activity-based contracts can be desirable from an equity perspective, in that people feel like being treated fairly (Clark and Oswald, 1996). These conflicting outcomes are reflected in the form of the ambiguous effect of FFS in our model. One explanation might be heterogeneous preferences depending on the stage of your career; Holte et al. (2015) shows that older GPs have a higher preference for FFS compared to younger and female GPs, suggesting that the income effect dominates.

An important reason for using FFS is that it induces effort for those who are not sufficiently motivated. While we do see this prediction in our model for GPs with below-than-average effort – which could be the target group of the policy maker, we also see that the same group are deterred from becoming GPs altogether. The intuition behind this finding is that if you find yourself in a job that you are not very motivated to do, you will either quit or need strong inducements to work hard. This may also be why the average age of GPs quitting their job has decreased (Norwegian Ministry of Health and Care Services, 2020); assuming that switching costs increases with age (Hyatt and Spletzer, 2016), an older GP will respond by working harder, while a younger GP will choose to quit the profession altogether.

Like with FFS, differing preferences for leisure makes the effect of practice quality on effort ambiguous. For low-income earners, we would expect that the income effect dominates, leading them to capitalize on its complementarity on effort. Interestingly, this income effect is only present when the remuneration scheme includes an FFS component where effort is rewarded, suggesting that such GPs would be early adopters of practice quality measures (i.e. technology) that attracts patient demand (Jha et al., 2008).

The above findings have important policy implications. First, attracting altruistically motivated GPs is often less costly than inducing effort amongst those who are not. As shown by Brekke and Nyborg (2010) an intrinsically motivated person needs less external inducement for a given level of effort. Second, one should try to identify those who appreciate existing working conditions before changing them. Third, for negligible tax increases of an investment in practice quality, we would see a positive effect on recruitment; either because it helps new entrants attract patients, or because it allows them to reduce their burden of work. Fourth, while both utility and effort are monotonically increasing with improved working conditions, changes to remuneration schemes have either ambiguous or opposing effects. Since this ambiguity often results from heterogeneous preferences, one could perceivably solve the problem of heterogeneity in preferences by offering different contracts to different GPs.

Our model has provided a theoretical framework for understanding the effects of policies on increasing supply of GP services, but there are some limitations. We have assumed that all individuals are qualified to work as doctors. Hence, reforms targeting medical education are not

explicitly included. While this is a limitation, this could be analyzed through working conditions by making students more positively predisposed to the attributes of the profession that are otherwise seen as inferior by others (i.e., rurality) (Aaraas et al., 2015). Moreover, the assumption of no transaction costs of entering general practice is a simplification given that some countries require a specialization. While this would require extra effort from the doctor, GPs would still be exposed to the same payment mechanisms. Additionally, the specialization requirement can be waived for locums who have not set up their own practice (Norwegian Directorate of Health, 2017).

Another criticism relates to that the attributes of alternative employment for the non-GPs may defy assumptions of our model with regards to remuneration and altruism. For remuneration, we would argue that the mixed payment system for GPs is unique compared to other professions and the associated effects hold independent of altruism. Moreover, a GP considering switching jobs will only do so if she believes the associated wage adequately reflects her expected marginal productivity in the new job, and that overtime work is paid according to expected marginal productivity. Equally, an employer will set the wage according to his/her expectation of the worker's productivity. This means that former GPs will be paid according to the value of their services in their new professions as e.g., health bureaucrats, doctors in private practice, academics, or hospital doctors. Similar arguments hold for candidates at the end of the medical university education in her choice say between becoming a hospital doctor (paid by a fixed wage) or becoming a GP.

An alternative employment may involve patient care and therefore involve an altruistic benefit, which is not explicitly included in our model. However, studies have shown that people who choose general practice have higher rates of altruism than for other specialties (Deci, 1971, Borges and Savickas, 2002; Mullola et al., 2018), lending support for differences in utility increasing in altruism. Moreover, a doctor may find it easier to capitalize or altruistic preferences as a GP because she can affect patient demand. Our altruistic parameter could thus be interpreted as the extra utility GPs get additional to alternative employment involving patient care, e.g. as a hospital doctor. Lastly, our model assumptions imply that GPs have an equal number of patients on their list. While this is a simplification, we believe that a non-GP wanting to enter general practice would expect a list equal to the average length. Moreover, the sustained GP deficit and prolonged working hours for the active workforce, suggests that new GPs would cater to unmet patient demand rather than compete for patients.

Acknowledgement
We thank three reviewers and the editor for valuable comments on earlier drafts. We thank the Research Council of Norway for funding the project Modernizing the GP Scheme (project number 288592).

4 References

Aaraas, I.J., Halvorsen, P.A., and Aasland, O.G. (2015). Supply of doctors to a rural region: Occupations of Tromsø medical graduates 1979–2012. *Medical Teacher*, 37 (12), 1078–1082.

Abelsen, B. and Olsen, J.A. (2012). Does an activity based remuneration system attract young doctors to general practice? *BMC Health Services Research*, 12 (1), 1–8.

Aboulghate, A., Abel, G., Elliott, M.N., Parker, R.A., Campbell, J., Lyratzopoulos, G., and Roland, M. (2012). Do English patients want continuity of care, and do they receive it? *British Journal of General Practice*, 62 (601), 567–575.

Akerlof, G.A. and Kranton, R.E. (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, 19 (1), 9–32.

Allan, G.M., Lexchin, J., and Wiebe, N. (2007). Physician awareness of drug cost: a systematic review. *PLOS Medicine*, 4 (9), e283.

Borges, N.J. and Savickas, M.L. (2002). Personality and medical specialty choice: A literature review and integration. *Journal of Career Assessment*, 10 (3), 362–380.

Brekke, K.A. and Nyborg, K. (2010). Selfish bakers, caring nurses? A model of work motivation. *Journal of Economic Behavior & Organization*, 75 (3), 377–394.

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

Brekke, K.R., Holmås, T.H., Monstad, K., and Straume, O.R. (2020). How does the type of remuneration affect physician behavior? *American Journal of Health Economics*, 6 (1), 104–138.

Clark, A.E. and Oswald, A.J. (1996). Satisfaction and comparison income. *Journal of Public Economics*, 61 (3), 359–381.

Danish Ministry of Health (2017). Lægedekning i hele Danmark.

Deci, E.L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18 (1), 105.

Eneroth, M., Gustafsson Sendén, M., Schenck Gustafsson, K., Wall, M., and Fridner, A. (2017). Threats or violence from patients was associated with turnover intention among foreign-born GPs–a comparison of four workplace factors associated with attitudes of wanting to quit one's job as a GP. *Scandinavian Journal of Primary Health Care*, 35 (2), 208–213.

Finnish Ministry of Social Affairs and Health (2016). *National programme for training medical and dental specialists 2017–2019*. Helsinki: Finnish Ministry of Social Affairs and Health.

Gibbons, R. (1992). *A primer in game theory*. Prentice Hall.

Godager, G., Iversen, T., and Ma, C.A. (2015). Competition, gatekeeping, and health care access. *Journal of Health Economics*, 39, 159–170.

Government proposition 2019:164 (2019). Inriktningen för en nära och tillgänglig vård – en primärvårdsreform.

Haraldsdóttir, H.B.S.& S. (2010). Skortur á heimilislæknum. *Fréttabréf Landlæknis um Heilbrigðistölfræði* .

Holte, J.H., Abelsen, B., Halvorsen, P.A., and Olsen, J.A. (2015). General practitioners' altered preferences for private practice vs. salaried positions: a consequence of proposed policy regulations? *BMC Health Services Research*, 15 (1), 1–10.

Hyatt, H.R. and Spletzer, J.R. (2016). The shifting job tenure distribution. *Labour Economics*, 41, 363–377.

Icelandic Directory of Health (2016). Heilsuvera: mínar heilbrigðisupplýsingar. Available from: https://www.landlaeknir.is/gaedi-og-eftirlit/heilbrigdisthjonusta/rafraen-sjukraskra/heilsuvera-minar-heilbrigdisupplysingar/ (Accessed: 2 June 2020).

Jha, A.K., Doolan, D., Grandt, D., Scott, T., and Bates, D.W. (2008). The use of health information technology in seven nations. *International Journal of Medical Informatics*, 77 (12), 848–854.

Keskimäki, I., Tynkkynen, L.-K., Reissell, E., Koivusalo, M., Syrjä, V., Vuorenkoski, L., Rechel, B., Karanikolos, M., and Organization, W.H. (2019). Finland: Health system review. *Health System Review*, 21(2), 1–166.

KS (2019). *Kommunesektorens arbeidsgivermonitor 2019*. The Norwegian Association of Local and Regional Authorities.

Kuusio, H., Heponiemi, T., Vänskä, J., Aalto, A.-M., Ruskoaho, J., and Elovainio, M. (2013).

Psychosocial stress factors and intention to leave job: differences between foreign-born and Finnish-born general practitioners. *Scandinavian Journal of Public Health*, 41 (4), 405–411.

Lafortune, G. (2016). Analytical framework of health labour markets. *In*: OECD, ed. *Health Workforce Policies in OECD Countries: Right Jobs, Right Skills, Right Places*. Paris: OECD Publishing.

Mullola, S., Hakulinen, C., Presseau, J., Gimeno Ruiz de Porras, D., Jokela, M., Hintsa, T., and Elovainio, M. (2018). Personality traits and career choices among physicians in Finland: employment sector, clinical patient contact, specialty and change of specialty. *BMC Medical Education 2018 18:1*, 18 (1), 1–12.

National Audit Office (2017). *Heilsugæsla höfuðborgarsvæðisins*. Reykjavík.

Norwegian Directorate of Health (2017). Kompetansekrav for leger i kommunene. Available from: https://www.helsedirektoratet.no/nyheter/krav-til-spesialisering-for-leger-i-kommunene (Accessed: 17 April 2022).

Norwegian Ministry of Health and Care Services (2020). *Handlingsplan for allmennlegetjenesten*. Oslo.

OECD (2016). *Health Workforce Policies in OECD Countries*. OECD Health Policy Studies. Paris: OECD Publishing.

Rachlin, H. and Jones, B.A. (2008). Social discounting and delay discounting. *Journal of Behavioral Decision Making*, 21 (1), 29–43.

Sigurgeirsdóttir Sigurbjörg Maresso Anna, W.J. (2014). Iceland: Health system review. *Health Systems in Transition*, 16(6): 1-182.

Sivey, P., Scott, A., Witt, J., Joyce, C., and Humphreys, J. (2012). Junior doctors' preferences for specialty choice. *Journal of Health Economics*, 31 (6), 813–823.

SOU 2018:39 (2018). *God och nära vård*. Stockholm: Norstedts Juridiks.

The Danish Ministry of Health (2018). *En læge tæt på dig: En plan for fremtidens almen praksis*.

The National Board of Health and Welfare (2018). *Bättre resursutnyttjande i hälso- och sjukvården*. Stockholm: The National Board of Health and Welfare,.

Tirole, J. (2002). Rational irrationality: Some economics of self-management. *European Economic Review*, 46 (4–5), 633–655.

Vårdanalys 2018:5 (2018). Allmän tillgång?

Zanaboni, P. and Fagerlund, A.J. (2020). Patients' use and experiences with e-consultation and other digital health services with their general practitioner in Norway: results from an online survey. *BMJ Open*, 10 (6), e034773.

5. Appendix

Table 1: Overview of selected reforms and policy measures, 2010-2020

Type of policy measure	Denmark	Finland	Iceland	Norway	Sweden
Remuneration	Differentiated capitation fees (i.e. severity-based)[a] DKK 1500 **compensation** per patient for deficit (compared with target list length)[c]	**FFS and capitation introduced** in addition to basic salary for subset of GPs[g]	**Incentive payments** as a bonus-supplement for extra effort[h]	Proposed policies on **income guarantees** up to a 500 patient list[f] Differentiated capitation, with lower capitation when list size > 1000[f] Expansion of salaried positions for new graduates (ALIS)[f]	Locums are generally offered higher salaries[g]
Working conditions	Earmarked funding to incentivise establishment of **group practices**[a] Older GPs have been **relieved** of the requirement to work in the emergency ward. They're also provided an **incentive payment** after turning 62 years to stay in the job[b]	Policies implemented to reduce demand for GP services through **task shifting**[b]	Post graduate studies for advanced nurses introduced, including to induce **task shifting**[h]	Removal **administrative tasks**, including through task shifting and use of IT[f]	The reform "Profesjonsmilliarden" seeks to induce **task shifting**[i] Better coordination and clarification of roles between primary and specialist care to reduce stress among GPs [j]
Practice quality	Proposed to renovate/ improve GP offices used for training[a] GP offices subsidized to to be accredited in accordance with the Danish Quality Model[a]	IT system Apotti launched in 2018, enabling information sharing and consultations between practitioners and patients in selected regions [d] Proposed to improve specialist training [k]	Electronic health systems put in place to enable electronic communication between patients and GPs and data-sharing across primary and specialist care[el]	Online platform established and e-consultation enabled for GP-practices[m] Specialist requirement to become GP[n]	Investments in effort to reduce **waiting time** for primary care. [o] Recommended that the government **subsidizes specialization** for GPs[g]

Source: a: (The Danish Ministry of Health, 2018) b: (OECD, 2016) c: (Lafortune, 2016) d: (Keskimäki et al., 2019) e: (Sigurgeirsdóttir Sigurbjörg Maresso Anna, 2014) f: (Norwegian Ministry of Health and Care Services, 2020) g:(Vårdanalys, 2018:5) h: (National Audit Office, 2017) i:(The National Board of Health and Welfare, 2018) j:(SOU, 2018:39) k: (Finnish Ministry of Social Affairs and Health, 2016) l: (Icelandic Directory of Health, 2016) m: (Zanaboni and Fagerlund, 2020) n: (Norwegian Directorate of Health, 2017) o: (Government proposition, 2019:164)

II

III

Sending emails to reduce medical costs?

The effect of feedback on general practitioners' claiming of fees

Ole Kristian Aars[a]*, Geir Godager[a, b], Oddvar Kaarboe[a, c], Tron Anders Moger[a]

Abstract:

Audit and feedback is used as a strategy to guide practices of health care professionals towards certain targets. The outcome of interest can be quality improvements, but also ensuring that health care workers adhere to relevant regulations. We conducted a nationwide field experiment in the Norwegian primary care sector to study the behavioral responses from giving general practitioners feedback (GPs) on their claiming of fees. The email-based feedback intervention targeted GPs who most frequently claimed fees for double consultations and provided them with a reminder of the formal regulations for double consultations. The intervention caused a 2-5 percentage point reduction in the use of the double-consultation fee, reducing the yearly health care spending of the Norwegian government by approximately €877 000 (or €1 270 per GP). This substantial and durable behavioral response found in our study sample comprising 15 % of Norwegian GPs, shows that low-cost interventions via email can have significant financial impact.

a: Institute of Health and Society, Department of Health Management and Health Economics, University of Oslo, Norway
b: Health Services Research Unit, Akershus University Hospital, Oslo, Norway
c: Department of Global Public Health and Primary Care and Department of Economics, University of Bergen, Norway
*Corresponding author: email: olekaar@uio.no
Declarations of interest: none.

1 Introduction,

Audit and feedback is used as a strategy to align professional practice with professional targets or standards. This is widely used for health care professionals, including general practitioners (GPs). The belief is that they are prompted to modify their practice when given performance feedback showing that their clinical practice is inconsistent with a desirable target (Hysong et al., 2006). The outcome of interest can be quality improvements and/or adherence to relevant regulations when prescribing. The purpose of this paper is to evaluate a field experiment on professional feedback, following a financial audit on Norwegian GPs.

GPs in Norway are remunerated with a combination of capitation (i.e. a fixed amount per patient) and fee-for-service (FFS). Beyond patient co-payments, GPs are reimbursed by HELFO – the government agency tasked with paying health professionals. One of the most used fees is a double-consultation fee (DCF), which can be claimed for consultations with a long duration (lasting more than 20 minutes). In 2019, the DCF was used more than five million times, amounting to €110M. This constituted more than 20% of the total fee reimbursement GPs received from the Norwegian health insurance scheme (HELFO, 2020). Based on a previous audit by HELFO (2021), it was hypothesized that DCF was claimed for more consultations than it was intended for.

In collaboration with HELFO, we conducted a nationwide field experiment in the Norwegian primary care sector to study the causal effects of giving GPs feedback on their claiming of DCF. The email-based feedback intervention targeted GPs who most frequently claimed fees for double consultations and provided them with a reminder of the formal regulations for double consultations. GPs in our study sample were randomly assigned to a control group (Control) or one of the two intervention groups Mild or Strong. GPs in the two intervention groups would receive different versions of a feedback email, whereas GPs in Control would not receive any feedback email. The feedback emails contained information about the total cost of DCFs reimbursed by HELFO in 2018 and stated that the GP used DCFs more frequently than the average GP. The information received by the two different intervention groups only differed in the

heading[2]. Mild received a feedback email with the heading "Information about DCF" and Strong had the heading "Regarding your use of DCF". The letter refers to moral aspects of a GP's practice style. However, given that HELFO can financially penalize GPs that overuse DCF, it may be that the letter is interpreted as a warning rather than a moral nudge.[3] We hypothesize that GPs in the two intervention groups reduce their use of DCF in response to the letters, and that the effect is more pronounced for the "Strong" group.

With HELFO distributing the email, we were able to use HELFO's established infrastructure to conduct the experiment, thus minimizing administrative costs and costs for patients and providers (Ivers et al., 2014), while enhancing the external validity of the study (Harrison et al., 2004). The field experiment enabled us to compare the behavioral responses to the different formulations used when messaging GPs, similar to Bott et al. (2019) who studied the effects of differently worded emails to reduce tax evasion. We also adjusted for seasonal effects and studied the impact of the Covid-19 pandemic on the estimated effects. Finally, and due to the possibility of internal communications among the GPs in a restricted Facebook group for GPs in Norway[4] and coverage of the intervention in the (national) media (Brandtzæg Clausen, 2019; Hafstad, 2019; Storvik, 2019), we analyzed the effect of the intervention on ***Control*** to quantify possible spillover effects. We quantified the causal effect of our field intervention on the monthly claims of the DCF relative to the regular consultation fee, i.e., the fee for a consultation lasting less than 20 minutes, in the 14-month period after the intervention.

The intervention had a statistically significant effect: Both the descriptive analysis of mean differences and the regression models showed a 2-5 percentage point drop in the use of DCF for both mild and strong intervention groups compared to the pre-treatment period. This relatively large short-term effect diminished over time. However, the effect remained statistically significant one year after the intervention. We also find that the observed difference between the

[2] See Appendix A1 for a translation of the emails

[3] Previous studies on tax compliance have shown that interventions referring to personal moral are less effective than those that include some likelihood of penalty on the recipient following from audit measures (Antinyan and Asatryan, 2019).

[4] According to Gronseth et al. (2020), a restricted Facebook group for GPs in Norway exists. As of spring 2018, the list included 3,357 members, of whom approximately 50 participated regularly in discussions.

Mild and ***Strong*** intervention groups is not statistically significant. Our interpretation is that receiving feedback has a larger impact than the specific wording used.

The intervention effects in this study are durable and have economic significance: The reduction in the use of DCF for the intervention groups added up to €877 279 per year or €1 270 per GP in the sample.

Previous studies on the effects of auditing and feedback on clinical practice report small to moderate effects (Eccles et al., 2001; Ivers et al., 2012; Jamtvedt et al., 2006). These studies have primarily been concerned with high intensity feedback – i.e., peer-to-peer, interviews, telephone calls, visits, educational components, and seminars. Likewise, the use of reminder messages to reduce radiology referrals has shown to be effective, but without reference to a GPs relative performance (Shojania et al., 2009). Moreover, these studies have primarily been concerned with – implicitly or explicitly – making accurate assessments of diagnoses. As we do not make any reference to diagnostic choices in our intervention, any changes in the use of DCF is not expected to be a result of changes in treatment decisions amongst GPs. Rather, a reduction in DCF would be generated from exerting more caution when coding consultations. The implication is that GPs either do not know how to use DCF properly, or that they knowingly use it for financial gain – often referred to as "upcoding". While the existence of gaming payment systems has been documented in the hospital sector (Dafny, 2005) and the GP setting (Gravelle et al., 2010), less is known about how interventions can alter such behavior.

We contribute to the literature on field experiments in health care by implementing two relatively minor email interventions to influence clinical practice and generate cost savings. Related studies focusing on antibiotics have shown that information letters (Meeker et al., 2016; Schwartz et al., 2021) and a mystery shopper scheme (Cheo et al., 2020) can reduce prescribing. Laboratory experiments have also been used for studying the behavioral responses caused by disclosing information about providers' performance, and by reminding providers about professional norms. Godager et al. (2016) found that compared to a regime with private information, a regime with performance disclosure was more likely to result in maximum benefits for patients. Experimental results reported by Kesternich et al. (2015) indicate that raising the saliency of

professional norms affect patient-regarding preferences and improve health outcomes. A common approach in the literature is to refer to a recommended clinical practice. In our study, the informational email only contained reference to a consultation fee without benchmarking it to standardized clinical care.

The paper proceeds as follows: Section 2 provides an overview of the Norwegian study setting. The randomization and interventions in the field experiment is presented in Section 3. Data and empirical methods are presented in Section 4. The results from nonparametric and parametric analysis are presented in Section 5, followed by a discussion of the findings, limitations, and implications for GP practice in Section 6.

2 Study setting

Norway has a National Health Service system financed through general taxation. Norwegian health care is organized into primary and secondary health care sectors. The former is the responsibility of municipalities while the latter is the responsibility of the central government. Since 2001, every Norwegian is listed with a GP, who also acts as a gatekeeper to access specialized care. In 2019, there were approximately 4,800 GPs, and only 0.2 % of the inhabitants had opted out of the system (Gaardsrud, 2020). Patients may switch GPs twice a year, and about 3 % of the patients do so annually. Most GPs (85 %) are self-employed and contract with a municipality. All fees and co-payments are set at the national level, without any geographical variations. The fee schedule specifies patient co-payments and fees reimbursed by HELFO - the government agency tasked with paying health professionals who contract with the National Health Service. The fee schedule includes a DCF, a fee that can be claimed in addition to a regular consultation fee (RCF) by GPs when consultations exceed a duration of 20 minutes and can be repeated per each started 15 min.[5] The RCF is higher for GPs who have qualified as specialists in general medicine. A typical GP consultation would result in claims for one RCF plus DCFs for long consultations.

[5] I.e. One DCF can be claimed for a 30 min consultation and two DCFs can be claimed for 35 min.

Table 1 GP fees for three consultation durations in 2019.

	Consultation duration in minutes		
	0 – 20	21 – 35	36 – 50
RCF [RCF for specialist in general med.]	€16 [€26]	€16 [€26]	€16 [€26]
DCF	0	1*€21	2*€21
Total claim	**€16 [€26]**	**€38 [€48]**	**€59 [€69]**

HELFO is also the financial auditor of GPs and may independently determine sanctions against health professionals that have failed to comply with the relevant regulations. HELFO performs financial audits of health professionals to ensure that the reimbursements to the health professionals are in line with the financial regulations of the public national insurance scheme. They also organize courses for GPs as part of their specialization. Following an audit, HELFO may independently determine sanctions against health professionals that have failed to comply with the relevant rules. The sanctions include instructions to adjust current practice, refund, loss of the right to practice at the expense of the Norwegian government, and reporting to the police. In 2020 (2019), HELFO reported 9 (3) cases to the police, 8 (10) health professionals lost the right to practice at the expense of the Norwegian government, and health care professionals had to return €5.6M (€2.8M) of payments received (Norwegian Directorate of Health, 2020). DCF has been a focus-area for HELFO in the lead up to our intervention, including information to GPs on fee regulations through newsletters and courses and by implementing automating rejection of excess use of DCFs.

3 The field experiment

3.1 Sampling and randomization

In total there are 4,800 GPs. The inclusion criteria for the field intervention sample were as follows:

- GPs had to claim reimbursement from HELFO during the first six months of 2019 (“be active”).
- GPs had to claim at least 500 RCF during this period.
- GPs had to rank among the top 700 GPs based on their frequency of DCFs relative to RCFs.

The frequency of DCFs relative to RCFs in our sample varied from 59 % to 173 % during the inclusion period. In comparison, the average relative frequency for all GPs was around 38 %. Nine GPs were dropped since they already were being audited by HELFO, resulting in a study sample of 691 GPs. This constitutes around 15% of all GPs. To avoid contamination between study arms, we made sure that GPs located at the same GP practice address were allocated to the same arm of the experiment[6]. This meant assigning a unique practice id-number to each GP, prior to performing the randomization. Randomization was performed by first randomly sorting the id-numbers, then sequentially assigning values 1, 2 or 3 to each practice.

After sampling, GPs were randomly assigned to one of three study arms by practice dummies: ***Control*** (no feedback) or either of two intervention groups ***Mild*** or ***Strong***. ***Mild*** and ***Strong*** received identical feedback emails except for the different headings. GPs in ***Strong*** received feedback email where the heading was "*Regarding your use of DCF*". In contrast, GPs in ***Mild*** received feedback emails where the heading was simply: "*Information about DCF*". Hence, we followed Bryan et al. (2013) and used active language to address the reader as "you" in the ***Strong*** arm and passive language in the ***Mild*** arm.

3.2 The feedback intervention

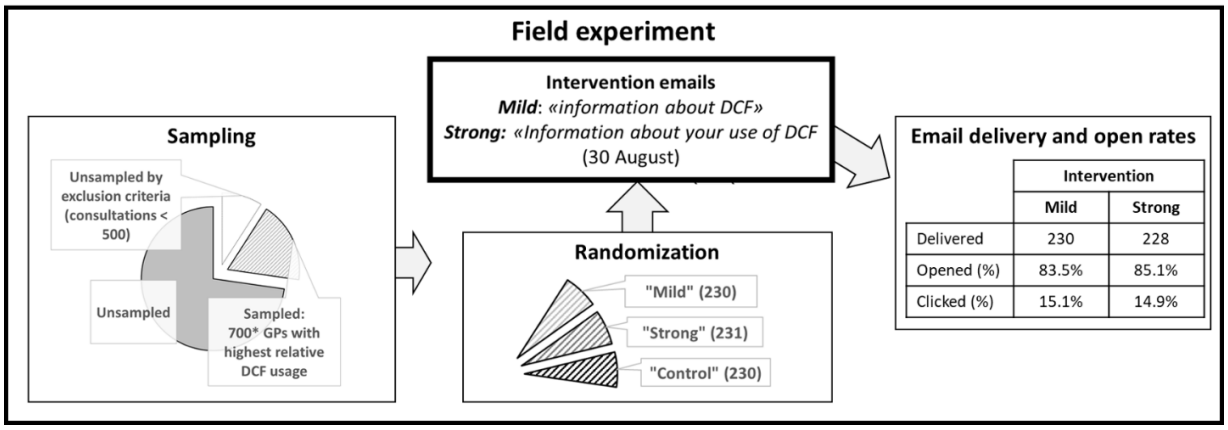
In the email body, the first paragraph stated the total amount (in NOK) that HELFO reimbursed for DCFs in 2019. The second paragraph gave details about the information campaign "Do you know" that was meant to increase the GP's awareness about claiming fees (HELFO, 2019). The third paragraph stated actively that "you" were receiving the email since statistics showed that "you" had used DCFs significantly more that the average GP. The fourth paragraph provided information about where the documentation for the claim in the former paragraph was taken from and where the GP could find more information about his or her claiming of fees. The fifth paragraph stated that the email was for information and guidance that the GP need not answer.

[6] Approximately 49 % of GPs in the sample belonged to the same GP-practice, of which 26 % were sharing practice with one other GP in the sample, while 23 % were in practices with 2-4 other sample GPs.

It also provided an email address, a phone number, and a reference number to be used if the GP would like to contact HELFO. The last paragraph provided a link to the "Do you know" campaign and a link to the HELFO newsletter.

An overview of the process for arriving at the control and intervention groups is provided in Flow Chart 1.

Flow Chart 1: Process for sampling, randomization, and intervention



4 Data and methods

4.1 Variable definitions

Data on the use of DCFs and RCFs per month from Jan 2017 to Nov 2020 were extracted for each GP in the three study arms. We could then compute the percentage of DCF relative to RCF henceforth denoted by ***%DCF***. ***%DCF*** is the main outcome variable in our empirical analyses. Using the indexes *ijt* to represent the fee *j* claimed by GP *i* in month *t*, our outcome variable ***%DCF*** is defined in Equation (1).

$$\%DCF_{it} \stackrel{\text{def}}{=} 100 * \frac{\sum_j DCF_{ijt}}{\sum_j RCF_{ijt}} \qquad (1)$$

In the descriptive analysis, within group differences in the ***%DCF*** claims over time periods after vs. before the intervention were examined by means of Wilcoxon signed rank tests. Overall differences in the ***%DCF*** claims across the intervention groups were examined by Kruskal-Wallis

tests. Tests were performed both as a simple comparison before vs. after the intervention and as before the intervention vs. three 5-month time periods after the intervention (0-4 months, 5-9 months, 10-14 months) to study whether effects were reduced in the short, medium and long run. Non-parametric tests were used due to skewness in the ***%DCF*** claims in the sample.

4.2 Model specification

We specify linear regression models with random GP specific effects. Results from a preliminary analysis of the effect before vs. after the intervention are presented in the text. To further examine how effects change over time, our specification let post intervention effects vary across the three time periods. We control for seasonal effects by using dummy variables for months and studied the impact of the Covid-19 pandemic on the estimated effects by adding a dummy variable equal to 1 from March 2020 onwards. Our model is specified as:

$$\%DCF_{it} = \beta_0 + \boldsymbol{\beta}_1 * \boldsymbol{Int}_i * \boldsymbol{Post}_t + \boldsymbol{\beta}_2 * \boldsymbol{Month}_t + \beta_3 * Covid_t + u_i + \varepsilon_{it}\,, \qquad (2)$$

where $\boldsymbol{Int_i}$ is a vector of dummy variables for the three arms. $\boldsymbol{Post_t}$ is a vector of dummy variables for the periods Sep 2019-Jan 2020 (0-4 months after intervention), Feb-Jun 2020 (5-9 months after) and Jul-Nov 2020 (10-14 months after). $\boldsymbol{Int_i} * \boldsymbol{Post_t}$ is the interaction between the two, giving rise to nine combinations, three terms per arm.[7] $\boldsymbol{Month}$ is a vector of eleven dummy variables Feb,…, Dec with January being the reference category. $Covid$ is a dummy variable taking the value 1 starting from March 2020. The latter two were included to assess the robustness of the intervention effect by taking into account changes in activity across seasons and following the pandemic. Finally, u_i is a GP specific random effect, while ε_{it} is a noise term.

Note that in our specification, we assume no difference in DCF use between the three arms prior to the intervention (observations score 0 on all dummies in $\boldsymbol{Int_i}$ and $\boldsymbol{Post_t}$ variables). This assumption is reasonable given that GPs are randomized to one of the three arms. The assumption was also supported by testing for difference using observations prior to Sep 2019 only (both by Kruskal-Wallis and by using a regression model with dummy variables for the study arms). Also note that the model will enable us to study effects in the control arm over time and

[7] These nine coefficients correspond to the nine first coefficients in **Table 3**.

hence the possible impact of information leaks via personal communication between GPs, in Facebook groups, and to the media: Effects in the control arm are represented by the three coefficients in β_1 that are assigned to ***Control***. Similarly, the three coefficients in β_1 that are assigned to ***Mild*** (and the three coefficients in β_1 that are assigned to ***Strong***) will reflect the effects of the ***Mild*** (***Strong***) intervention in each time period post intervention compared to pre intervention.

The distribution of ***%DCF*** is skewed in the sample. Using a gamma generalized linear mixed model with log link yielded similar effects and test results (Appendix 1), thus we opted to present the linear model. The small differences in results across model specifications is likely to be a result of the sample size (691 GPs contributing to 27,304 observations).

The choice between a fixed or random effects model will in general involve a trade-off between robustness (fixed effects) and efficiency (random effects). A random effect model provides efficient slope estimates when the random effects are uncorrelated with the regressors, and a random effect model is preferred when this assumption is met. Since all regressors are deterministic in the case at hand, the random effect assumptions are not restrictive. Our choice was also supported by a Hausman test (p-value of 0.13 for the model above), and there were negligible differences between the coefficients from the fixed and random effect models. In two sensitivity analyses, we first added to the model a random effect by practice id-number to capture dependence in doctors' behaviors within the same practice and variation between practices in ***%DCF*** claims. However, the variation between practices was close to zero (<0.01) when including the GP random effect. Second, GPs with low activity in regular consultation fees will either have lower activity in general or have shorter total follow-up time. Removing the 50 GPs with the lowest activity in regular consultation fees during the study period did not alter the results.

5 Results

5.1 Descriptive results

Table 2 present descriptive results. The arms are approximately balanced, considering the large standard deviations for the fees per month per GP. As is clear, the unadjusted effects seem stable through the follow-up time period. There is a small reduction in ***%DCF*** after the intervention in

the control arm; from 70.5 percent before the intervention, to 69.6 (0-4 months), 68.6 (5-9 months) and 69.5 (10-14 months) for the periods after the intervention. However, these reductions were not significant at the 5 % level compared to before the intervention according to Wilcoxon signed rank tests (p>0.2 for all). For the mild and strong intervention, however, the reductions in ***%DCF*** for all time periods after the intervention compared to before are statistically significant (p<0.01 for all). There are also significant differences between the control and intervention groups in ***%DCF*** for both the total time period 0-14 months vs. before (p=0.01, Table 2), and for each 5-month time period after intervention vs. Before.

Table 2 Descriptive statistics. Mean (SD) monthly claims for RCF and DCF and Mean (SD) of %DCF for the three arms.

Study arm	Variable	Before intervention:	After: 0-4 mths (p<0.001)*	5-9 mths (p=0.027)*	10-14 mths (p=0.025)*	0-14 mths (p=0.01)*
Control N=230	***%DCF***	70.5 (16.7)	69.6 (20.9)	68.6 (20.4)	69.5 (20.8)	68.8 (20.2)
	#DCF	136 (53)	142 (63)	131 (62)	139 (65)	137 (63)
	#RCF	193 (60)	206 (81)	194 (81)	207 (89)	200 (79)
Mild N=230	***%DCF***	70.2 (18.0)	65.9 (21.1)	66.7 (22.2)	67.6 (23.4)	66.2 (19.5)
	#DCF	144 (49)	148 (58)	142 (59)	148 (65)	226 (81)
	#RCF	209 (65)	230 (87)	222 (86)	227 (94)	145 (56)
Strong N=231	***%DCF***	70.3 (16.7)	65.8 (19.9)	66.3 (22.0)	66.8 (21.8)	66.2 (19.6)
	#DCF	147 (61)	152 (72)	141 (72)	147 (75)	225 (78)
	#RCF	211 (70)	234 (83)	213 (84)	220 (84)	148 (70)

*The p-values refer to Kruskal-Wallis tests on differences across intervention and control arms in the %DCF at 0-4 months after vs. before, 5-9 months after vs. before, and 10-14 months after vs. Before. §P-value refers to Kruskal-Wallis test on difference in the %DCF across study arms before the intervention.

Figure *1* presents the ***%DCF*** by arm and months. In September 2019 – the first month GPs would have been able to alter their claiming of fees – we see a distinct reduction in the use of DCFs for

all three study arms. The largest reduction in DCF claims is observed in the two intervention groups, Strong and Mild.

Figure 1: %DCF by study arm and month. N=691 GPs contributing to 27,304 observations in total

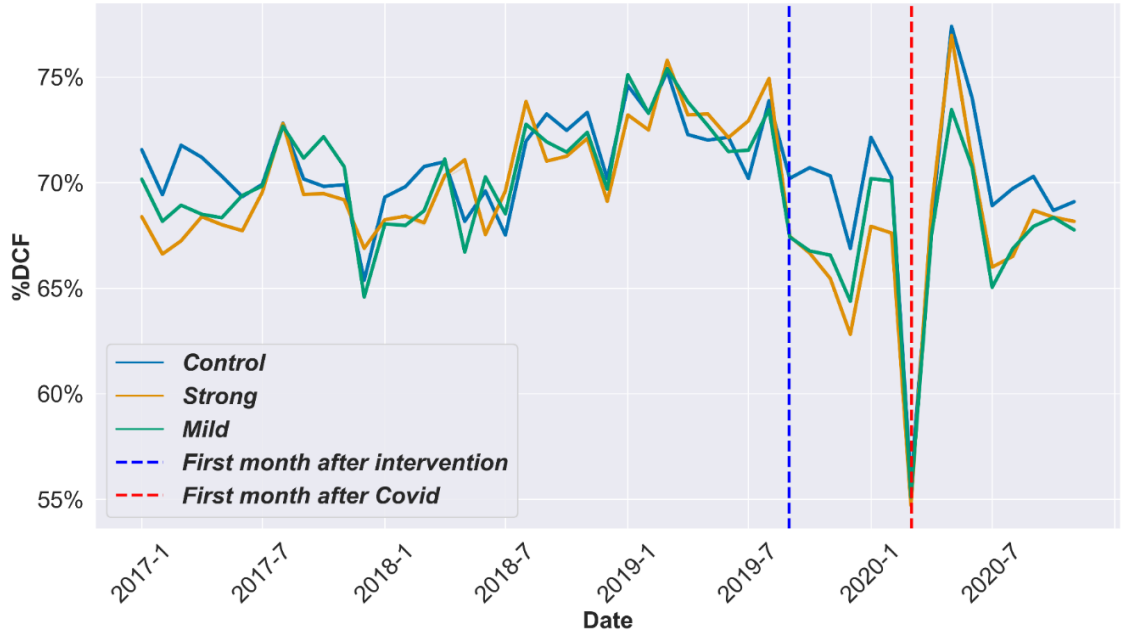
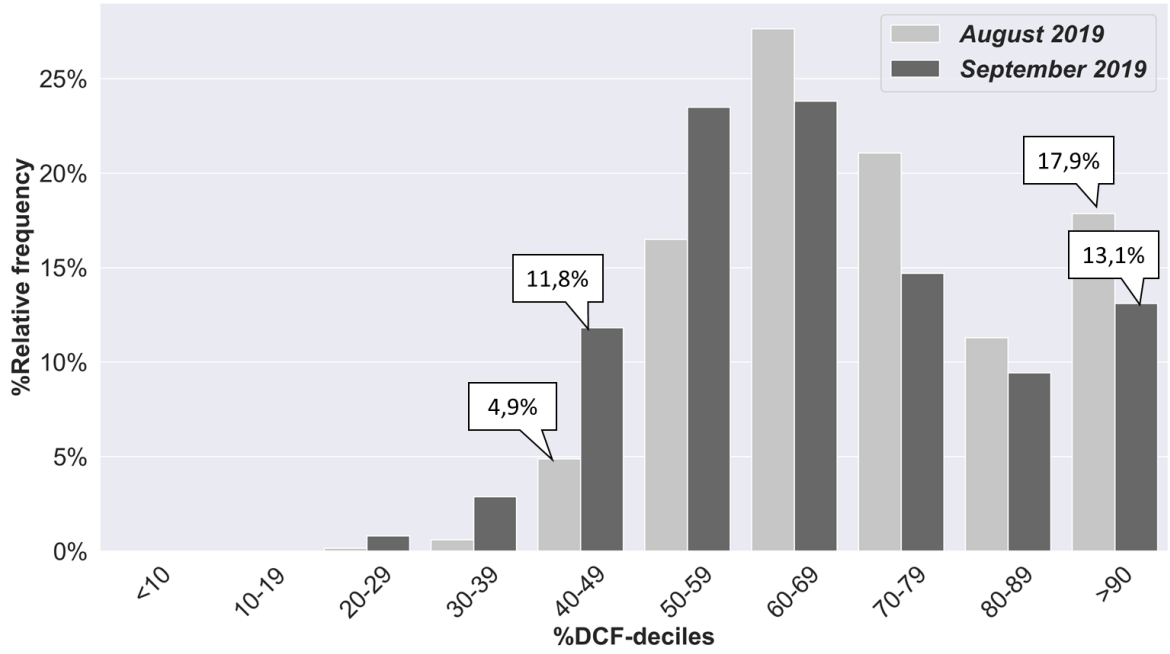


Figure 2 presents a histogram of relative frequencies of GPs in each of ten deciles for ***%DCF*** claims in August and September 2019. The intervention was implemented in the end of August 2019. When comparing the distribution in September with August, one can see clearly how the probability mass was shifted to the left. For example, we see that the relative frequency of GPs with ***%DCF >*** 90% was reduced from 17.9 % in August 2019 to 13.1 % in September 2019. We also see that the relative frequency of GPs with ***%DCF*** between 40 and 49 was more than doubled, from 4.9 % in August 2019 to 11.8 % in September the same year.

Figure 2: Histogram of relative frequencies in the months pre- and post- intervention. N=691 GPs contributing to 27,304 observations in total



5.2 Regression analysis

The preliminary regression analysis of the total time period after vs. before, adjusted for month and Covid-19 dummy variables, showed a non-significant reduction of 0.8 percentage points in ***%DCF*** for the control group (p=0.07), and reductions of 4.0 and 4.1 percentage points in the ***%DCF*** for the Mild and Strong intervention groups. The latter were both significantly different from the control group effect (p<0.001). **Table 3** presents the estimation results from two time period-dependent regression models; the unadjusted model showing the effects of each intervention group in the different periods and an adjusted model where dummy variables are added for months and the Covid-19 outbreak.

*Table 3: Estimation results from linear regression models. N=691 GPs contributing to 27,304 observations in total. Dependent variable %DCF=no. double-consultation fees/no. regular consultation fees*100*

	Unadjusted analysis		Adjusted analysis	
Variable:	**Coefficient (95% CI):**	**p-value**	**Coefficient (95% CI):**	**p-value**
Before intervention	Reference		Reference	
Control 0-4 months after	-1.05 (-2.08, -0.02)	0.05	-1.04 (-2.10, 0.03)	0.06
Control 5-9 months after	-1.14 (-2.19, -0.08)	0.03	0.58 (-1.08, 2.24)	0.49
Control 10-14 months after	-0.38 (-1.46, 0.71)	0.49	1.14* (-0.79, 3.08)	0.25
Mild 0-4 months after	-4.01 (-5.01, -3.00)	<0.001	-3.98 (-5.02, -2.93)	<0.001
Mild 5-9 months after	-3.97 (-4.99, -2.93)	<0.001	-2.25 (-3.89, -0.61)	0.01
Mild 10-14 months after	-4.34 (-5.39, -3.26)	<0.001	-2.82 (-4.73, -0.90)	<0.001
Strong 0-4 months after	-4.96 (-5.95, -3.96)	<0.001	-4.93 (-5.96, -3.88)	<0.001
Strong 5-9 months after	-3.79 (-4.83, -2.75)	<0.001	-2.06* (-3.71, -0.41)	0.01
Strong 10-14 months after	-3.86 (-4.93, -2.80)	<0.001	-2.34* (-4.26, -0.41)	0.02
Dummies for months			YES	
Before Covid-19			Reference	
After Covid-19			-1.96 (-3.54, -0.37)	0.02
Constant	70.52 (69.23, 71.80)	<0.001	71.50 (70.08, 72.91)	<0.001
R2	0.0047		0.0085	
Share variance due to GP random effect	56%		55%	

*Marks significant differences across time periods within the study arms at the 5% level.

There is a borderline significant effect of around one percentage point reduction in %DCF in the control arm 0-4 months after the intervention in both unadjusted and adjusted analyses. However, this seems to disappear after 5 months in the adjusted analyses.

There is a statistically significant drop in ***%DCF*** for both ***Mild*** and ***Strong*** intervention arms – and substantially higher than for the control group. While the ***Mild*** and ***Strong*** intervention groups also have an apparent reduction over time following intervention after adjusting for seasonal effects and Covid-19, the reduction is significant only for the ***Strong***. Significance follows from tests of the $\boldsymbol{Int_i * Post_t}$ interaction, using 0-4 months after the intervention as reference for each of the study arms. For ***Strong*** in the adjusted analyses, for example, the initial effect is a reduction of 4.93 percentage points in ***%DCF***. The size of the effect is reduced over time but remains significant; that is, a medium- and a long-run effect exists, respectively, of 2.06 and 2.34 percentage points, 5-9 months and 10-14 months after the intervention. Switching reference category for the $\boldsymbol{Int_i * Post_t}$ interaction will also show significant differences for each time period between controls after the intervention and for both the mild and strong intervention groups. There are no significant differences between the mild and strong interventions, however. The large variation in the outcome is apparent from the low R2-values. Due to collinearity, it is not feasible to further separate exogenous calendar time trends from the effects of the intervention.

6 Discussion and conclusion

We conducted a nationwide field experiment in the Norwegian primary care sector to study the behavioral responses of GPs receiving feedback on their claiming of fees. The interventions focused on the mode of treatment and whether the use of fees was aligned with regulatory and not clinical guidelines.

We found statistically significant effects of a simple feedback intervention via email. The effects are relatively large and long-lasting, and observable in both intervention groups through the 14-month follow-up period. We also find a short-run borderline significant effect in the control group, most probably caused by sharing of information in a closed Facebook groups and media coverage that followed the intervention. While our findings inform the debates around drawing causal conclusions from randomized control trials, the known challenges concerning the upscaling of experiment results into real world settings remain (Al-Ubaydli et al., 2021; Al-Ubaydli and List, 2015; Banerjee et al., 2017; Deaton, 2020; Harrison, 2021, 2011).

There are multiple channels through which the intervention could have caused behavioral responses for the ***Mild*** and ***Strong*** intervention groups. The intervention highlighted the existence of HELFO's auditing measures. While there was no reference to consequences for the GP, the mere existence of an audit may influence a GP's belief about the likelihood that the auditor has knowledge about his or her income-generating activities. This may affect the decision to claim the DCF. This information is given to both intervention groups, but not to the control group, and the GPs that learned about the audit through the media leak. However, the mild intervention group would only learn that HELFO possessed this knowledge if they opened the feedback email, as the passive language used in the heading did not reveal that the email contained individual information of the GPs use of the DCF. Having said so, the mild intervention group might have learned about the individual information given in the email through the media leak, as the media reported that this information was given in the email, and that the division director of HELFO confirmed that the receivers of the emails had used the DCF significantly more than the average GP (Storvik, 2019). Hence the shift in the claiming of DCFs can be understood as being a resulting of shift in beliefs.

Beyond any fear of financial retributions, a GP may also have preferences for assimilating to DCF usage of other GPs upon receiving information about their fee usage. If GPs preferences include a concern for their *self-image*, *honor* or *stigma* (Bénabou and Tirole, 2006) GPs may have changed fee usage as a response to a perceived change in the moral cost of deviating from the norm. Similar to studies on water consumption (Ferraro and Price, 2013), knowledge about using DCF significantly more than their peers may trigger a moral obligation to alter behavior. This mechanism may especially be at work in the ***Strong*** intervention group where the letter appealed to the economic consequences of overuse.

We conclude by noting that the effects of the field experiment constituted a significant reduction in reimbursement of the DCF, hence contributing to savings for Norwegian taxpayers. The reduction observed in ***%DCF*** among the mild and strong intervention groups in the study sample add up to €877 279 per year or €1 270 per GP. How big the savings would be if the field experiment were to be scaled up is difficult to estimate due to issues of external validity and scalable policies (Al-Ubaydli and List, 2015; Banerjee et al., 2017). One issue that might lower the

effect of a large-scale implementation is that honest GPs might feel they are unfairly treated due to the audit and respond by being less inclined to follow the financial regulations of the public National Insurance Scheme (Houser et al., 2012; Hu and Ben-Ner, 2020).

Lastly, without being able to determine the exact mechanism behind the reduction in use of DCF, we cannot say whether the estimated savings constitute a net welfare gain. On the one hand, it may be that the reduction in DCF observed in the intervention groups is a result being more cautious in coding consultations correctly; on the other hand, the reduction may be a reflection of GPs undertaking fewer long consultations – which assumes that GPs were using the DCF correctly in the first place. If the former mechanism holds, the estimated cost savings would be accrued from reduced up-coding amongst GPs. If the latter mechanism holds, the same savings would be offset by reduced patient welfare, resulting from shorter consultations, possibly at the expense of patients' health outcome. We cannot say which effect dominates, but we note that previous audits by HELFO (2021) have uncovered the existence of incorrect – or even fraudulent coding practices. Together with the wording in the intervention referring to "use of fees [being] correct [according to consultations undertaken]", and not to reducing a certain type of consultations, we'd be inclined to conclude that at least the mechanism of reducing incorrect coding is at work. Irrespectively, we view the magnitude of the effect as substantial and financially significant given the mild type of intervention.

Appendix

A: Emails

Box shows the emails that were sent to GPs. The letters to the ***Mild*** and ***Strong*** intervention groups only differed in the subject field, as shown in brackets.

Box 1: Email to GPs from HELFO

From: informasjon@helfo.no

To: <recipient>

Subject: [**Mild:** "Information about DCF"; **Strong:** "Regarding your use of DCF"]

Did you know that Helfo paid approximately 1.1 billion Norwegian kroners in reimbursements for the fee DCF in the year 2018?

Helfo has recently launched the campaign "did you know", aimed at increasing doctors' awareness around fee usage.

You are receiving this email because our analyses show that during the first half of 2019, you have used the fee DCF substantially more than the average. This does not necessarily mean that your use of the fee is incorrect, as there are various characteristics with your practice that may influence fee usage. As a doctor, you are responsible for reimbursements being correct, and in accordance with guidelines. We are asking you to consider whether your use of fees is correct. For information about the fee, please see helfo.no.

This information is based on statistics from your reimbursements. You can find more information regarding your fee usage in your reimbursement summary.

This email is for guidance and informational purposes. You do not need to reply. If you have questions you can reach out to Helfo at post@helfo.no or 23327040. State the reference 19/74113.

You can find the website of the "did you know"-campaign here. To learn more about the campaign and other important information regarding guidelines and fees, subscribe to Helfo's newsletter.

Kind regards,
Helfo

B: Robustness checks

Table 4: Estimation results from gamma generalized linear mixed model with log link. Table shows marginal effects interpreted as the percentage points change in %DCF. N=691 GPs contributing to 27,304 observations in total. Dependent variable %DCF=no. double-consultation fees/no. regular consultation fees*100

	Unadjusted analysis		**Adjusted analysis**	
Variable:	**Marginal effect (95% CI):**	**p-value**	**Marginal effect (95% CI):**	**p-value**
Before intervention	Reference		Reference	
Control 0-4 months after	-0.80 (-1.84, 0.23)	0.13	-0.74 (-1.81, 0.32)	0.17
Control 5-9 months after	-1.32 (-2.38, -0.26)	0.01	0.62 (-1.07, 2.32)	0.47
Control 10-14 months after	-0.63 (-1.73, 0.46)	0.25	1.32 (-0.66, 3.32)	0.19
Mild 0-4 months after	-4.27 (-5.24, -3.29)	<0.001	-4.16 (-5.17, -3.16)	<0.001
Mild 5-9 months after	-4.41 (-5.40, -3.42)	<0.001	-2.53 (-4.13, -0.91)	<0.01
Mild 10-14 months after	-4.42 (-5.43, -3.40)	<0.001	-2.54 (-4.43, -0.65)	<0.01
Strong 0-4 months after	-5.45 (-6.40, -4.50)	<0.001	-5.34 (-6.32, -4.36)	<0.001
Strong 5-9 months after	-4.10* (-5.10, -3.09)	<0.001	-2.20* (-3.84, -0.57)	0.01
Strong 10-14 months after	-3.78* (-4.81, -2.74)	<0.001	-2.18* (-4.09, -0.27)	0.03
Dummies for months			YES	
Before Covid-19			Reference	
After Covid-19			-2.38 (-3.94, -0.81)	<0.01

*Marks significant differences across time periods within the study arms at the 5% level.

Acknowledgements: We thank HELFO for providing the data and the Research Council of Norway for funding the projects Modernizing the GP Scheme (project number 288592) and Norwegian Centre for Health Services Research (project number 296114).

References

Al-Ubaydli, O., Lee, M.S., List, J.A., Mackevicius, C.L., Suskind, D., 2021. How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling. Behavioural Public Policy 5, 2–49. https://doi.org/10.1017/bpp.2020.17

Al-Ubaydli, O., List, J., 2015. On the Generalizability of Experimental Results in Economics: With A Response To Camerer, in: Handbook of Experimental Economic Methodology. https://doi.org/10.3386/w19666

Antinyan, A., Asatryan, Z., 2019. Nudging for Tax Compliance: A Meta-Analysis. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.3500744

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017. From proof of concept to scalable policies: Challenges and solutions, with an application, in: Journal of Economic Perspectives. American Economic Association, pp. 73–102. https://doi.org/10.1257/jep.31.4.73

Bénabou, R., Tirole, J., 2006. Incentives and Prosocial Behavior. American Economic Review 96, 1652–1678. https://doi.org/10.1257/AER.96.5.1652

Bott, K.M., Cappelen, A.W., Sørensen, E.O., Tungodden, B., 2019. You've Got Mail: A Randomized Field Experiment on Tax Evasion. Management Science 66, 2801–2819. https://doi.org/10.1287/MNSC.2019.3390

Brandtzæg Clausen, V., 2019. Reagerer på at 450 fastleger får brev fordi de bruker for mye tid med pasienter [Letters to 450 GPs regarding excess consultation length causes reactions]. TV2.

Bryan, C.J., Adams, G.S., Monin, B., 2013. When cheating would make you a cheater: Implicating the self prevents unethical behavior. Journal of Experimental Psychology: General 142, 1001–1005. https://doi.org/10.1037/a0030655

Cheo, R., Ge, G., Godager, G., Liu, R., Wang, J., Wang, Q., 2020. The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment. Health Economics Review

2020 10:1 10, 1–19. https://doi.org/10.1186/S13561-020-00290-Z

Dafny, L.S., 2005. How Do Hospitals Respond to Price Changes? American Economic Review 95, 1525–1547. https://doi.org/10.1257/000282805775014236

Deaton, A., 2020. Introduction: Randomization in the Tropics Revisited, a Theme and Eleven Variations. Randomized Control Trials in the Field of Development 29–46. https://doi.org/10.1093/oso/9780198865360.003.0002

Eccles, M., Steen, N., Grimshaw, J., Thomas, L., McNamee, P., Soutter, J., Wilsdon, J., Matowe, L., Needham, G., Gilbert, F., Bond, S., 2001. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: A randomised trial. The Lancet, 357, 1406–1409. https://doi.org/10.1016/S0140-6736(00)04564-5

Ferraro, P.J., Price, M.K., 2013. USING NONPECUNIARY STRATEGIES TO INFLUENCE BEHAVIOR: EVIDENCE FROM A LARGE-SCALE FIELD EXPERIMENT 95, 64–73. https://doi.org/10.1162/REST_a_00344

Gaardsrud, P.Ø., 2020. Styringsdata for fastlegeordningen, 4 kvartal 2019 [GP management data, 4th quarter 2019]. Norwegian Directorate of Health, Oslo.

Godager, G., Hennig-Schmidt, H., Iversen, T., 2016. Does performance disclosure influence physicians' medical decisions? An experimental study. Journal of Economic Behavior and Organization 131, 36–46. https://doi.org/10.1016/J.JEBO.2015.10.005

Gravelle, H., Sutton, M., Ma, A., 2010. Doctor behaviour under a pay for performance contract: Treating, cheating and case finding? Source: The Economic Journal 120, 129–156. https://doi.org/10.1111/j

Gronseth, I.M., Malterud, K., Nilsen, S., 2020. Why do doctors in Norway choose general practice and remain there? A qualitative study about motivational experiences. Scandinavian Journal of Primary Health Care 1–8.

Hafstad, A., 2019. Helfo-kampanjen – en bommert [The Helfo campaign - off target]. Dagens Medisin.

Harrison, G.W., 2021. Field experiments and public policy: festina lente. Behavioural Public Policy 5, 117–124. https://doi.org/10.1017/bpp.2020.28

Harrison, G.W., 2011. Randomisation and Its Discontents. Journal of African Economies 20, 626–652. https://doi.org/10.1093/JAE/EJR030

Harrison, G.W., List, J.A., Burks, S., Camerer, C., Carpenter, J., Gerking, S., Isaac, R.M., Krueger, A., Mcmillan, J., Ortmann, A., Plott, C., Reiley, D., Rutström, E.E., Wilcox, N., 2004. Field Experiments. Journal of Economic Literature XLII, 1009–1055.

HELFO, 2021. Effektvurdering informasjon til leger - takst 2cd. Tønsberg.

HELFO, 2020. Helserefusjon 2019- Statistikk og nøkkeltall for legeområdet [Health care claims 2019 - statistics and data on doctors].

Houser, D., Vetter, S., Winter, J., 2012. Fairness and cheating. European Economic Review 56, 1645–1655. https://doi.org/10.1016/j.euroecorev.2012.08.001

Hu, F., Ben-Ner, A., 2020. The effects of feedback on lying behavior: Experimental evidence. Journal of Economic Behavior and Organization 171, 24–34. https://doi.org/10.1016/j.jebo.2019.12.019

Hysong, S.J., Best, R.G., Pugh, J.A., 2006. Audit and feedback and clinical practice guideline adherence: Making feedback actionable. Implementation Science 1, 9. https://doi.org/10.1186/1748-5908-1-9

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J.M., Odgaard-Jensen, J., French, S.D., O'Brien, M.A., Johansen, M., Grimshaw, J., Oxman, A.D., 2012. Audit and feedback: Effects on professional practice and healthcare outcomes. Cochrane Database of Systematic Reviews 2012. https://doi.org/10.1002/14651858.CD000259.pub3

Ivers, N.M., Grimshaw, J.M., Jamtvedt, G., Flottorp, S., O'Brien, M.A., French, S.D., Young, J., Odgaard-Jensen, J., 2014. Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. Journal of General Internal Medicine 29, 1534–1541. https://doi.org/10.1007/S11606-014-2913-Y

Jamtvedt, G., Young, J.M., Kristoffersen, D.T., Ann, M.O., Oxman, A.D., 2006. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. BMJ Quality & Safety 15, 433–436. https://doi.org/10.1136/qshc.2006.018549

Kesternich, I., Schumacher, H., Winter, J., 2015. Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? Journal of Public Economics 131, 1–11. https://doi.org/10.1016/J.JPUBECO.2015.08.009

Meeker, D., Linder, J.A., Fox, C.R., Friedberg, M.W., Persell, S.D., Goldstein, N.J., Knight, T.K., Hay, J.W., Doctor, J.N., 2016. Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among

Primary Care Practices: A Randomized Clinical Trial. JAMA 315, 562–570. https://doi.org/10.1001/JAMA.2016.0275

Norwegian Directorate of Health, 2020. Årsrapport 2019 [Annual report 2019].

Schwartz, K.L., Ivers, N., Langford, B.J., Taljaard, M., Neish, D., Brown, K.A., Leung, V., Daneman, N., Alloo, J., Silverman, M., Shing, E., Grimshaw, J.M., Leis, J.A., Wu, J.H.C., Garber, G., 2021. Effect of Antibiotic-Prescribing Feedback to High-Volume Primary Care Physicians on Number of Antibiotic Prescriptions: A Randomized Clinical Trial. JAMA Internal Medicine. https://doi.org/10.1001/JAMAINTERNMED.2021.2790

Shojania, K.G., Jennings, A., Mayhew, A., Ramsay, C.R., Eccles, M.P., Grimshaw, J., 2009. The effects of on-screen, point of care computer reminders on processes and outcomes of care. Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.CD001096.pub2

Storvik, A.G., 2019. Leger reagerer på Helfo-brev om tidsbruk [GPs react to Helfo-letter regarding consultation length]. Dagens Medisin.

IV