The Impact of Financial Incentives on Demand and Supply of Health Care Services in Norway

Empirical Studies on Co-Payments and Activity-Based Hospital Financing

Camilla Beck Olsen

Thesis submitted for the Degree of Philosophiae Doctor (PhD)

Department of Health Management and Health Economics
Institute of Health and Society
Faculty of Medicine
University of Oslo
Norway
June 2018

Summary

The financing of health care systems is constantly evolving in order to best achieve objectives such as access to and efficient use of resources. To be able to use financial incentives efficiently, there is need for evidence on to what extent users and providers respond to these incentives. The aim of this thesis is to provide more evidence across different populations, settings and outcomes on how behaviour in health care systems is affected by financial incentives. *Paper I* estimates the effect of a co-payment exemption on adolescent visits to the General Practitioner (GP). *Paper II* tests to what extent hospitals focus on the diagnostic groups that are most financially favourable given costs and reimbursement rates, while, *Paper III* evaluates to what extent a 10% price increase was successful in inducing hospitals to provide more day surgery.

Overall, the findings in this thesis support the existing empirical literature that financial incentives do affect behaviour. The results suggest that adolescents were sensitive to having to pay a fee for visiting their primary care physician and that exempting them from this fee increased their use of health care services by 22.1% among females and 13.8% among males. Hospitals also respond to changes in reimbursement for Diagnostic Related Groups (DRGs). The percentage increase in number of admissions was on average four times higher when the DRG reimbursement was increased, relative to the percentage increase in DRGs with reduced reimbursement rates. However, a 10% increase in DRG weight for day surgery did not seem to provide a strong enough incentive for hospitals to provide more day surgery. The results of the analyses are relevant for both policy-makers and researchers who are interested in the further development of both demand-side and supply-side incentives to achieve health policy objectives such as access and efficiency.

Acknowledgements

No project is ever completed without the help of others, and this thesis is no exception. For the past three years, I have been blessed by having many people who have encouraged and supported me in various ways.

For my main supervisor, Hans Olav Melberg: Thank you for giving me this opportunity and for having confidence in me. Thank you for sharing your valuable time and for listening to me trying to make sense of everything; I will miss our discussions. Also, thank you for introducing me to Python and very patiently teaching me how to use it.

I would also like to thank my supervisor Terje P. Hagen. Thank you for your valuable insights, especially in relation to the financing of hospitals. You have an understanding of the Norwegian health care system that I hope to achieve myself one day.

Not only have I learned a lot about research from my supervisors, but also through informal discussions as well with researchers such as Henning Øien, Tor Iversen and Eline Aas. Thank you all for spending some of your valuable time talking to me. And, thanks to Ole Røgeberg for help with Bayesian modelling and getting my model to converge.

For my fellow colleagues at the Department of Health Management and Health Economics: Thank you for a collaborative and helping atmosphere and for our daily lunches. You have all encouraged me in your own way, but I would like to specifically thank Ge Ge for brainstorming with me, talking through difficult concepts and uncovering new insights (well new to us) related to theory on demand-side and supply-side cost sharing. Also, thanks to Eliva for graciously sharing your office with me for two years, making chapatis and for being a friend.

Finally, for friends and family; you have each in your own way provided encouragement, support and fun times, without which this thesis would have been a more difficult and lonely endeavour. For my sister, Emilie, and brother-in-law, Daniël: thank you for keeping me nourished with delectable food!

List of Papers

Paper I: Did Adolescents in Norway Respond to the Elimination of Co-payments for General Practitioner Services?

Olsen CB, Melberg HO. *Health Economics*. 2018. 7(27):1120–1130.

Paper II: Did Hospitals Respond to Changes in Weights of Diagnosis Related Groups in Norway between 2006 and 2013?

Melberg HO, Olsen CB, Pedersen K. *Health Policy*. 2016. 120(9):992–1000

Paper III: Hospitals' Response to Changes in Reimbursement for Day Surgery: Evidence from Norway

Olsen CB, Melberg HO, Røgeberg O

Submitted

Table of Contents

1 Introduction

The aim of this thesis is to provide more evidence on to what extent financial incentives in health care affect behaviour. Financial incentives can be defined as any monetary factor "that provides motivation for a particular course of action, or counts as a reason for preferring one choice compared to alternatives"(Flodgren *et al.*, 2011). Within health care, both consumer and provider incentives influence decisions on how much health care is consumed and supplied, which in turn affects the resource allocation and size of health care expenditure (Zweifel and Manning, 2000).

Optimal financing systems involve protecting patients against financial risk, supporting providers in providing health care services, while at the same time ensuring efficiency in the level of services consumed and provided (Ellis and McGuire, 1993). Both demand-side and supply-side cost-sharing systems attain efficiency by transferring some of the costs and risks to patients and providers. However, in both cases there is some trade-off between attaining efficient levels of health care and the degree of risk protection provided for patients and providers. This trade-off depends on how sensitive patients and providers are to changes in financial incentives and financing systems. To be able to use financial incentives to attain the goals of efficiency, access and cost containment, policy makers need evidence on to what extent, and under which circumstances, these incentives will work. In general, the existing empirical evidence has demonstrated that incentives affect choices of patients and providers of health care services; nevertheless, the magnitude of the response varies. This thesis will complement and extend the literature by adding further estimates on responses to financial incentives in health care and thus "add to the accumulation of evidence across settings, populations and study designs" (Angrist and Pischke, 2010).

Specifically, *Paper I* contributes to the demand-side cost sharing literature by assessing the effect of a policy that exempted adolescents (between the ages of 12 and 15) from paying a co-payment for consultations with a General Practitioner (GP). The effect of co-payments on demand for health care services for this age group has not been studied extensively in the literature and the aim is to assess to what extent this age group is sensitive to demand-side cost sharing. Since the policy resulted in a natural experiment, other age groups were used as control groups to estimate the effect. In addition, the paper assesses to what degree there might be gender differences in response to co-payments.

Paper II and *Paper III* focus on hospital reimbursement systems and evaluate to what extent changes to prices within the Activity-Based Financing system affect prioritisation between Diagnostics Related Groups (DRGs). The system is meant to reimburse the hospitals fairly for costs associated with each treatment or hospital admission and to not affect prioritisation between patients, treatments or diseases. *Paper II* tests to what extent hospitals focus on the diagnostic groups that are most financially favourable given actual costs and reimbursement rates. The paper identifies DRGs that are financially favourable in this context, and compares changes in activity levels for these DRGs to changes in activity levels for non-favourable DRGs. The paper provides evidence of how differences in profitability across DRGs affect hospital activity levels.

Paper III exploits an exogenous price increase for day surgery DRGs, where the DRG weight was increased by 10 %. The price for inpatient surgery was reduced by 1.4 %, thus resulting in a change in the relative price between day and inpatient DRGs, for a given surgical procedure. The paper evaluates to what extent this price incentive was successful in inducing hospitals to provide more day surgery. Other countries have also experimented with ways to incentivise increased provision of day surgery treatments, and the Norwegian experience could provide additional insights on the extent to which financial incentives can be used as a policy tool to achieve health policy goals such as quality and efficiency.

Each paper estimates the effect of causes, i.e. effects of changes in financial incentives, and in order to get unbiased estimates, the methods in each paper attempt, in various ways, to reduce confounding. In addition, *Paper I* contributes to the methods literature by applying a modified Synthetic Control Method using Elastic Net regression (as suggested by Doudchenko and Imbens (2016)), for weighting control groups to get results that are more robust.

Chapters 2 and 3 will provide some background and context for the papers by first describing the organisation and financing of the Norwegian health care system and then by presenting a short theoretical outline of how demand and supply-side cost sharing may affect patient and provider behaviour. Chapter 3 will also provide an overview of the existing literature related to each paper. Following this, Chapters 4 and 5 will describe the data and methods used in each of the papers, and in particular describe the identification strategies, which determine to what extent a causal effect can be estimated. Chapter 6 provides a summary of the paper and Chapter 7 will briefly discuss the results and methods. This chapter also pinpoints some policy implications and suggestions for further research. Lastly, some concluding thoughts on the thesis will be provided in Chapter 8.

2 The Organisation and Financing of the Norwegian Health Care System

The Norwegian health care system delivers universal access to health care services and is predominantly financed through taxation (Ringard *et al.*, 2013). Services provided are organised at the regional, county and municipal level, but the overall responsibility for the health care sector rests at the national level with the Ministry of Health and Care Services (Ringard *et al.*, 2013). Health care expenditure constitutes 10.0% of GDP (Statistics Norway, 2017), which is close to the OECD average of 9.0% for 2015, but Norway has a high GDP and hence a markedly higher per capita spending on health ($6647) compared to the OECD average ($3997)(OECD, 2017).

The health care system is predominantly publically funded (85.2%), which is higher than the OECD average of 72.9% (Statistics Norway, 2017), but this share varies across the different health care sectors. For primary care services, there is a larger degree of financing through patient co-payments, whereas inpatient secondary care is almost completely financed through public spending.

2.1 Primary Care

As a result of the Municipalities Health Services Act (passed in 1982), the organisation and provision of the primary care system was, from 1984, decentralised to the municipalities. Primary care is mainly financed through municipal taxes, block grants and earmarked grants from the central government (Ringard *et al.*, 2013). The municipalities are responsible for providing primary care services such as GPs, rehabilitation, physiotherapy, home-based care and nursing homes to their inhabitants. They are autonomous in how the services are provided, although the central government controls the overall organisation and financing of primary care (Ringard *et al.*, 2013).

In 2001, a Regular GP scheme was introduced where every person registered in the Population Registry is entitled to register with a regular primary care physician. 99.6% of those entitled make use of this scheme (Norwegian Directorate of Health, 2016b). In terms of resources, there are 10.6 full-time equivalent physicians for primary care services per 10,000 inhabitants (Statistics Norway, 2016). A key role for GPs is their gatekeeping function. Patients need a referral from a GP to access elective secondary health care services and prescription drugs.

Most general practitioners (80%) are independent self-employed primary care physicians who contract with a municipality (Statistics Norway, 2016). These independent self-employed GPs are financed through a mix of capitation (35%), fee-for-service (35%) and patient co-payment (30%)

(Lindahl, 2017). Approximately 10% of general practitioners receive a fixed salary and 2% provide private care without receiving any remuneration from the municipality being financed solely through patient co-payments. The last 8% are temporarily employed as interns, i.e. physicians in training (Statistics Norway, 2016).

Co-payments

During the 1980s, the co-payment system in Norway saw several significant changes. For example, co-payments for prescription drugs were introduced and a national fee was set for GP consultations (Ramm, 2008). The aim of using co-payments has been to reduce growth in public spending and to reduce demand for unnecessary health care (Ringard *et al.*, 2013). Currently, patients make co-payments for general practitioner services, prescription drugs, x-ray and laboratory services, psychologist services, outpatient services at hospitals, rehabilitation, and physiotherapy. For a standard consultation with a general practitioner, patients contributed a co-payment fee of 136 NOK (€18.2)[1] in 2012, which has subsequently increased to 152 NOK in 2017. For general practice specialists the co-payment is slightly larger at 201 NOK (2017). At the same time, there exists a cap on annual co-payments to shield patients from some of the cost (Ramm, 2008). Per 2017, the annual limit was set to 2205 NOK (236 EUR[2]) and patients who reach this limit before the end of the calendar year are eligible for a co-payment exemption card and receive free services for the rest of the year. In 2010, this was registered automatically, resulting in a substantial increase (approx. 36%) in patients who were eligible (Norwegian Directorate of Health, 2011). There is also a second co-payment cap (at 1990 NOK (213 EUR)), which applies to physiotherapy services, some dental treatments and treatment abroad. Approximately 44 800 patients were eligible for the second co-payment exemption card in 2015, whereas 1 140 000 patients (about 21% of the population) were eligible for the first exemption card (Ministry of Health and Care Services, 2016b). The co-payment caps are independent of income.

2.2 Secondary Care

In 2002, the responsibility for providing secondary care services was transferred from the counties to the central government, that is the Ministry of Health and Care Services, as a result of the 2001 Act of State Owned Health Enterprises (Hagen and Kaarbøe, 2006). The aim was to consolidate the responsibility of ownership and financing. The ministry owns four regional health authorities who are responsible for providing services to the inhabitants in their health regions. These services include somatic and psychiatric hospital services (both emergency and elective care),

---

[1] The amount in Euros is calculated using the average exchange rate of 7.47 for 2012 (http://www.norges-bank.no/en/Statistics/exchange_rates/currency/EUR)
[2] The amount in Euros is calculated using the average exchange rate of 9.32 for 2017 (http://www.norges-bank.no/en/Statistics/exchange_rates/currency/EUR)

ambulance, laboratory and radiology services and specialised rehabilitation and substance abuse treatment (2001). The four RHAs are: Northern Norway RHA, Central Norway RHA, Western Norway RHA and South-Eastern Norway and they serve populations varying from 480 000 (Northern RHA) to 2.9 million (South Eastern RHA) (Ministry of Health and Care Services, 2016b). These regions in turn own 19 public hospital trusts. In addition, the RHAs procure secondary care services from private non-profit hospitals and private for-profit hospitals, but the proportion of services procured from private hospitals varies from year to year (Hagen *et al.*, 2017).

Financing of the secondary care system

Financing of secondary care in Norway has undergone several changes in the past decades in order to find the right incentives and optimal contract to achieve health policy goals. During the past 35 years, two major reforms have altered the financing of Norwegian hospitals. In 1980, to control increasing costs, the reimbursement system was changed from a retrospective reimbursement based on length of stay to a prospective block grant financing system based on county tax revenues, the age composition of the population and population density (Johnsen and Bankauskaite 2006). The prospective block grant financing system reduced overall costs and length of stay; however, there was an increase in waiting times for treatment (Hagen and Kaarbøe, 2006). In 1997, a part of the counties' reimbursement for secondary care services was made contingent on its activity[3] (Johnsen and Bankauskaite, 2006). The government introduced the activity based financing system (ABF), where activity was reimbursed prospectively based on the Diagnostic Related Groups (DRG) system. The goal was to refine and adjust the financing system in order to stimulate increased activity and efficiency and thus decrease the waiting times that came about as a result of the prospective system introduced in the 1980s (Ministry of Health and Care Services, 1995). In the years after the introduction of ABF (1997 to 2000), activity increased by 3.2% per year and technical efficiency increased by 2% (Biørn *et al.*, 2003).

With the new system, the secondary care sector in Norway is currently financed partly (50%) through a prospective ABF system and partly (50%) through a prospective block grant system.[4] The block grant is based on the patient case-mix in the hospital's catchment area. The ABF system is based on the DRG system. A DRG code represents a group of diagnoses or treatments, which are intended to be similar both medically speaking and in terms of resource use. In the ABF system for 2017, there were 866 DRG codes and each code is assigned a cost weight. The cost weight represents each DRG code's average cost relative to the cost of the average patient

---

[3]It was intended that the new financing system would form the basis for activity-based contracts between the county and its hospitals; however, it was not until 2000 that all counties introduced activity-based financing for their hospitals (Biørn *et al.*, 2003).

[4]As of 2014, approximately 0.5 per cent of the hospitals' budget is determined by how well they perform on a set of quality indicators, however, this amount is negligible (Helsedirektoratet, 2014b).

and is based on patient and accounting data from a select number of hospital trusts (for 2015, data from 16 hospital trusts was included in the calculations). This is updated annually based on cost data from two years prior (Norwegian Directorate of Health, 2014). The cost of the average patient gives a DRG weight of one. This is also referred to as the unit price, which was approximately 41 462 NOK (4631 EUR) in 2015. The ABF portion of the total reimbursement is then based on hospital activity within each DRG, multiplied by the DRG weight, the unit price and the percentage reimbursed through ABF (50% from 2014)

The aim of the ABF reimbursement system is to support the regional health authorities in their provision of health care services to the Norwegian population. The financing model entails that the RHA distributes the block grant and the ABF reimbursement to the individual hospital trusts in their region (Norwegian Directorate of Health, 2015). Because each weight associated with a DRG is based on average costs collected from a sample of the hospital trusts, it does not have the granularity required for distribution further down to each individual department at a hospital or for each individual patient treated (Norwegian Directorate of Health, 2015). Despite this constraint, the DRG points system is in practice used to distribute resources to each department in the hospitals based on the DRG points they produced (Kjekshus and Harsvik, 2007). Although the incentives from a DRG based financing system exists at department level in Norwegian hospitals, individual physicians and other medical staff are salaried, and their income is not linked to hospital activity.

3 Theoretical Background and Empirical Literature

This chapter will first briefly describe patient and provider behaviour under demand and supply-side cost sharing, followed by a review of the empirical literature identifying what is currently known about patient and provider responsiveness to changes in financial incentives. Lastly, it will summarise how the papers in this thesis will contribute to the empirical literature.

3.1 Demand-side cost sharing

Due to uncertainty in health care outcomes and the financial risk for patients consuming health care services, some form of health insurance (taxes, social health insurance or private health insurance) is socially optimal (Arrow, 1963). With insurance, individuals pay a premium that reflects their probability of being sick and the cost of being sick. However, patients may have an incentive to use care beyond the premium, because the price of these services to them is equal to zero. This implies that these additional services would not necessarily be used if there was a price to be paid by the consumer. This is often referred to as moral hazard and is defined by Pauly (1968) as "demanding more at zero price than at a positive one."

There are two different ways in which moral hazard may arise, as a consequence of patients becoming more insured against risk and costs. These were first referred to as *self-protection* and *self-insurance* by Ehrlich and Becker (1972), but are most often referred to as *ex-ante* and *ex-post* moral hazard in the literature. *Ex-ante* moral hazard refers to patient behaviour prior to falling sick and states that having full insurance will reduce patient incentives to participate in preventative efforts that reduce the probability of falling sick (Barros and Martinez-Giralt, 2012). This is because insurance reduces the benefits from investing in prevention, since there is less income loss when the patient is sick. Nevertheless, *ex-ante* moral hazard may be mitigated because there is a utility loss from being sick and the loss of health is not necessarily fully compensated by insurance (Cutler and Zeckhauser, 2000, Kenkel, 2000). *Ex-post* moral hazard describes patient behaviour once a sickness episode has occurred and states that having full insurance will increase demand for health care services (Barros and Martinez-Giralt, 2012).

To minimise the moral hazard behaviour of patients, some form of cost sharing is introduced. Cost-sharing is any financial mechanism by which the patient shares the cost of the services (Barros and Martinez-Giralt, 2012). Cost sharing can often be found in the form of deductibles (amount paid before insurance covers expenses), co-insurance rates (percentage rate of costs incurred per service) or co-payments (flat fixed fee per service). By introducing cost sharing, the patient becomes more sensitive or aware of the cost of the treatment and thus it gives some incentive to reduce demand for health care. On the other hand, as mentioned in the beginning of

the chapter, in financing health care there is a trade-off between efficiency and risk protection for patients. There are concerns that having too much cost sharing reduces patient access to services and their financial risk protection and especially that it may reduce use of necessary health care services.

The optimal level of insurance depends on how sensitive patients are to cost sharing. The general conclusion from the literature is that increases in cost sharing are associated with a decrease in demand for health care services, however, the magnitude of this effect varies (Kiil and Houlberg, 2014, Remler and Greene, 2009, Skriabikova *et al.*, 2010, Pendzialek *et al.*, 2016). According to the systematic reviews, patient sensitivity may depend on type of cost sharing, institutional settings such as availability of alternative health care services, health status, income, age and gender. Given that differences in response to cost sharing are context-dependent, empirical evidence from a wide variety of contexts is needed. There is extensive research on cost sharing; however, there are at least three areas in which the empirical literature is lacking: 1) evidence from the Norwegian health care system 2) evidence on adolescent response to cost sharing and 3) evidence on gender differences. This will be explored in the section below.

Empirical literature

The majority of studies on demand-side cost sharing and its effect on demand for health care services stem from the USA. One of the first and most comprehensive studies, which involved a randomised experiment was the RAND Health Insurance Experiment conducted in the USA in the 70s. Approximately 6000 people were randomly assigned to health insurance plans, which differed in terms of levels of coinsurance rates (0%, 25%, 50% or 95%) and upper limits on out-of-pocket expenses (5%, 10% or 15% of family income) (Manning *et al.*, 1987). For outpatient services, families on the insurance plan with 25% coinsurance rate reduced their number of visits by 27% compared to those on the free plan, and families with 95% coinsurance rates reduced number of visits by 31% (Manning *et al.*, 1987).

The RAND experiment reported important results with respect to response to varying levels of insurance coverage, yet there are limitations on the extent to which the results can be generalised to current health care systems, both in the US and in Europe. Zweifel and Manning (2000) point out that the health care delivery systems have changed considerably since the 1970s. In particular, the current US system has a more extensive use of non-price rationing of health care demand in the form of managed care than it did previously and this might affect how patients respond to changes in price for health care services. In addition, the types of health care services available today (more diagnostic procedures, management of chronic conditions, etc.) are different compared to what was available in the 70s, and the demand structure for these services might

differ (McGuire, 2011). Nevertheless, recent US studies have estimated similar effects of health insurance on physician visits (Freeman *et al.*, 2008). While the RAND experiment focused on differences in health insurance levels, a more recent randomised experiment that took place in Oregon, USA in 2008, investigated the impact of introducing insurance coverage and showed that access to Medicaid increased the number of outpatient visits by 55% (Finkelstein *et al.*, 2012).

Several recent changes in cost-sharing arrangements in Europe have resulted in papers evaluating demand responses in health care systems with universal health care coverage. Focusing on results for GP visits, there seems to be mixed evidence as to the size of the estimated response, but overall a smaller response compared to the US studies. For example, in Germany, a co-payment of ten Euros for the first GP consultation per quarter was introduced in 2004. Adjusting for the non-linear price effect that arises when the co-payment is only eligible for the first consultation per quarter, the number of GP visits was estimated to have decreased between 4% and 8% (Farbmacher and Winter, 2013). In Sweden, on the other hand, an increase in co-payment for GP services by approx. 5.5 EUR in 2012 was found to have no effect on utilisation (Jakobsson and Svensson, 2016). Similarly, a French study evaluated a co-payment increase of 10% in private health insurance contracts, but also found no evidence on changes in demand for GP visits (Chiappori *et al.*, 1998). In Belgium, a 48 % increase in co-payments for GP office visits was found to have a price elasticity of -0.06 for men and -0.01 for women (Cockx and Brasseur, 2003).

Furthermore, several studies have focused on estimating the effect of co-payments on children's demand for health care services rather than adults. The RAND experiment in the USA was also one of the first studies to estimate children's (age 13 years or younger) use of both outpatient and inpatient health care services in response to cost-sharing. They found that cost-sharing was mostly effective in reducing outpatient care, where children receiving a free care plan had 67% more treatment episodes than children with the 95% care plan (Leibowitz *et al.*, 1985). Both the Czech Republic and Sweden have introduced policies where children have been exempted from paying co-payment fees for outpatient and GP services. The effect estimates varied from 0% to 10%, reflecting differences in the size of co-payment exemption (Paul and Nilsson, 2014, Votapkova and Zilova, 2015, Zápal, 2010)[5]. At the local government level in Japan, there have been several changes to the subsidies for children's use of outpatient health care services. Kato and Goto (2017) utilise variations at the local level in expanding the upper age limit for subsidies and found that outpatient care subsidies decreased the number of inpatient admissions. Miyawaki *et al.* (2017), on the other hand, focus on children being disqualified from the medical subsidy and having to pay a coinsurance rate of 30% for outpatient services. They found that the

---

[5] *Paper* I describes these studies in more detail.

disqualification for 10 year olds resulted in a 19% decrease in outpatient visits, but did not affect the expenditure per visit.

However, the previous studies have either estimated effects on young children or included both children and adolescents between the ages of 0 and 20. None of them has focused specifically on adolescents. As we argue in *Paper I*, adolescents may be more sensitive to co-payments due to greater financial constraints and confidentiality concerns. A recent study from Norway evaluated use of GP services for 16 year olds (Landsem and Magnussen, 2018). Since the 2010 co-payment reform raised the age threshold for exemption from 12 to 16, the paper estimated to what extent this threshold led to a decrease in use of GP services using a regression discontinuity (RD) design. The results indicate that having to pay a co-payment of 172 NOK at age 16 led to a reduction in GP visits of between 10 to 15%.

Lastly, very few studies have assessed to what extent there are gender differences in response to co-payments. The systematic reviews by Remler and Greene (2009) and Pendzialek *et al.* (2016) identified 3 and 2 studies, respectively, that differentiate between male and female response. Remler and Greene (2009) concluded that the limited evidence suggested a greater response to changes in cost-sharing for women, whereas the studies included in Pendzialek *et al.* (2016) gave mixed evidence and thus no general gender effect could be deduced. Only one of the studies in the systematic reviews assessed gender differences amongst children concerning co-payments and outpatient physician visits (Cherkin *et al.*, 1989). They found that for children and young adults, a co-payment of $5 had twice as large an effect for females compared to males on outpatient visits, however this difference was only evident amongst the younger age groups and no gender differences were found for adults over 40 years of age.

In summary, the previous empirical literature on demand-side cost-sharing has provided evidence that patients respond to cost-sharing, but there are important differences between countries in how demand-side cost-sharing is organised and other system characteristics can impact the response. For example, previous US studies provide evidence on cost sharing and use of outpatient services, which includes primary care level services. However, in Norway, patients need to see a GP for referrals to outpatient specialist care. Therefore, the types of services that are included in US outpatient studies may differ from services provided at the primary care level in Norway. Moreover, recent studies from the US and the Netherlands have showed that the types of cost sharing employed may affect responsiveness, contrary to the theory that consumers should respond equally to all dollars of marginal out-of-pocket payments (Remmerswaal *et al.*, 2017, Stockley, 2016). For example, health care consumers seem to be more responsive to changes in co-payments compared to insurance deductibles. The argument is that "consumers have better information on co-payment costs and are thus more price responsive to costs they can observe"

(Stockley, 2016). This shows that we need empirical evidence from different systems. In addition, few studies have focused on adolescents as a patient group or estimated any gender differences. Following from this, *Paper I* contributes to the literature by estimating the effect of a co-payment exemption on adolescent use of GP visits for males and females separately in a health care system with universal health care coverage.

3.2 Supply-side cost-sharing

Provider reimbursement systems in the health care sector aim to incentivise cost efficiency, allocative efficiency and cost containment (Olsen, 2009) and can do so by transferring some of the financial risk from payers to providers. There are several different types of payment systems and each gives different incentives according to how risk is shared between payers and providers. Jegers *et al.* (2002) provide a typology of the different payment systems and classify systems along two dimensions; retrospective vs. prospective and variable vs fixed. Most European countries use some form of prospective payment systems (PPS) to reimburse providers of secondary care (Cots *et al.*, 2011). Drawing on both theoretical and empirical studies, provider behaviour under a PPS is described below.

There are two main features of a variable PPS (also referred to as activity-based financing in some countries). First, hospital activity is measured using a classification system and most countries employ the Diagnostic Related Groups (DRG) system. A DRG code represents a group of homogenous medical diagnoses or treatments with similar lengths of hospital stay for patients of similar age and gender. Second, the price per DRG is fixed and set prospectively (often based on lagged national average costs for each DRG). The net revenue or profit providers receive is then determined by the price, P, for each DRG, j, multiplied by the quantity, Q, of services supplied within each DRG minus the cost, C, per DRG[6].

$$\Pi = \sum_{j=1}^{n} P_j Q_j - \sum_{j=1}^{n} C_j Q_j \tag{1}$$

The providers bear the risk if their costs are higher than the price, but at the same time, they keep the profits if the costs are below the price. In general, this type of system gives incentives for cost and technical efficiency since providers aim to keep the costs below the set price.

Several empirical studies have estimated the extent to which reimbursing hospitals using a variable prospective financing system has resulted in increased efficiency. In Norway, the introduction of Activity-Based Financing (ABF) resulted in improvements in technical efficiency, but with no significant improvement in cost-efficiency (Biørn *et al.*, 2003, Hagen *et al.*, 2006). A comparison of the evidence on efficiency from the Scandinavian countries (Norway, Sweden and

---

[6] Notation is borrowed from Kasper H & Sloan F A. 2008. *Incentives and Choice in Health Care*, MIT Press.

Denmark) concludes that there is mixed evidence on increased efficiency as a result of ABF; most studies report clear gains in technical efficiency, but small gains to made in terms of cost-efficiency and stronger results in Sweden compared to Norway and Denmark (Jakobsen, 2010). In a Nordic perspective, Norway and Sweden have experienced smaller cost-efficiency gains compared to Finland (Linna *et al.*, 2010). There is also no conclusive evidence internationally on the effect of ABF on hospital efficiency. Street *et al.* (2011) provide an overview of some of the international literature and report mixed evidence. They argue that a reason for the mixed evidence may be the different starting points and contexts within which the reforms were implemented.

Nevertheless, the quest for efficiency improvements in hospital provision of treatments is subject to a trade-off as well. A pure PPS will incentivise hospitals to provide treatment below the set price, but unintended consequences may exist from the cost-reducing efforts. These unintended consequences have been derived in several theoretical papers (Ellis, 1998, Ellis and McGuire, 1986, Ellis and McGuire, 1996, Hodgkin and McGuire, 1994, Dranove, 1987). These studies show that to maximise profit a hospital may seek to attract profitable, low severity patients and thus unnecessarily admit such patients, or to undersupply health care and reduce intensity of treatment supplied per episode. Other concerns are gaming or DRG up-coding effects, where providers increase their revenues by assigning higher revenue-generating DRGs to the illnesses suffered by their patients (Ellis and McGuire, 1986).

The magnitude of this type of response is predicted to be greater in situations where the reimbursement is not cost-neutral, i.e. for each additional unit of care the relationship between expected marginal costs and marginal returns is not identical (Dranove, 1987, Jegers *et al.*, 2002). Ellis and McGuire (1986) also show that these incentives are greater when the provider or physician is not a perfect agent for the patient, i.e. the provider cares about profit to some extent. There is no consensus in the literature on to what extent hospitals (both public and private) are profit maximisers or motivated by altruism. It seems that both public and private providers are profit maximisers, but public providers may be more altruistic than private ones. See Barros and Siciliani (2011) for a brief overview of the existing literature on public and private differences in motivation.

Although the majority of the empirical literature, which has evaluated the effect of implementing DRG financing systems has been concerned with efficiency concerns, other studies have examined other outcomes to assess its effects on activity levels, length of stay and health related medical outcomes such as mortality. For Norwegian hospitals, Kjerstad (2003) found that the introduction of the ABF reform, in some counties, led to increased number of patients admitted and production of DRG points compared to counties who had not yet introduced ABF. In terms

of quality, Hagen *et al.* (2006) report increased patient satisfaction as a result of the ABF reducing waiting times. An international systematic review summarised the effects of implementing ABF on outcomes such as mortality, readmissions, discharge destination, severity of illness and volume of care (Palmer *et al.*, 2014). In general, the studies reported no negative effects on mortality or readmissions, but there is mixed evidence in terms of severity of illness and volume of care.

Within existing PPS, any DRG price change will affect the marginal revenue relative to the marginal cost and if marginal revenue increases relative to the marginal cost for a DRG, then providers have incentives to increase activity levels to receive more revenue. Recent literature has focused on assessing hospital responsiveness to these price changes. Estimating hospital response to price changes is important because it determines to what extent financial incentives (in terms of DRG prices) can be used to achieve health policy goals by, for example, increasing overall hospital activity or prioritising DRGs that will increase efficiency and patient outcomes (Maynard and Bloor, 2012). It is also important to assess how hospitals will respond. Depending on the nature of the price change, hospitals can increase their revenue by up-coding, decreasing treatment intensity or increasing number of admissions.

The results are mixed as to the extent of responsiveness. Although most studies find some degree of response to price changes (which is consistent with hospitals being concerned with profits and revenue), other studies describe the response as negligible and argue that patient benefit is also of concern to hospitals. The following section provides a short summary of some studies evaluating hospital response to price changes within the setting of PPS.

The following studies have examined the effect of changes in profitability of DRGs and its effect on both hospital activity and treatment intensity. Januleviciute *et al.* (2015) investigated changes in DRG prices over time in Norway. The results suggest that a 10% increase in price leads to an increase of 0.49% in number of patients treated. This result was true only for medical DRGs; they found no effect for surgical DRGs. In the US, Lindrooth *et al.* (2007) examined whether exogenous cuts in reimbursements for patient admissions, as a result of the Balanced Budget Act (BBA), affected treatment intensity. They compared treatment intensities in the pre and post BBA periods. Treatment intensity was reduced for diagnoses that were more generously reimbursed before BBA; however, there was no change in treatment intensity for unprofitable diagnoses. Similar experiences regarding changes in profitability of DRGs can also be found in Taiwan. Increases in profitability of orthopaedic surgical DRGs was associated with increases in the share of the respective DRGs (Liang, 2015). Similarly, an Italian study found significant changes in volume for surgical DRGs following a policy change that increased the prices of a subset of

DRGs by 6% on average in 2007. A 1% increase in price for surgical DRGs was followed by an increase in volume of 1.7% in the short term and 4% in the medium term (Verzulli *et al.*, 2016).

Other studies have focused on the effect of changes in relative prices between similar, substitutable DRGs. Many DRG systems differentiate between similar diagnoses or procedures, but with different intensity of services provided for a particular diagnoses or procedure. For example, hospitals often receive a higher price for patients that are admitted with either comorbidities or who experience complications, or they may receive a higher price for procedures that are more expensive, require more services or involve an overnight stay.

By refining the DRGs and thus increasing the price in favour of the more intensive treatment, an unintended consequence may be that hospitals have incentives to game the system by up-coding patients into higher paying DRG categories (Ellis, 1998, Simborg, 1981). In Norway, several studies have found evidence of up-coding. Januleviciute *et al.* (2015) found that a 10% increase in the price ratio between DRGs with and without complications increased the proportion of patients coded with complications by 0.3-0.4% points. A recent, similar, study found that a one DRG point difference in price between a complicated and uncomplicated group was associated with an increase of 14 percentage points in the share of complicated discharges within a DRG pair (Anthun *et al.*, 2017). However, they found no effect of changes in the share of ABF on coding behaviour. The international literature also confirms the Norwegian experience in terms of up-coding. In the US, Dafny (2005) investigated the effect of an administrative change in the DRG coding, which resulted in a relative price increase of 7% for DRGs with complications compared to DRGs without complications. She found that hospitals responded by up-coding patients into the diagnoses with largest price increases and the response was largest for the oldest patients, since it was easier to justify complications in this group. Barros and Braun (2017) report a similar experience in Portugal comparing DRGs with and without complications. DRGs with larger price increases were associated with a larger proportion of patients assigned to complication DRGs.

Moreover, DRG refinements may also lead to actual changes in hospital activity or treatment intensity (Hafsteinsdottir and Siciliani, 2010). For example, the introduction of Payment by Results in England in 2003 resulted in a more generous reimbursement for uncemented Hip Replacement compared to the cemented treatment procedure, although the guidelines recommended the latter. Comparing activity with Scotland where no such financial differential existed, the financial incentives in the UK were linked with a higher uptake of the more expensive uncemented procedure (Papanicolas and McGuire, 2015). In the US, Gilman (2000) investigated how the introduction of procedure-based DRGs for HIV treatment affects the intensity of services provided for the new high paying procedural DRGs compared to the low

paying non-procedural DRGs. Hospitals responded to the lower average payment for non-procedural DRGs by reducing the average intensity of resources and vice versa for the procedural DRGs. However, a Norwegian DRG refinement that resulted in substantial incentive shifts favouring inpatient surgery in 2010, did not seem to result in any significant change in treatment behaviour (Huitfeldt, 2017). The author argues that hospitals care about patient welfare and thus do not necessarily respond to price incentives once patients are admitted (i.e. at the intensive margin).

The literature has focused on a variety of different price changes and different outcomes. Some have focussed on overall DRG price increases, whereas others have focussed on changes in relative price differences between similar or substitutable DRGs. In terms of outcomes there seems to be consensus regarding evidence of up-coding as a result of relative price changes, but less overwhelming evidence concerning changes in activity or volume or treatment decisions.

Similarly to the discussion of differences in results related to demand-side cost sharing, some of the differences in empirical results on hospital responsiveness to price changes may stem from differences in the organisation and implementation of the ABF schemes and how DRG prices are calculated. For example, the strength of the incentives may vary depending on existence of other sources of funding such as a global budget. Countries where the majority of hospital revenue is related to DRGs may experience stronger incentives (Cots *et al.*, 2011). Other differences relate to the level at which the ABF system is implemented. As mentioned in the description of the Norwegian health care system, the ABF finances the activity at the RHA level and the extent to which individual hospitals are financed according to DRG activity differs among RHAs. This may dilute incentives at the hospital level compared to other countries.

Most DRG classifications systems in use in Europe for ABF are based on the Health Care Financing Administration System introduced in 1983 under the US Medicare system. However, the collection of cost data and calculation of reimbursement rates differs substantially among European countries (Schreyogg *et al.*, 2006). Some countries are more heterogeneous in terms of setting prices. For example, in Finland and Sweden the base rates for DRG prices are calculated separately for each hospital, while in Norway the base rate is applied to all hospitals (Cots *et al.*, 2011). This means that in Norway, there is potentially a greater discrepancy between reimbursement and costs at the hospital level and thus there are greater incentives for adjusting activity levels to maximise profits.

Given these differences across countries in the implementation of ABF, it is important to evaluate Norwegian responses to price changes. Some aspects of the Norwegian health care system may dilute the effects, whereas other aspects may create stronger incentives.

Both *Paper*s *II* and *III* assess the effects of price changes, which in some way alter the marginal revenue relative to the marginal costs. In a cost-neutral system, where reimbursement is equal to the marginal cost, the financing should not give incentives to prioritise between different treatments.

However, reimbursement is often imperfectly related to costs simply because the DRG prices are updated with a lag. This provides the motivation for *Paper II*. During a given year, costs related to producing the DRG might change while the reimbursement is fixed, and a DRG may become marginally financially favourable. Given that hospitals operate within a budget constraint, *Paper II* is interested in estimating whether the quantity of admissions increased more for DRGs that are financially favourable compared to DRGs that are not.

For surgical DRGs, the Norwegian ABF system has refined the DRGs to reflect the average costs of either admitting patients for inpatient surgery or day surgery. *Paper III* assesses an increase in the DRG weight for day surgery. This means that the marginal revenue for day surgery is greater compared to previous years and suggests that this should induce increased activity for day surgery. At the same time, the treatment decision for the average patient depends in part on the relative price difference between inpatient surgery and day surgery. It is therefore also important to take into account the impact of the price changes on the relative price difference between inpatient and day surgery.

4 Data

The aim of the following sections is to complement the description of the data employed in the papers by describing in more detail the registries as data sources. Sections 4.1-4.4 will provide a brief description, while section 4.5 considers overall data quality reported to the registries.

The data for the analyses are based on data from two national registries: Control and Payment of Health Reimbursement (KUHR) and the Norwegian Patient Registry (NPR). This gives us the advantage of doing analysis on population registry data and reduces concerns about sampling issues.

A common feature of the data is that they are aggregated and not estimated at the individual level. The data are aggregated up to the unit of interest for each paper and we are thus evaluating policies at the level at which they target units of interest. This means that for *Paper I*, the data are aggregated at the age group level, since the policy targeted different age groups. For *Papers II* and *III*, the data are aggregated at the DRG level and DRG-RHA level since we are interested in policy effects at the DRG level. For *Paper III* specifically we are also interested in the RHA response to financial incentives.

4.1 Control and Payment of Health Reimbursement (KUHR)

For information on number of GP visits amongst adolescents, we contacted the Norwegian Health Economics Administration (HELFO) and received data from their administrative systems (KUHR). The KUHR database collects data for reimbursing the GPs and has information on every patient contact with a GP for which the GP receives a tariff. GP visits were electronically registered as of 2006 and the database does not include contacts with the GP with paper-based reimbursement claims or with private-practicing GPs[7] who are not entitled to reimbursement from the municipalities (Secretariate for National Health Registry Project, 2009). According to HELFO, about 3.2% of the data was missing for 2006, due to some lag in GPs converting from a paper-based to electronic-based system, but this has improved. For *Paper I*, the dataset consisted of aggregated numbers of GP consultations for each age group and gender in the period 2006 to 2013. The number of visits is based on consultations that generated a reimbursement tariff for standard consultations, emergency contacts with the GP office or visits related to psychotherapy and group therapy. Gender was unidentified for 0.1 % of the observations in the dataset. It was assumed that the missing consultations were randomly distributed and not systematically biased with respect to different age groups or gender

---

[7] About 2% of GPs are not registered in KUHR because they are in private practice without municipal contracts.

4.2 Norwegian Patient Registry (NPR)

*Paper*s *II* and *III* used data from the Norwegian Patient Registry (NPR). NPR is owned and managed by the Norwegian Directorate of Health and has registered both administrative and medical information on every hospital admission since 1997 (Norwegian Directorate of Health, 2016a).

Hospital admission data was obtained by accessing the publicly available aggregated hospital admissions data for all somatic hospitals, reported to the NPR. Hospital activity for the analysis in *Paper II* was aggregated at the DRG level, whereas for *Paper III* the activity was aggregated at DRG and RHA level. Since we used aggregate data published on the Directorate of Health's website, observations were missing if a DRG within an RHA was registered with fewer than three admissions. In addition, not all hospitals admit patients for all possible DRG codes. The data was coded so that it was not possible to determine whether missing observations were due to an RHA not providing a certain surgical procedure or if they had registered fewer than three admissions for that DRG code.

4.3 Norwegian Directorate of Health - Activity-Based Financing guidelines

Information on DRG weights for both *Paper II* and *III* was obtained from the Activity-Based Financing guidelines report, which is published every year and includes information about any changes to the financing system for the next year. Appendix A in the report has a list of all DRGs, DRG categories, DRG weights and threshold values for average length of stay. Both *Paper II* and *III* exclude observations for which DRG weights were missing, due to changes in the DRG classification system.

4.4 Statistics Norway

Data on the population for year, age and gender was obtained from Statistics Norway, table 10211 (Statistics Norway). These numbers from Statistics Norway were used for *Paper I* to construct the outcome variable, defined as GP consultations per capita. Population numbers are measured 1st January each year.

4.5 Data quality

The NPR is generally deemed to be of good quality, at least in terms of data completeness and coverage (Secretariate for the National Health Registry Project, 2016). There have been some concerns regarding the quality of the coding practice and thus the correctness of the classification of patients to DRG groups. A recent report indicates that 28 % of hospital episodes should have been classified to a different DRG group, which in turn will affect the reimbursement hospitals

receive (Office of the Auditor General, 2017). For *Paper III*, issues regarding coding practice should not be of great concern since there is little discretion in classifying a patient stay as a day treatment or inpatient treatment. For *Paper II*, on the other hand, this might be of concern if an increase in hospital activity is a result of opportunistic coding practices rather than the reimbursement. First, the paper attempts to exclude DRGs that might be subjected to up-coding. Second, the report only investigated hip fractures and pneumonia, so it is uncertain to what extent this is pervasive across all diagnostic groups.

KUHR receives data based on activity and services for which the GPs' receive a tariff. It is considered that this ensures data completeness in terms of reporting of the services that are reimbursed (Ministry of Health and Care Services, 2016a). There have been some reports in the media of fraud regarding reimbursement claims for consultations that never took place[8]. However, this a small percentage of the total number of consultations registered in KUHR and they did not seem to affect some age groups more than others.

Table 4-1 Table of data sources and variables used in each paper

Paper	Data Sources	Variables	Years	Level of observation
Paper I	KUHR	GP visits, year, age, gender	2006-2013	Year, age, gender
	Statistics Norway, table: 10211	Population, year, age, gender		
Paper II	NPR	Activity, DRG code, year, DRG category, DRG weights	2006-2013	Year, DRG
	Norwegian Directorate of Health	DRG code and DRG weights	2006-2013	
Paper III	NPR	Day surgery activity, inpatient surgery activity, DRG code, year, RHA, DRG weights	2011-2016	Year, DRG, RHA
	Norwegian Directorate of Health	DRG code and DRG weights	2011-2016	

---

[8] See for example https://www.nrk.no/dokumentar/fastlegekaka-blir-ulikt-fordelt-1.14032216 . For 2015 and 2016 13 mill NOK had to be repaid back to HELFO. This represents 0.26% of the total reimbursements that are registered in KUHR per year.

5 Methods

As mentioned in the introduction, the topic of this thesis is estimating to what extent financial incentives in health care affect behaviour and thus attempt to estimate causal relationships. In econometrics and policy evaluation, causal relationships are often defined in terms of a counterfactual (Angrist and Pischke, 2010, Heckman, 2008) although there are different approaches to identifying and estimating the causal effect through this definition.

The intuition behind the counterfactual definition is to determine causality in terms of 'what if' questions and define causal relationships in relation to counterfactual or alternative outcomes. To illustrate, assume we have a group of individuals who are exposed to a treatment. To find the causal effect of this treatment, we need, for each individual, the outcome observed if the individual is exposed to treatment, Y_i^1, and the outcome observed if the individual is not exposed, Y_i^0. The causal effect is calculated by subtracting Y_i^0 from Y_i^1. The problem is that we can at most observe one of these outcomes because the same individual cannot be both exposed and not exposed to treatment at the same time. This is often referred to as the fundamental problem of causal inference (Holland, 1986). Nevertheless, a causal effect can still be estimated at the aggregate or population level, under certain assumptions. The expected causal effect δ is thus $E(\delta) = E(Y^1) - E(Y^0)$. This is often referred to as the average treatment effect (ATE).

Randomised experiments are often proposed as the ideal solution to estimating causal effects. The counterfactual framework determines that a causal relationship can be established if you have two groups who are as similar as possible, but only differ in treatment assignment and this can be achieved through randomised experiments. If the non-treated individuals are similar to the treated individuals, the effect of other characteristics that might affect the observed outcome are removed, and thus any change in the outcome variable must be due to the cause of interest.

Randomised experiments in health economics are relatively rare; however, recent empirical studies have made use of naturally occurring experiments. These types of experiments are also referred to as quasi-experiments. The terminology used and definition differs among research disciplines, but the main idea is that there is an exogenous change in the variable of interest, which leads to a treated and a non-treated group, but the researcher has no control over who is assigned to treatment or not. Thus, there is variation in treatment assignment, but this is not perfectly randomised. Methods exploiting these exogenous variations are sometimes referred to as design-based approaches (Angrist and Pischke, 2010). This approach forms the basis for the method applied in *Paper I*.

For *Paper*s *II* and *III* on changes in financial incentives for hospitals there is no policy reform that led to a clear natural experiment in terms of treated and control groups. Nevertheless, the papers use the idea of comparing similar observations and keeping everything else constant to reduce potential confounding in the estimates.

Any analysis that attempts to say something about a causal effect needs an identification strategy. An identification strategy states to what extent it is possible to learn the true value of an effect by describing the assumptions needed and proof that if assumptions hold then a causal effect can be identified (Keele, 2015). For example, the design-based approach determines that a causal relationship can be established if you have two groups who are as similar as possible, but only differ in treatment assignment. However, the challenge is to ensure that we have an appropriate non-treated group, or counterfactual. Further assumptions need to be made to ensure that all meaningful differences between the two groups are controlled for and that the effect of other factors that might affect the outcome has been removed. Each section below describes the identification strategies for each paper in more detail, followed by a brief description of the estimation procedures.

5.1 Paper I

5.1.1 IDENTIFICATION: SYNTHETIC CONTROL METHOD

Paper I exploits a natural experiment and the existence of potential control groups to construct a counterfactual and estimate a causal effect. However, since the treatment assignment was not random, the analysis needs to ensure that the control groups are a valid counterfactual. This involves a choice about what observed quantity is a good counterfactual for the treated units (Keele and Minozzi, 2013).

To help identify relevant control groups, the paper used the Synthetic Control Method (SCM) developed by Abadie and Gardeazabal (2003) and Abadie *et.al* (2010, 2015). Abadie *et.al* (2010, 2015) argue that for aggregate data a weighted combination of control units is better at depicting the characteristics of the treated unit than any single unit alone. The identification strategy of the SCM is that groups with similar outcomes in the pre-treatment period are assumed to have similar treatment-free outcomes in post-treatment periods. This is referred to as the *independence conditional on past outcomes* assumption (O'Neill *et al.*, 2016). Thus units with similar past outcomes are likely to also be similar in terms of unobserved confounders (O'Neill *et al.*, 2016). By combining several units and allowing the weights of the control units to vary, the SCM allows for the effects of observed and unobserved predictors of the outcome to change over time because they are controlled for by including linear combinations of pre-intervention outcomes (Abadie *et al.*, 2010). The synthetic control group, or counterfactual, is thus defined as the sum of the weighted observations of the outcome, *Y*, for the control groups, *j*:

$$Y_{1,t}^0 = \sum_{j=2}^J w_j Y_{jt} \tag{2}$$

The average treatment effect, δ, is found by subtracting the observed outcome for the treated unit Y_{1t} from the counterfactual outcome $Y_{1,t}^0$ in the post-treatment period:

$$\delta = Y_{1t} - Y_{1,t}^0 \tag{3}$$

For cases where the available control groups do not satisfy the parallel trends assumption, studies show that the SCM can provide less biased estimates than for example Differences-in-Differences (Kreif *et al.*, 2016, O'Neill *et al.*, 2016).

5.1.2 ESTIMATION: ELASTIC NET REGRESSION

There are various ways to choose weights and construct a relevant synthetic control group. The original SCM proposed by Abadie *et al.* (2015) chooses weights by minimising the distance between the pre-intervention outcomes of the treated and control groups[9]:

$$Q\left(\omega \middle| Y_{t,pre}^{obs}, Y_{c,pre}^{obs}\right) = \left\| Y_{t,pre}^{obs} - \omega^T Y_{c,pre}^{obs} \right\|_2^2 \tag{4}$$

This means that given $Y_{t,pre}^{obs}$ and $Y_{c,pre}^{obs}$ i.e. the observed outcomes in the pre-reform period for the treated, t, and control group, c, we want to estimate values for the weights, ω, such that the distance between the observed outcomes are minimised. It is important to note that the SCM places some constraints on the weights in order to find unique weights. As shown in *Paper I*, the SCM constraints makes it difficult to find relevant control groups for our male treated group and is thus not able to find a good pre-treatment fit. A reason for this is that imposing the "no-intercept" constraint, i.e. $\mu = 0$, will force the treated and control group to have same means, something that is inappropriate if the treated group is systematically smaller or larger than the control groups. Second, constraining these weights to equal to one ($\sum_{i=1}^J \omega_i = 1$), will inflate the slope of the synthetic control group larger than the actual data (Li, 2017). Doudchenko and Imbens (2016) argue that these constraints should be considered based on their merits, rather than being implemented as a matter of routine. They also show that these constraints may be relaxed, but by regularising the estimates, one can still find unique weights.

Therefore, *Paper I* follows the modified SCM approach (as proposed by Doudchenko and Imbens (2016)), where the "no-intercept" and "adding-up" constraints are relaxed, i.e. $\mu \neq 0$ and $\sum_{i=1}^N \omega_i \neq 1$. The counterfactual is then defined as the following:

$$Y_{1,t}^0 = \mu + \sum_{j=2}^J w_j Y_{jt} \tag{5}$$

---

[9] The equation assumes that only pre-intervention outcomes are used to construct the control group. Other covariates may also be included and used to construct a synthetic control group. See the papers by Abadie et.al. for more details.

This then allows for differences in mean between treated and control group. The objective function for constructing the synthetic control group in *Paper I* is thus:

$$Q\left(\mu, \omega \middle| Y_{t,pre}^{obs}, Y_{c,pre}^{obs}\right) = \left\|Y_{t,pre}^{obs} - \mu - \omega^T Y_{c,pre}^{obs}\right\|_2^2 + \lambda\left(\frac{1-\alpha}{2}\|\omega\|_2^2 + \alpha\|\omega\|_1\right) \tag{6}$$

The second part of the equation is the elastic net penalty term (developed by Zou and Hastie (2005)), where the parameter λ is the ratio of *l1* (Lasso) and *l2* (Ridge) type penalties and α represents the degree of regularisation. For data settings where the number of potential control units is larger than the number of time periods (in *Paper I:* 17 potential age groups and 4 years prior to reform), regularisation is suggested for estimating the weights to ensure that the minimisation procedure calculates unique weights (Athey and Imbens, 2017, Doudchenko and Imbens, 2016, O'Neill *et al.*, 2016). Regularising the estimates using both *l1* and *l2* penalties shrinks the coefficients to zero and thus minimises overfitting and variance. The elastic net penalty balances out the weaknesses and strengths of the two types of penalties. For example, for *Paper I*, the outcomes for the control groups are correlated. The *l1* type penalty will then only select one of variables in the groups that are correlated and may lead to a model that is too parsimonious. On the other hand, the *l2* penalty will keep all of the predictors (control groups) in the model.

5.2 Paper II

5.2.1 IDENTIFICATION

For *Paper II* several steps were taken in order to analyse the effect of changes in reimbursement rates on hospital activity. The first step was to identify a relevant treated group and outcome in order to observe a causal effect operating in relative isolation from threats of confounding (Keele, 2015). Since the analysis is not based on a natural experiment, the treatment assignment is less transparent. The treatment of interest here is whether increased reimbursement leads to increased activity. A DRG was identified as being financially favourable in year *t+1* if it experienced a price increase in year *t+1* compared to year *t*. Due to the lag in DRG price adjustments this means that for parts of year *t*, the DRG price was too low. Thus, the DRG in year *t+1* is financially favourable compared to itself in year *t*. This does not necessarily mean that the DRG is profitable compared to other DRGs. It may still be unprofitable in terms of reimbursement and marginal costs; however, it is more profitable relative to the year before. This treated group was then compared to earlier observations of itself and to financially unfavourable DRGs. Financially unfavourable DRGs are DRGs without price increases in year *t+1* compared to year *t*. Both financially favourable and unfavourable DRGs experienced increases in activity, but by comparing the two groups the causal effect is then the increase in activity for financially favourable DRGs relative to financially unfavourable DRGs.

Representing the causal relationship between DRG weights and hospital activity in a diagram can help determine which variables must be observed to estimate the size of the causal effect, and thus whether identification assumptions are reasonable (Morgan and Winship, 2015). Figure 1 in *Paper II* illustrated which factors might have affected both the reimbursement rates and the treatment volume and which were merely disturbances or noise. Changes in the administrative framework and technology were identified as potential confounders. The problem with confounders is that instead of estimating the effect of a price change, the model estimates the effect of both a price change and change in the administrative framework or technology on treatment volume. One way of controlling for confounding is to reduce the heterogeneity of the dataset by limiting the sample size to a smaller, more comparable and homogenous subset (Keele, 2015). In practice, this involved eliminating observations where administrative changes had taken place. By excluding these observations, the analysis evaluates DRGs for which there is no change in volume due to changes in the administrative framework. See *Paper II* for details on which observations were eliminated.

Secondly, there are factors that can affect the precision of the estimated causal effect because they affect the treatment volume. This was identified as being disease patterns and political priorities. Because they only affect the outcome - treatment volume, they do not confound the relationship between reimbursement and hospital activity. They do not then need to be controlled for to estimate a causal relationship (Morgan and Winship, 2015).

5.2.2 ESTIMATION

The main results were estimated using descriptive analysis. The average change in activity levels for the marginally favourable DRGs was compared to the marginally unfavourable DRGs. This was also compared across different sub groups. The difference in changes in activity levels between favourable and unfavourable DRGs was tested using a two-sample t-test for unequal variances.

To obtain elasticities and to assess which other factors might affect changes in activity levels, the analysis included three regression models. The simple model regressed yearly percent change in DRG weight on the yearly percent change in activity. This was expanded in the full model to include variables related to different types of DRGs, including dummy variables for whether the DRG was outpatient or inpatient, surgical or medical, and continuous variables representing the DRG weight in levels and number of treatments per DRG. In addition, time dummy variables were included to account for events in the years 2011, 2012 and 2013[10]. The last model included interactions between the change in DRG weight and the outpatient variable, medical variable and DRG weight.

---

[10] The time dummy for year 2008 was omitted due to collinearity

5.3 Paper III

5.3.1 IDENTIFICATION

The policy of increasing the price for day surgery was implemented for all day surgery DRGs and for all hospitals and regional health authorities reimbursed as part of the ABF scheme. The policy has a clear treatment assignment and an exogenous increase in the cause of interest (i.e. the price). However, the existence of a relevant control group as a counterfactual was less clear. Again, the methodological approach involved identifying an outcome and subset that would ensure that we were comparing similar observations. To do this we used within-pair comparisons; the sample size was reduced to DRGs where there existed both an inpatient and day surgery DRG code for a surgical procedure. By analysing the probability of day surgery within a DRG-pair, the method reduces potential confounding in the form of changes in the ABF share, changes in the DRG unit price and other factors related to treatment of specific diagnoses that may affect overall surgery rates. In addition to a within-DRG comparison, the outcome for each DRG was compared to previous years, thus previous outcomes for the DRGs were used as a counterfactual for the outcomes in the treatment period.

5.3.2 ESTIMATION: PARTIAL POOLING AND BAYESIAN MODELLING

Due to the data structure, the model in *Paper III* was estimated using multilevel methods. For *Paper III*, the data were structured in such a way that the values of some of the parameters depend on the values of other parameters (Kruschke, 2015). The dependency results from observations being correlated within groups and across time. This means that the observations for each DRG are correlated over time and the probability of observing day surgery for a DRG is related to DRG specific properties, such as type of surgical procedure. In addition, the probability of observing day surgery may be related to regional properties, such as availability of resources, which affect all DRGs similarly within a region. This type of data structure may be referred to as multilevel or hierarchical structures. There are several different ways to write and estimate multilevel models and Gelman and Hill (2007) provide a concise overview. The motivation and estimation procedure for the model in *Paper III* uses the partial pooling framework. The partial pooling method can be seen as a compromise between a complete-pooling model and a no-pooling model. With complete-pooling, the model assumes there is no systematic variation or correlation between the observations and it just estimates an overall mean.

$$y_{ij} = a + X_{ij}B + \epsilon_{ij} \tag{7}$$

Not considering the correlation in the complete-pooling model means that it may overestimate the precision of the regression estimates; each new observation contains less information than if the sample were completely random and not clustered in groups (Angrist and Pischke, 2015). The

effective sample size for calculating the standard errors is a lot smaller when the observations are not completely independent. On the other hand, a no-pooling model estimates separate intercepts α_j for each group and thus may overstate the variation between the groups. It assumes that the groups are independent and that no information from other groups affects the estimate for group j.

$$y_{ij} = a_j + X_{ij}B + \epsilon_{ij} \tag{8}$$

Partial pooling takes into account that the groups are similar in some ways and shares information that exists both within and between groups. The partial pooling model is thus similar to eq. 8, but instead of estimating a separate α for each group, j, the α_j are assumed to come from a distribution $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$ and can be thought of as a weighted average of no-pooling estimate and pooled estimate mean. Groups with a larger number of observations contribute more towards the overall mean, whereas groups with fewer observations contribute less and their mean is pulled closer to the overall mean. Thus, the model is useful for performing inference for groups with small sample sizes.

The model was estimated in a Bayesian framework using Markov Chain Monte Carlo (MCMC) algorithms. Following is a brief description of the intuition behind Bayesian analysis and MCMC. This section will end with a brief note on why this approach was selected.

Bayesian statistical analysis is based on conditional probabilities expressed using Bayes formula:
$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} \tag{9}$$

The outcome of the Bayesian model is the posterior distribution, which represents the conditional probability of the parameter of interest given the data $P(\theta|x)$. It is essentially a weighted combination of prior information $P(\theta)$ about the true value for the parameter and the likelihood $P(x|\theta)$, which represents the process in which the data are generated or distributed, given the parameter values. The prior distribution can reflect previous information about the parameter or, if little prior information exists, a noninformative prior can be chosen.

Paper III is thus analysing what the probability of day surgery is, given the observed data. This is estimated using the assumption that the data on day surgery admissions are generated by a Bernoulli distribution ($P(x|\theta)$) and that the estimate for day surgery probability lies between 0 and 1, but follows a beta distribution $P(\theta)$. $P(x)$ is a normalising constant ensuring that the posterior distribution is true. Calculating $P(x)$ involves integrating over all possible parameter values and for continuous data, this is not possible to solve analytically[11]. Nevertheless, MCMC algorithms can approximate the posterior distribution. This is done by randomly proposing a

---

[11] The theorem can be solved analytically if the prior distribution and the posterior distribution are conjugate.

posterior value for the parameter and then jumping to a new value. The jump to the new proposed value is accepted if this explains the data better than the current value. This is evaluated by calculating the ratio of the proposed posterior distribution to the current posterior distribution. If this is greater than one, then the proposed posterior distribution is larger compared to the current posterior distribution. The equation below shows how this leads to the $P(x)$ being cancelled out. This process is done iteratively until the posterior distribution converges.[12]

$$\frac{\frac{P(x|\theta_p)P(\theta_p)}{P(x)}}{\frac{P(x|\theta_c)P(\theta_c)}{P(x)}} = \frac{P(x|\theta_p)P(\theta_p)}{P(x|\theta_c)P(\theta_c)} \tag{10}$$

The NUTS (No U-Turn Sampler) algorithm used in Pystan works in a similar way, but has a smarter and more efficient method of suggesting proposed values. It builds a set of likely values that spans a wide swath of the target distribution, stopping automatically when it starts to double back and retrace its steps (Hoffman and Gelman, 2014).

Hierarchical models are in general estimated using either Bayesian or maximum likelihood methods. The choice between the two methods is usually made based on matters of convenience or philosophical preferences. For multilevel logit models, there seems to be no difference in the estimates between the two, especially for panel data (Elshiewy *et al.*, 2017, Huber and Train, 2001). Nevertheless, maximum likelihood approximations sometimes have difficulties in finding the right starting values and this may lead to convergence problems. The estimation of the model for *Paper III* in STATA 14 using the *melogit* command (which uses maximum likelihood) resulted in poor convergence. In addition, a method that could incorporate three levels (year, DRG and regional levels) was needed. The Pystan software, which uses the Bayesian framework, gave this flexibility. Therefore, the model was estimated using the MCMC and Bayesian framework instead of maximum likelihood methods.

---

[12] Thomas Wiecki (2015) provides a more complete explanation, which can be found here
http://twiecki.github.io/blog/2015/11/10/mcmc-sampling/

6 Summary of Results

6.1 Paper I

Did Adolescents in Norway Respond to the Elimination of Co-Payments for General Practitioner Services?

Paper I aimed to evaluate to what extent adolescents are sensitive to co-payments for GP consultations. The paper exploits a policy reform in 2010 where adolescents between the ages of 12 to 15 were exempted from a co-payment of 17.5 Euro. Aggregate data on the number of GP consultations for males and females from 2006 to 2013 was obtained from the Norwegian Health Economics Administration (HELFO). The effect was estimated using a modified Synthetic Control Method to compare GP visits among the treated age groups to relevant control age groups.

The results suggest that adolescents were sensitive to having to pay a fee for visiting their primary care physician and that exempting them from this fee increased their use of health care services by 22.1% and 13.8% among females and males respectively. There are approximately 120,000 adolescents of each gender in the age bracket 12 – 15 years in a given year; thus, effect estimates of 22.1% and 13.8% will result in an average of 24,000 and 13,500 extra GP visits for girls and boys respectively per year. Adolescent females were more responsive to the co-payment reform, suggesting that for this age group, prior to the reform; girls were more vulnerable than boys were to the effect of co-payments.

6.2 Paper II

Did Hospitals Respond to Changes in Weights of Diagnosis Related Groups in Norway between 2006 and 2013?

The aim of *Paper II* was to test to what extent hospitals focus on the diagnostic groups that are most financially favourable given costs and reimbursement rates. DRGs were defined as financially favourable if they received an increase in reimbursement one year, on the assumption that the reimbursement for a DRG had been too low for parts of the previous year. The net impact would be that the DRG is reimbursed more favourably for parts of the current year than for parts of the previous year. In order to isolate the causal effects, the dataset was restricted to a sample that was not affected by potential confounders. The main source of confounding in this context is administrative changes that result in changes in volume, but are not the result of changes in reimbursement. In addition, the robustness of the results was evaluated across different subgroups.

For the main sample, activity increased by 3.16% for the DRGs with increased reimbursement, and increased by 0.74% on average for DRGs with decreased reimbursement. Thus, the percentage increase in hospital admissions was on average four times higher when the reimbursement was increased, relative to the percentage change for DRGs with reduced rates. The direction of the result was consistent across different time periods and sub-groups such as surgical, medical, and inpatient DRGs. The effect was smaller, but remained significant after eliminating DRGs that were most likely to be affected by up-coding. The average elasticity of response to a 10% change in price was 0.087.

6.3 Paper III

Hospitals' Response to Changes in Reimbursement for Day Surgery: Evidence from Norway

The background for the study is a policy that was introduced in Norway in 2015, which aimed to incentivise the hospitals to provide more day surgery procedures in an effort to increase quality and efficiency. The incentive involved an exogenous 10% increase in the reimbursement hospitals receive for day surgery admissions combined with a 1.4% decrease in reimbursement for inpatient procedures. This resulted in a change in the relative price for procedures where both day and inpatient surgeries are options. The objective was to estimate to what degree hospitals responded to the price increase by increasing the probability of performing day surgery, and thus to what extent the policy was successful in achieving its objective of stimulating day surgery in Norwegian hospitals.

The annual change in the probability of day surgery is estimated using a random effects model in a Bayesian framework that accounts for the hierarchical nature of our data. The results show that on average the relative odds of day surgery admissions in the post-policy period, were 1.05 and 1.09 larger compared to the start of the period (2011). However, there was also increase in day surgery activity of almost the same magnitude in 2014 and 2013. While the data is consistent with a shift towards day surgeries following the reform, the lack of a clear break in the time trend means that the results provide only weak evidence of an incentive-driven shift in how cases are allocated between day and inpatient surgery.

7 Discussion

The following sections will discuss the results in relation to the existing empirical literature, provide some thoughts on policy implications and suggest avenues for further research. In addition, for each paper an assessment is made as to what extent the estimated effects are justified. The identification and estimation methods chosen for each paper were described in chapter 5, however a discussion of their assumptions and limitations and to what extent a causal relationship could be established is still warranted (Keele and Minozzi, 2013). Because *Paper*s *II* and *III* both estimate incentives within the activity-based financing system for hospitals these are discussed together.

7.1 Paper I

Results

Paper I estimated a considerable degree of responsiveness in terms of GP visits following exemption from co-payments. This result indicates that co-payments were an important determinant of health care use for this age group. In general, the result confirms the results from other studies on co-payments and adolescents or children, at least in the direction of the effect. However, the estimated effect in the Norwegian experience is larger compared to effects estimated in other countries (Kato and Goto, 2017, Paul and Nilsson, 2014, Votapkova and Zilova, 2015, Zápal, 2010).

A potential reason for the large effect is the size of the incentive coupled with the age of the targeted group and how it relates to other barriers to access. The size of the co-payment reduction may be large relative to an adolescent's personal budget constraint, and this constraining factor may be accentuated by the consideration that this is an age group with a growing need for private consultations without parent involvement compared to younger children. Several studies highlight that confidentiality concerns are a barrier to access for adolescents (Gleeson *et al.*, 2002, Tylee *et al.*, 2007). The co-payment exemption reduces adolescent financial dependency and thus enables them to see a GP without parents being involved. This is also evident in the Norwegian co-payment study on 16 year olds, where a co-payment reduced their use of GP services, suggesting that adolescents are particularly sensitive to paying for health care services (Landsem and Magnussen, 2018).

Furthermore, it is important to consider how the institutional setting may affect the results and relevant factors include the existence and degree of access to other health care services. For example, Norway has a gatekeeping system and a visit to the GP is necessary for access to prescription drugs and secondary care services; especially outpatient services at hospitals

(Ringard *et al.*, 2013). Countries such as the Czech Republic and Japan do not require referrals from the GP for use of these services and can thus receive health care elsewhere (Alexa *et al.*, 2015, Sakamoto *et al.*, 2018). Since Norwegian GPs are gatekeepers, the demand for their services may be larger than in other countries.

Related to this is the degree of access to potential substitute services such as school nurses and public health nurses at health clinics. Availability of alternative health care services was found to have been an important determinant of price elasticity of demand in several empirical studies (see (Skriabikova *et al.*, 2010)). Although access to these services is limited (due to few nurses and limited opening hours), and they are not perfect substitutes for GP services, the existence of these services may reduce the demand for some GP visits and reduce the effect of the co-payment reform. However, there is currently no data available in Norway on adolescent use of school or public health clinic services, and thus it is difficult to evaluate to what extent there are any spillover effects between GP services and health clinics.

As pointed out in the literature review in chapter 3, few studies have estimated gender differences in response to cost sharing, and there is currently no consensus on whether males or females are more responsive. *Paper I* estimated a larger effect for females, and this is consistent with the US based study on co-payments in a health care maintenance organisation (Cherkin *et al.*, 1989). For adults, a survey from New Zealand reveals that women are more likely than men to defer primary health care because of cost regardless of individual deprivation or income levels (Jatrana and Crampton, 2012). Nevertheless, it is difficult to state a priori why female adolescents in general should be more responsive to changes in co-payments and to what extent female and male adolescents differ in their preferences for spending money. *Paper I* discusses the possibility of confidentiality being of greater concern to female adolescents compared to males. Another possible reason could be differences in why the adolescents visit the GP, and a lower responsiveness to co-payments could be explained by males requiring services for more urgent issues. However, the data for *Paper I* did not include any information on diagnosis.

Can we trust the causal estimate?

For *Paper I*, there were a large number of available control groups, which led to the question of which age group was most relevant as a control group. Although SCM provides a data driven method of selecting and matching appropriate control units, it is still based on assumptions that cannot be tested. The identification assumption for SCM is: independence conditional on similar levels in past outcomes (O'Neill *et al.*, 2016), so control groups that are similar to the treated group in terms of the past outcomes will accurately provide an estimate of the outcome for the treated group, if they had not been exposed to treatment. One way of ensuring similarity of the

groups was to reduce the pool of potential groups and this is described in detail in *Paper I*. In addition, SCM also allows for the inclusion of other covariates that have predictive power to match the treated and control groups. Since the data is analysed at the age level, there is no need to adjust for individual level differences that might affect demand for GP services, such as income, socioeconomic status and health status. At the age level, the overall health status of each age group could be an important determinant of GP services. Data on health status at age level was not included in the analysis. In most cases the weighting of the control groups is driven by the lagged outcomes rather than matching on covariates, and excluding covariates is rarely important (Athey and Imbens, 2017). Excluding health status data is only a problem if it would have led to a different weighting of the potential control groups.

A limitation of the SCM to bear in mind is that it is sensitive to the effects of idiosyncratic shocks, especially with few time periods (Ferman and Pinto, 2016, Kreif *et al.*, 2016). This means that with few time periods, the method might construct the synthetic control groups from units that appear to be similar and thus have similar unobserved characteristics in the pre-treatment period, but then are not similar in the post-treatment period. There were only 4 years prior to the reform available for the application of the SCM in *Paper I*; however, it is not possible to test to what extent this is a limitation in *Paper I*.

Placebo analyses

To strengthen the story about the causal relationship and to assess the credibility of the analyses one can undertake sensitivity or placebo analysis (Athey and Imbens, 2017). For *Paper I*, a placebo analysis was performed as suggested by Abadie et.al. (2010, 2015). The analysis was applied iteratively to each control group in the donor pool and the distribution of resulting effects was plotted and compared to the effect estimated for the treated group. Ideally, no other age group would have an estimate as large as the one observed for the treated group. Groups with a poor pre-treatment fit were excluded, resulting in few groups with which to compare the effect. This limits the inference available through the placebo analysis since the resulting p-value is a function of the number of control groups. This is a general limitation of the SCM; asymptotic inference and standard t-tests are not possible to estimate, thus placebo methods,[13] as described above, are necessary for inference. However, a sufficient number of control units for estimating p-values is often not available. More research is needed on alternative methods for inference under SCM.

---

[13] Also referred to as permutation tests in the literature

Policy implications and further research

There is a current debate in Norway as to whether a similar co-payment exemption should be implemented for 16 to 18-year old adolescents as well[14]. The results from *Paper I* indicate that adolescents are sensitive to co-payments and that one can expect an increase in demand for GP services for the older age groups as well following a similar co-payment exemption. A previous assessment of the co-payment reform assessed the impact on the individual age groups, and found that 15 year olds seemed to be more responsive than 12 year olds (e.g. 24.5% vs 13.1% respectively for females) (Olsen and Melberg, 2016). If this trend in age response can be extrapolated, then any future policy recommendation regarding exempting older adolescents from co-payments should also consider to what extent they may be more responsive than the age group evaluated in *Paper I*. Concerning gender differences: if a greater response amongst female adolescents can be interpreted as a greater success of the policy reform, then the smaller response amongst males suggests that other measures are necessary for increasing their use of GP services.

A related discussion is the existence, magnitude, and nature of the moral hazard response, which is claimed to be a key input into the optimal design of private or public health insurance contracts (Einav and Finkelstein, 2017). The theoretical literature claims that under full health insurance coverage (which would be case when adolescents are exempted from co-payments) some inefficiency exists due to moral hazard (Arrow, 1963, Pauly, 1968). However, this literature has been based on the assumption that the demand curve correctly reflects the marginal benefits of services and that patients are fully-informed (Ellis and McGuire, 1993). Thus to assess to what degree exempting adolescents from co-payments and thus receiving health care for free[15] has resulted in inefficient use of resources, requires an assessment of to what degree the level of health care use prior to the reform was optimal for adolescents. The point here is that even though the co-payment exemption led to a large increase in use of GP services, it is difficult to assess to what degree this has led to more inefficient use of GP services in terms of increase in demand for unnecessary services.

The study by Landsem and Magnussen (2018) on the effect of co-payments on 16 year olds attempts to address the issue of unnecessary consumption by evaluating to what extent the co-payment reduced admission for different types of diagnoses. The results indicate that the co-payment reduces visits for patients with "general complaints" – i.e. what they term as unnecessary services, but it also reduces visits for necessary services for chronic and psychological diagnoses. Although this indicates that the co-payment prior to the reform for 12 to 15 year olds

---

[14] see e.g. https://www.stortinget.no/no/Saker-og-publikasjoner/Sporsmal/Sporretimesporsmal/et-sporretimesporsmal/?qid=67809
[15] Free in terms of monetary costs; there will always be some degree of travel and time costs.

was inefficient at targeting solely unnecessary services, a limitation is that it is difficult to define what a medically unnecessary visit is.

Alternatively, one can assess to what degree exemption from co-payments led to health improvements for adolescents. The RAND experiment showed no detrimental effects on health status for children due to increased cost-sharing (Valdez *et al.*, 1985). The study in Japan on medical subsidy disqualification for 10 year olds argued that this resulted in reduced outpatient visits for mild or chronic conditions, however, this did not necessarily lead to negative health outcomes in the short term (Miyawaki *et al.*, 2017). However, if the above results by Landsem and Magnussen are true - that co-payments reduce visits for necessary services, then it would be interesting to evaluate health improvements following the co-payment exemption. Thus, an important avenue for further research would be to assess to what extent increased access to primary care has both short and long-term health effects. If differences between males and females also persist in terms of health improvements, this will help assess to what degree other measures are necessary for increased visits amongst males.

7.2 Papers II and III

Results

Paper II used price changes in the DRG prospective payment system to identify DRGs where the reimbursement exceeds the marginal cost compared to the year before. Theory predicts that hospitals have incentives to admit more patients where this is the case (Ellis and McGuire, 1986). The results confirm that hospitals may increase treatment volume for these DRGs, although of slightly different magnitude compared to other studies. The elasticity estimated in *Paper II* (0.087) is larger than the elasticity of 0.049[16] estimated in the Norwegian study by Januleviciute *et al.* (2015). One potential reason for the difference in size of the effect between the two Norwegian studies is the time periods in which they assess the effects. Prior to 2005, the financing of Norwegian hospitals involved elements of retrospective cost reimbursement, due to soft budget constraints (Tjerbo and Hagen, 2018). Thus, hospitals had reduced incentives to focus on profitable DRGs to increase revenue.

The Italian study also showed increases in treatment volume for a one time change in DRG prices, however, they found a much larger effect for surgical DRGs compared to medical DRGs; the estimated elasticity was 1.7 vs -0.24 (Verzulli *et al.*, 2016). The elasticity of price was found to be larger for medical DRGs in the paper by Januleviciute *et al.* (elasticities between 0.079 and 0.129). *Paper II* also estimates larger price elasticity for medical DRGs compared to surgical DRGs, however, this effect was not significant. The argument for the effect differences in the

---

[16] This result was incorrectly reported in *Paper* II. It reported that the elasticity was 0.094 and not 0.049.

Norwegian studies is that hospitals can exercise more discretion in terms of admitting patients for medical treatment compared to surgical treatment. Verzulli *et al.* (2016) on the other hand argues that the hospitals had a higher proportion of occupied beds for the medical diagnoses prior to the tariff increase and had less response for these DRGs.

Given that hospitals seem responsive when prices increase, and the prevailing argument is that prices can be used to influence hospital behaviour, the hypothesis for *Paper III* was that the 10% increase in price for day surgery would lead to an increase in day surgery provision. However, there was no significant increase. This suggests that the increase in price for day surgery, compared to its price the previous year, was less important to hospital decisions than the relative prices between day and inpatient surgery. This conclusion is further strengthened when comparing to the UK experience. The UK has also implemented financial incentives within their prospective payment system to stimulate day surgery and this was evaluated in a study by Allen *et al.* (2016). A 24% increase in the tariff for cholecystectomy day surgery led to a 6% point increase in activity. The main difference between this incentive and the Norwegian incentive was that surgical DRGs in the UK were not refined into a day surgery and inpatient surgery DRG. In effect, the day surgery tariff was increased by 24% compared to inpatient surgery. The Norwegian incentive, however, only reduced the difference between day and inpatient surgery DRGs by approximately 10%[17]. Thus, policy makers may still use changes in DRG prices to pay for "best practice", but relative prices between substitutable DRGs need to be taken into account.

Can we trust the causal estimate?

In contrast to the design-based method of *Paper I*, for *Paper II* one cannot assume that all confounding factors are accounted for based on the design and method. A large part of the method involved ensuring that the observed confounders in the data set were eliminated (see chapter 5). However, there may still be unobserved confounders that could affect reimbursement levels and treatment volume. The paper discusses the degree to which unobserved confounding due to technology may be of concern, but in all scenarios it shows that the estimated effects may be conservative rather than overestimated.

Similar to *Paper I*, *Paper III* uses a policy intervention as a source of exogenous change for the cause of interest. However, the identification and estimation is based on a before/after analysis, comparing the outcome in the years prior to the intervention with the outcome in the years after the intervention. An assumption of this approach is that there are no unobserved variables that also affect the outcome at the same time. It does not necessarily separate the effect of the treatment from any underlying trend of other factors that happen at the same time as the

---

[17] Taking into account the yearly DRG adjustments, the price increase was only 5% on average.

intervention. This is partially accounted for by using DRG pairs (as described in chapter 5), however the analysis could have been strengthened by applying an interrupted time series (ITS) approach to account for prior trends. An ITS uses the trend in the outcome prior to treatment as a counterfactual for the outcome in the absence of treatment (Morgan and Winship, 2015). However, for *Paper III* there was not enough time periods to establish a time trend and doing so might have resulted in not enough power to detect any effect (Zhang *et al.*, 2011).

There may be several reasons why *Paper III* doesn't estimate a significant effect of the policy, but the main concern is that the 10% increase was not big enough compared to other changes happening with the DRG weights both prior to and during the reform. From a policy evaluation perspective, it would have been better if the reform had been implemented differently. For example, different day surgery DRG codes could have received different levels of the incentive and thus one could assess how large the incentives would have to be in order to give an effect.

Placebo analyses

To check the credibility of the results for *Paper III*, the effect of the cause could have been assessed on outcomes that should not be affected by the cause (Morgan and Winship, 2015). For example, since the policy targeted surgical DRGs, one could potentially assess if there was a similar increase in medical day treatment DRGs. However, this was not estimated in the paper because there are very few DRGs that differentiate between a medical day treatment and inpatient treatment DRG. Thus, assessing the change in proportion would have been difficult. Another robustness check could have been to assess to what the extent the causal effect varies across subgroups in predictable ways (Morgan and Winship, 2015). The policy targeted all day surgery DRGs, however, there could be subgroup differences as to the extent to which the policy should have an effect. One such difference could have been the proportion of the surgery performed as day surgery prior to the policy year. For example, a large proportion performed as day surgery could indicate that there is less capacity for performing more; however, this does not necessarily mean that for DRGs where a small proportion is performed as day surgery would experience a larger increase. Instead, these DRGs might be indicative of DRGs that are not suitable for increases in day surgery activity.

For *Paper II*, several subgroup robustness checks were conducted. This involved excluding DRGs that were more susceptible to up coding, and checking whether the results held for the remaining DRGs. In addition to excluding these DRGs, the analysis was performed on surgical DRGs for which there should be no effect of up coding and only an effect of a change in reimbursement, thus assessing the effect on multiple outcomes that should be affected by the cause (Morgan and Winship, 2015). The results seemed to hold for these DRGs as well. To assess the effect on

outcomes that should not have been affected by changes in reimbursement, the analysis was performed on the different main diagnostic disease categories. The results were as expected for the pregnancy/birth and new born/neonatal categories, since the hospital has less discretion over treatment volume with respect to births. This analysis was extended to other main diagnostic categories, although it is difficult to argue a priori to what extent one might expect an effect in either direction for the remaining groups.

Policy implications and further research

The activity-based financing system in Norway is intended to be neutral and not provide incentives for prioritising between DRGs or patients. This implies that only medical considerations should steer priorities and provision of hospital treatments (Norwegian Directorate of Health, 2017). The point of departure for *Paper II* and related studies is that DRG prices reflect average costs, but when DRGs are lagged, resulting in changes in profitability between DRGs, then the system is not neutral and affects how hospitals prioritise between DRGs. There are two sides to this, which should be taken into account by policy-makers. On the one hand, the observed response to changes in prices in *Paper II* may have unintended consequences if hospitals increase activity for the financially favourable DRGs and crowds out treatment for other DRGs. This in turn may have adverse impacts on access to care for patients in the financially unfavourable DRGs. More research is needed on to what extent changes in such allocation decisions are efficient and do not lead to unintended consequences in terms of relevant patient outcomes and quality. The systematic review by Palmer *et al.* (2014) investigates effects on patient outcomes, but focuses on comparing the effect of introducing ABF with alternative funding systems. This has not been systematically assessed in the studies on price changes within existing ABF systems, apart from Verzulli *et al.* (2016) examining the effect of a price change on patient access in terms of waiting times. They found no significant effect. On the other hand, by focusing on generating revenue the hospitals may be prioritising, so that more patients may receive treatment in the long-run. In addition, the paper shows that reimbursement fees may serve as a tool to steer prioritization and achieve health policy aims.

The price change evaluated in *Paper III* could be viewed as an attempt to signal that more day surgery is desirable and thus steer priorities for the hospitals. By letting the price reflect best practice, instead of national average costs, hospitals have incentives to improve performance, quality and supply a more optimal mix of services (Or and Häkkinen, 2011, Street and Maynard, 2007). Thus, from the trade-off perspective, this financial incentive had the advantage of both reducing costs and increasing efficiency while at the same time not being detrimental to health outcomes (Andersen and Jensen, 1993, Majholm *et al.*, 2012, Martinussen and Midttun, 2004). Although Norwegian hospitals are in general responsive to price changes (as evident in *Paper II*

and (Januleviciute *et al.*, 2015)), the results of *Paper III* suggest that future similar policies need to take into account how large the incentive is compared to the price of alternative treatment options. The results of the analysis are relevant for both policy-makers and researchers who are interested in the further development of the prospective reimbursement systems employed in many countries, however, more research is needed on how paying for best practice is best implemented for optimal results.

8 Concluding Remarks

Health care policy-making attempts to achieve multiple objectives at once and financial incentives are among the instruments that are often used to help achieve these goals. For example, in Norway, the financing systems aim to support the following health policy objectives: (1) create equal access to a good health care service, (2) stimulate efficient resource use and (3) provide the basis for cost containment (Ministry of Health and Care Services, 1995). All papers provide evidence of the extent to which patients and providers increase use and provision of health care services in response to changes in financial incentives, and thus indicate how financial incentives affect prioritisation and allocation of resources. *Paper*s *I* and *III* evaluated policy reforms that aimed to redistribute use and provision of health care services to achieve policy goals of access and cost efficiency. *Paper II*, on the other hand, examined existing incentives in the activity-based financing system, showing that these incentives also affect prioritisation decisions made by hospitals.

The papers in this thesis provide evidence from a Norwegian health care setting. Since there are some differences between the Norwegian and other health care systems (described and discussed in chapters 2 and 3), the size of the effects estimated are valid only for local Norwegian policy-making. Nevertheless, the papers still contribute to a growing empirical literature on the effects of demand and supply-side cost sharing across variations in contexts such as population, settings, treatments and outcomes (Shadish *et al.*, 2002) and provide additional contexts or mechanisms that might explain the variation in response (Morgan and Winship, 2015). Thus, *Paper I* demonstrated that adolescents are also responsive to changes in co-payments. *Paper II* demonstrated that hospitals respond to price changes by also changing activity levels and not just by up-coding, thereby providing evidence on changes in actual priority given to different DRGs. And, *Paper III* demonstrated the importance of the size of the incentive to stimulate provision of day surgery.

From a policy perspective, this thesis provides a piece of the puzzle in terms of providing evidence to balance the trade-offs in using financing to obtain health care objectives. The focus of the thesis has been to estimate responsiveness – and the results show that financial incentives can be used to steer behaviour. Even so, more research is needed on the extent to which the observed responses, especially for *Paper*s *I* and *II*, affected health outcomes and efficient use of resources.

Lastly, concerning the estimation of causal effects, each paper differed in terms of data available and the existence of exogenous variation and control groups, which in turn has determined the method that is most appropriate to use to estimate an unbiased effect. For each method an assessment was made as to what extent the estimated effects are justified. However, for future

research, it would be helpful if policy-makers designed policy implementations in such a way that effects may be found in observational data. The discussion highlighted ways in which additional insights could be gained from comparing the results to a control group. This is also recommended by a recent white paper on priority setting in Norway, which seeks more evidence on how financing systems affect behaviour in health care, stemming from policy interventions that are designed and implemented in such a way as to capture informative effects (NOU 2014:12, 2014).

9 References

2001. Lov om spesialisthelsetjenesten m.m. [Specialized Health Services Act §2-1]. Norway.

Abadie A, Diamond A & Hainmueller J. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association,* 105**,** 493-505.

Abadie A, Diamond A & Hainmueller J. 2015. Comparative politics and the synthetic control method. *American Journal of Political Science,* 59**,** 495-510.

Abadie A & Gardeazabal J. 2003. The economic costs of conflict: A case study of the Basque Country. *The American Economic Review,* 93**,** 113-132.

Alexa J, Rečka L, Votápková J*, et al.* 2015. Czech Republic: Health System Review 2015. *Health Systems in Transition,* 17(1)**,** 1-165.

Allen T, Fichera E & Sutton M. 2016. Can payers use prices to improve quality? Evidence from English hospitals. *Health economics,* 25**,** 56-70.

Andersen S & Jensen M. 1993. [Ambulatory anesthesia. An inquiry study of patient satisfaction and therapeutic quality]. *Ugeskrift for laeger,* 155**,** 2803-2806.

Angrist J D & Pischke J-S. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives,* 24**,** 3-30.

Angrist J D & Pischke J-S. 2015. *Mastering 'Metrics: The Path from Cause to Effect,* New Jersey, Princeton University Press.

Anthun K S, Bjørngaard J H & Magnussen J. 2017. Economic incentives and diagnostic coding in a public health care system. *International Journal of Health Economics and Management,* 17**,** 83-101.

Arrow K J. 1963. Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review,* 53**,** 941-973.

Athey S & Imbens G W. 2017. The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives,* 31**,** 3-32.

Barros P & Braun G. 2017. Upcoding in a national health service: The evidence from Portugal. *Health economics,* 26**,** 600-618.

Barros P & Martinez-Giralt X. 2012. *Health Economics: An Industrial Organization Perspective,* UK, Routledge.

Barros P P & Siciliani L. 2011. Chapter Fifteen - Public and Private Sector Interface. *In:* Pauly, M V, Mcguire, T G & Barros, P P (eds.) *Handbook of Health Economics.* Elsevier.

Biørn E, Hagen T, Iversen T*, et al.* 2003. The Effect of Activity-Based Financing on Hospital Efficiency: A Panel Data Analysis of DEA Efficiency Scores 1992–2000. *Health Care Management Science,* 6**,** 271-283.

Cherkin D C, Grothaus L & Wagner E H. 1989. The Effect of Office Visit Copayments on Utilization in a Health Maintenance Organization. *Medical Care,* 27**,** 669-679.

Chiappori P-A, Durand F & Geoffard P-Y. 1998. Moral hazard and the demand for physician services: First lessons from a French natural experiment. *European Economic Review,* 42**,** 499-511.

Cockx B & Brasseur C. 2003. The demand for physician services: Evidence from a natural experiment. *Journal of Health Economics,* 22**,** 881-913.

Cots F, Chiarello P, Salvador X*, et al.* 2011. DRG-based hospital payment: Intended and unintended consequences. *In:* Reinhard Busse, Alexander Geissler, Wilm Quentin & Wiley, M (eds.) *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals.* Mc Graw Hill.

Cutler D M & Zeckhauser R J. 2000. The Anatomy of Health Insurance. *In:* Culyer, A J & Newhouse, J P (eds.) *Handbook of Health Economics.* Elsevier.

Dafny L S. 2005. How do hospitals respond to price changes? *The American Economic Review,* 95**,** 1525-1547.

Doudchenko N & Imbens G W. 2016. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *National Bureau of Economic Research,* Working Paper:22791.

Dranove D. 1987. Rate-Setting by Diagnosis Related Groups and Hospital Specialization. *The RAND Journal of Economics,* 18, 417-427.

Ehrlich I & Becker G S. 1972. Market Insurance, Self-Insurance, and Self-Protection. *Journal of Political Economy,* 80, 623-648.

Einav L & Finkelstein A. 2017. Moral Hazard in Health Insurance: What We Know and How We Know It. National Bureau of Economic Research.

Ellis R P. 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics,* 17, 537-555.

Ellis R P & McGuire T G. 1986. Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of health economics,* 5, 129-151.

Ellis R P & McGuire T G. 1993. Supply-Side and Demand-Side Cost Sharing in Health Care. *Journal of Economic Perspectives,* 7, 135-151.

Ellis R P & McGuire T G. 1996. Hospital response to prospective payment: moral hazard, selection, and practice-style effects. *Journal of health economics,* 15, 257-277.

Elshiewy O, Zenetti G & Boztug Y. 2017. Differences Between Classical and Bayesian Estimates for Mixed Logit Models: A Replication Study. *Journal of Applied Econometrics,* 32, 470-476.

Farbmacher H & Winter J. 2013. Per-Period Co-Payments And The Demand For Health Care: Evidence From Survey And Claims Data. *Health economics,* 22, 1111-1123.

Ferman B & Pinto C. 2016. Revisiting the synthetic control estimator. *Munich Personal RePEc Archive,* No. 75128.

Finkelstein A, Taubman S, Wright B, *et al.* 2012. The Oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics,* 127, 1057-1106.

Flodgren G, Eccles M P, Shepperd S, *et al.* 2011. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *The Cochrane Library.*

Freeman J D, Kadiyala S, Bell J F, *et al.* 2008. The Causal Effect of Health Insurance on Utilization and Outcomes in Adults: A Systematic Review of US Studies. *Medical Care,* 46, 1023-1032.

Gelman A & Hill J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models,* New York, Cambridge University Press.

Gilman B H. 2000. Hospital response to DRG refinements: the impact of multiple reimbursement incentives on inpatient length of stay. *Health economics,* 9, 277-294.

Gleeson C R, Robinson M B & Neal R D. 2002. A review of teenagers' perceived needs and access to primary health care: implications for health services. *Primary Health Care Research & Development,* 3, 184-193.

Hafsteinsdottir E J G & Siciliani L. 2010. DRG prospective payment systems: refine or not refine? *Health economics,* 19, 1226-1239.

Hagen T P, Holom G H & Amayu K N. 2017. Outsourcing day surgery to private for-profit hospitals: the price effects of competitive tendering. *Health Economics, Policy and Law,* 13, 50-67.

Hagen T P & Kaarbøe O M. 2006. The Norwegian hospital reform of 2002: Central government takes over ownership of public hospitals. *Health Policy,* 76, 320-333.

Hagen T P, Veenstra M & Stavem K. 2006. Efficiency and patient satisfaction in Norwegian hospitals. *Health Organization Research Norway (HORN),* 2006.

Heckman J J. 2008. Econometric Causality. *International Statistical Review,* 76, 1-27.

Hodgkin D & McGuire T G. 1994. Payment levels and hospital response to prospective payment. *Journal of Health Economics,* 13, 1-29.

Hoffman M D & Gelman A. 2014. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research,* 15, 1593-1623.

Holland P W. 1986. Statistics and Causal Inference. *Journal of the American Statistical Association,* 81, 945-960.

Huber J & Train K. 2001. On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths. *Marketing Letters,* 12, 259-269.

Huitfeldt I M S. 2017. *Universal Health Care Supply: Queues, Prices and Regional Variation.* Ph.D., University of Oslo.

Jakobsen M L F. 2010. The Effects of New Public Management: Activity-based Reimbursement and Efficiency in the Scandinavian Hospital Sectors. *Scandinavian Political Studies,* 33, 113-134.

Jakobsson N & Svensson M. 2016. Copayments and physicians visits: A panel data study of Swedish regions 2003–2012. *Health Policy,* 120, 1095-1099.

Januleviciute J, Askildsen J E, Kaarboe O*, et al.* 2015. How do hospitals respond to price changes? Evidence from Norway. *Health economics.*

Jatrana S & Crampton P. 2012. Gender differences in financial barriers to primary health care in New Zealand. *Journal of Primary Health Care,* 4, 113-122.

Jegers M, Kesteloot K, De Graeve D*, et al.* 2002. A typology for provider payment systems in health care. *Health Policy,* 60, 255-273.

Johnsen J R & Bankauskaite V. 2006. Health systems in transition. *Norway: European Observatory on Health Systems and Policies,* 8.

Kasper H & Sloan F A. 2008. *Incentives and Choice in Health Care*, MIT Press.

Kato H & Goto R. 2017. Effect of reducing cost sharing for outpatient care on children's inpatient services in Japan. *Health Economics Review,* 7, 28.

Keele L. 2015. The statistics of causal inference: A view from political methodology. *Political Analysis,* 23, 313-335.

Keele L & Minozzi W. 2013. How much is Minnesota like Wisconsin? Assumptions and counterfactuals in causal inference with observational data. *Political Analysis*, mps041.

Kenkel D S. 2000. Prevention. *Handbook of Health Economics.* Elsevier.

Kiil A & Houlberg K. 2014. How does copayment for health care services affect demand, health and redistribution? A systematic review of the empirical evidence from 1990 to 2011. *The European Journal of Health Economics,* 15, 813-828.

Kjekshus L E & Harsvik T. 2007. Organisasjonsutvikling ved norske sykehus [Organisational development at Norwegian Hospitals]. *Tidsskrift for den Norske Lægeforening,* 127, 288-290.

Kjerstad E. 2003. Prospective Funding of General Hospitals in Norway—Incentives for Higher Production? *International Journal of Health Care Finance and Economics,* 3, 231-251.

Kreif N, Grieve R, Hangartner D*, et al.* 2016. Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics,* 25, 1514-1528.

Kruschke J. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*, Academic Press.

Landsem M M & Magnussen J. 2018. The effect of copayments on the utilization of the GP service in Norway. *Social Science & Medicine,* 205, 99-106.

Leibowitz A, Manning W G, Keeler E B*, et al.* 1985. Effect of cost-sharing on the use of medical services by children: interim results from a randomized controlled trial. *Pediatrics,* 75, 942-951.

Liang L L. 2015. Do Diagnosis-Related Group-Based Payments Incentivise Hospitals to Adjust Output Mix? *Health economics,* 24, 454-469.

Lindahl A K. 2017. The Norwegian Health Care System. *In:* Mossialos, E, Wenzl, M, Osborn, R & Anderson, C (eds.) *International profiles of health care systems.* The Commonwealth Fund.

Lindrooth R C, Bazzoli G J & Clement J. 2007. The effect of reimbursement on the intensity of hospital services. *Southern Economic Journal*, 575-587.

Linna M, Häkkinen U, Peltola M*, et al.* 2010. Measuring cost efficiency in the Nordic Hospitals—a cross-sectional comparison of public hospitals in 2002. *Health Care Management Science,* 13, 346-357.

Majholm B, Engbaek J, Bartholdy J*, et al.* 2012. Is day surgery safe? A Danish multicentre study of morbidity after 57,709 day surgery procedures. *Acta Anaesthesiologica Scandinavica,* 56, 323-331.

Manning W G, Newhouse J P, Duan N, *et al.* 1987. Health insurance and the demand for medical care: evidence from a randomized experiment. *The American economic review,* 77**,** 251-277.

Martinussen P E & Midttun L. 2004. Day surgery and hospital efficiency: empirical analysis of Norwegian hospitals, 1999–2001. *Health Policy,* 68**,** 183-196.

Maynard A & Bloor K. 2012. Health Economics and Policy: the Challenges of Proselytizing. *In:* Glied, S & Smith, P C (eds.) *The Oxford Handbook of Health Economics.* Oxford Handbooks Online.

McGuire T G. 2011. Demand for Health Insurance. *In:* Pauly, M V, Mcguire, T G & Barros, P P (eds.) *Handbook of Health Economics.* Elsevier.

Ministry of Health and Care Services. 1995. St.meld.nr.44(1995-1996) Ventetidsgarantien - kriterier og finansiering [White Paper no.44 (1995-1996) The Waiting-time Guarantee - Criteria and Financing]. *In:* Ministry of Health and Care Services (ed.).

Ministry of Health and Care Services. 2016a. Endringer i helseregisterloven m.m. (kommunalt pasient-og brukerregister m.m.) [Proposition to parliament: Changes to the Health Register Act (municipal patient and user register). *In:* Ministry of Health and Care Services (ed.).

Ministry of Health and Care Services. 2016b. Prop 1 S 2016-2017 [Budget Proposition to Parliament 2016-2017]. *In:* Helse Og Omsorgsdepartementet [Ministry of Health and Care Services] (ed.).

Miyawaki A, Noguchi H & Kobayashi Y. 2017. Impact of medical subsidy disqualification on children's healthcare utilization: A difference-in-differences analysis from Japan. *Social Science & Medicine,* 191**,** 89-98.

Morgan S L & Winship C. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research,* New York, Cambridge University Press.

Norwegian Directorate of Health. 2011. *Utviklingen i antall frikort for 2010 [The change in number of co-payment exemption cards for 2010]* [Online]. Available: http://sites.helsedirektoratet.no/finansiering/refusjonsordninger/tall-og-analyser/frikort/Sider/utviklingen-i-antall-frikort-for-2010.aspx [Accessed 25th October 2017].

Norwegian Directorate of Health. 2014. Innsatsstyrt finansiering 2015 [Activity-Based Financing 2015]. *Regelverk.* Helsedirektoratet.

Norwegian Directorate of Health. 2015. Innsatsstyrt finansiering 2016 [Activity-Based Financing 2016]. *Regelverk.* Helsedirektoratet.

Norwegian Directorate of Health. 2016a. *NPR - formål og bruk av registeret [NPR - purpose and use of the registry]* [Online]. Available: https://helsedirektoratet.no/norsk-pasientregister-npr/om-npr/npr-formal-og-bruk-av-registeret [Accessed 23rd November 2017].

Norwegian Directorate of Health. 2016b. Styringsdata for fastlegeordningen, 4. kvartal 2016 [Regular GP statistics: 4th Quarter 2016]. Available: https://helsedirektoratet.no/Documents/Finansieringsordninger/fastlegestatistikk/hovedtallsrapport%20fastlegeordningen%20landstall%202016_4.pdf [Accessed 1st November 2017].

Norwegian Directorate of Health. 2017. Innsatsstyrt finansiering 2018 [Activity-Based Financing 2018]. *Regelverk [Guidelines].* Norwegian Directorate of Health,.

NOU 2014:12. 2014. Åpent og rettferdig - prioriteringer i helsetjenesten [Transparent and Fair - priority setting in health care].

O'Neill S, Kreif N, Grieve R, *et al.* 2016. Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Services and Outcomes Research Methodology,* 16**,** 1-21.

OECD. 2017. *Health spending (indicator)* [Online]. [Accessed 10th November 2017].

Office of the Auditor General. 2017. Riksrevisjonens undersøkelse av medisinsk kodepraksis i helseforetakene [The Auditor General's survey of medical coding practices in the health trusts]. Office of the Auditor General,.

Olsen C B & Melberg H O. 2016. Påvirker egenandeler ungdoms legebesøk? *Tidsskrift for velferdsforskning,* 19**,** 62-75.

Olsen J A. 2009. *Principles in Health Economics and Policy*, Oxford University Press.

Or Z & Häkkinen U. 2011. DRGs and quality: For better or worse? *In:* Reinhard Busse, Alexander Geissler, Wilm Quentin & Wiley, M (eds.) *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals.* McGrawHill.

Palmer K S, Agoritsas T, Martin D*, et al.* 2014. Activity-Based Funding of Hospitals and Its Impact on Mortality, Readmission, Discharge Destination, Severity of Illness, and Volume of Care: A Systematic Review and Meta-Analysis. *PLOS ONE,* 9**,** e109975.

Papanicolas I & McGuire A. 2015. Do financial incentives trump clinical guidance? Hip Replacement in England and Scotland. *Journal of Health Economics,* 44**,** 25-36.

Paul A & Nilsson A. 2014. The Effect of Copayments on Children's and Adolescents' Demand for Medical Care. *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik,* Health III, No. C10-V2.

Pauly M V. 1968. The economics of moral hazard: comment. *The American Economic Review,* 58**,** 531-537.

Pendzialek J B, Simic D & Stock S. 2016. Differences in price elasticities of demand for health insurance: a systematic review. *The European Journal of Health Economics,* 17**,** 5-21.

Ramm J. 2008. Inntektsforholdene til storbrukere av helsetjenester [Income distribution of resource intensive users of health care services]. Statistics Norway.

Remler D K & Greene J. 2009. Cost-sharing: a blunt instrument. *Annual review of public health,* 30**,** 293-311.

Remmerswaal M, Boone J, Bijlsma M*, et al.* 2017. Cost-Sharing Design Matters: A Comparison of the Rebate and Deductible in Healthcare. CPB Netherlands Bureau for Economic Policy Analysis.

Ringard Å, Sagan A, Saunes I S*, et al.* 2013. Norway: health system review. *Health Systems in Transition,* 15**,** 1-162.

Sakamoto H, Rahman M M, Nomura S*, et al.* 2018. Japan Health System Review. *Health Systems in Transition,* 8.

Schreyogg J, Stargardt T, Tiemann O*, et al.* 2006. Methods to Determine Reimbursement Rates for Diagnosis Related Groups (DRG): A Comparison of Nine European Countries. *Health Care Management Science,* 9**,** 215-223.

Secretariate for National Health Registry Project. 2009. Gode helseregistre – bedre helse [Good Health Registries - Better Health]. Norwegian Institute of Public Health.

Secretariate for the National Health Registry Project. 2016. Status For Nasjonale Helseregistre [Status for National Health Registries]. Institute of Public Health.

Shadish W, Cook T D & Campbell D. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin Company.

Simborg D. 1981. DRG creep: a new hospital-acquired disease. *The New England journal of medicine,* 304**,** 1602.

Skriabikova O, Pavlova M & Groot W. 2010. Empirical Models of Demand for Out-Patient Physician Services and Their Relevance to the Assessment of Patient Payment Policies: A Critical Review of the Literature. *International Journal of Environmental Research and Public Health,* 7**,** 2708.

Statistics Norway. Tabell: 10211: Folkemengde, etter kjønn og ettårig alder [Table:10211: Population by gender and age]. Statistics Norway,.

Statistics Norway. 2016. Tabell: 09750: Legeårsverk i kommunehelsetenesta, etter sentralitet, virkeområde og avtaleform [Table 09750: Man-years for physicians in the municipal health services, by centrality, activity area and wage agreement]. Statistics Norway.

Statistics Norway. 2017. *Health expenditure accounts for more than 10 per cent of GDP* [Online]. Available: https://www.ssb.no/en/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/health-expenditure-accounts-for-more-than-10-per-cent-of-gdp [Accessed 10th November 2017].

Stockley K. 2016. Evaluating Rationality in Responses to Health Insurance Cost-Sharing: Comparing Deductibles and Copayments. *Working Paper.* Harvard University.

Street A & Maynard A. 2007. Activity based financing in England: the need for continual refinement of payment by results. *Health Economics, Policy and Law, 2*, 419-427.

Street A, O'Reilly J, Ward P*, et al.* 2011. DRG-based hospital payment and efficiency: Theory, evidence and challenges. *In:* Busse, R, Geissler, A, Quentin, W & Wiley, M (eds.) *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals.* Mc Graw Hill.

Tjerbo T & Hagen T P. 2018. The health policy pendulum: Cost control vs activity growth. *The International journal of health planning and management, 33*, e67-e75.

Tylee A, Haller D M, Graham T*, et al.* 2007. Youth-friendly primary-care services: how are we doing and what more needs to be done? *The Lancet, 369*, 1565-1573.

Valdez R B, Brook R H, Rogers W H*, et al.* 1985. Consequences of Cost-Sharing for Children's Health. *Pediatrics, 75*, 952.

Verzulli R, Fiorentini G, Lippi Bruni M*, et al.* 2016. Price Changes in Regulated Healthcare Markets: Do Public Hospitals Respond and How? *Health economics.*

Votapkova J & Zilova P. 2015. The abolition of user charges and the demand for ambulatory visits: evidence from the Czech Republic. *Health economics review, 6*, 1-11.

Zápal J. 2010. Doctor-Visit Co-Payment Exemption for Children: First Look at the Data. *Finance a Uver: Czech Journal of Economics & Finance, 60*.

Zhang F, Wagner A K & Ross-Degnan D. 2011. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. *Journal of Clinical Epidemiology, 64*, 1252-1261.

Zou H & Hastie T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*, 301-320.

Zweifel P & Manning W G. 2000. Moral Hazard and Consumer Incentives in Health Care. *In:* Culyer, A J & Newhouse, J P (eds.) *Handbook of Health Economics.* Elsevier.

PAPER II

Did hospitals respond to changes in weights of Diagnosis Related Groups in Norway between 2006 and 2013?

Hans Olav Melberg*, Camilla Beck Olsen, Kine Pedersen

Department of Health Management and Health Economics, University of Oslo, Box 1089, Blindern, 0317 Oslo, Norway

ARTICLE INFO

ABSTRACT

It has been argued that activity based payment systems make hospitals focus on the diagnostic groups that are most beneficial given costs and reimbursement rates. This article tests this hypothesis by exploring the relationship between changes in the reimbursement rates and changes in the number of registered treatment episodes for all diagnosis-related groups in Norway between 2006 and 2013. The number of treatment episodes can be affected by many factors and in order to isolate the effect of changes in the reimbursement system, we exclude DRGs affected by policy reforms and administrative changes. The results show that hospitals increased the number of admissions in a specific DRG four times more when the reimbursement was increased, relative to the change for DRGs with reduced rates. The direction of the result was consistent across time periods and sub-groups such as surgical vs. medical, and inpatient vs. outpatient DRGs. The effect was smaller, but remained significant after eliminating DRGs that were most likely to be affected by upcoding. Activities that the hospital had little control over, such as the number of births, had small effects, while activity levels in more discretionary categories, for instance mental diseases, were more affected. This demonstrates that contrary to the wishes of policy makers the economic incentives affect hospital reporting and priority setting behavior.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Hospitals in many countries are financed partly by payments for each treatment in a specific diagnosis-related group (DRG). By using a system of DRG-payments instead of fixed transfers, policy makers want to increase the number of treatments, reduce unit costs and stimulate innovation [1]. At the same time, it is often claimed that the incentives produce undesirable consequences [2]. For example,

it has been argued that the system encourages up-coding of patients into the DRGs that are most profitable [3], that it makes hospitals select patients who are relatively easy to treat [4,5] and that it makes hospitals focus too much on the DRGs with the highest financial rewards [6,7]. Policymakers want the system to reward efficiency, but they do not want the incentives to affect prioritization between patients, treatments or diseases. As expressed by the Norwegian health authorities: "the main aim is to make the funding system as neutral as possible in terms of decisions regarding choice of form of treatment" [8].

In this paper, we test the hypothesis that the payment system is neutral with respect to hospital reporting behavior and decisions about the number of treatments they provide in different diagnosis-related groups. We do this by estimating whether, and to what extent, changes in

---

* Corresponding author at: Department of Health Management and Health Economics, Oslo University Hospital, University of Oslo, Box 1089, Blindern, 0317 Oslo, Norway.

E-mail addresses: h.o.melberg@medisin.uio.no (H.O. Melberg), c.b.olsen@medisin.uio.no (C. Beck Olsen), kine.pedersen@medisin.uio.no (K. Pedersen).

the reimbursement for different DRGs affect the reported number of treatment episodes. The main result from our analysis is contrary to the explicitly stated policy that financial incentives should not influence hospital behavior. Instead we show that increasing the DRG-weight by ten percent will lead to an increase of about one percent in the reported level of activity and that the annual increase in the number of treatments for DRGs with an increase in reimbursement was four times larger than for the DRGs with a reimbursement decrease.

1.1. Background

The relationship between changes in DRG-weights (i.e. reimbursements) and the level of recorded activity in the DRG has previously been explored using many different approaches. Some studies have focused on changes over time in a single DRG, such as a large increase or reduction in the treatment of a specific diagnosis following a change in the reimbursement or the cost of treatment. For instance, one study showed that during the period of 1999–2002 the number of treatments for sleep apnea increased by 110% when the costs of the intervention sank while the reimbursement remained high [9]. In another case study, Kuwabara and Fushimi [10] have shown that the introduction of a new DRG affected the choice between surgery and chemotherapy for breast cancer patients in Japan.

A second approach in the literature has been to focus on up-coding and in particular on admissions for diagnoses that are closely related [11,12]. For instance, some diseases have separate reimbursement codes for the same condition depending on whether it is classified as with or without complications. By examining how coding practice changes in response to relative changes in reimbursements, it is possible to analyze how hospitals react to incentives. In this case the focus is mainly on changes in the coding, and the concern is that the hospitals adapt to the system by using the most profitable DRG. However, while changes in closely related DRGs are useful for identifying financially motivated switches in reporting practice, it does not capture the extent to which changes in DRG reimbursements affect the actual priority given to some diseases or treatments.

In contrast to the focus on changes in a single DRG category, our aim is to examine overall average changes in hospitals' behavior in response to changes in reimbursement rates. This will consist partly of reporting changes and partly of changes in actual priorities. The analysis captures the net effect of both mechanisms, but in order to learn more about the importance of reporting change vs. prioritization, we also provide an analysis of sub-groups of DRGs that are less likely to be affected by upcoding.

The results are relevant for the design of financial systems in general since upcoding causes a financial burden and changes in prioritization affects patients and waiting times. In addition, the effect of reimbursement changes is of particular policy interest in systems such as the Norwegian one, which explicitly state that hospitals should not let DRG reimbursement affect the reporting or prioritization of treatments.

2. Data and method

2.1. Data and identification

We collected data on rates of reimbursement for all DRGs in every year between 2002 and 2013 in Norway. For the same time period we gathered the annual number of recorded events in the different DRGs at the national level from the Norwegian Directorate of Health and the Norwegian Patient Registry. The registry contains aggregate information about all hospital treatment episodes, inpatient and outpatient, from all hospitals in Norway.

The monetary reimbursement hospitals receive for each hospital stay is the product of three factors: the specific DRG-weight assigned to the DRG for the stay, the general monetary value of one unit of the DRG-weight, and the importance placed on activity-based payment relative to the global budget. Since the introduction of activity-based financing in Norway in 1997, the activity-based share of the budget has varied between 30% and 60%. In order to avoid biases introduced by changes in the share of the budget using activity based funding, we use the time period from 2006 to 2013 when the share remained constant at 40%. The monetary reimbursement for a unit of the DRG-weight increases every year, but the increase reflects the average cost of a hospital stay, and increases in the general reimbursement affects all DRGs equally. Consequently, when investigating the effect of changes in reimbursement rates on hospital prioritization, the key remaining component of relevance of the reimbursement is the DRG-weight.

A major problem for identifying the effect of financial incentives is that the financially favorable and unfavorable DRGs are not directly observable. The weight for a DRG in a given year is known, but since the hospitals' true costs are unknown, we do not know which DRGs have the most favorable relationship between costs and reimbursements. Because of this, it was necessary to find a more indirect way of identifying the effect of financial incentives. Instead of relying on information about the absolute levels of the (un)profitability of DRGs, we will use information about changes in the profitability. The key that makes it possible to identify these empirically, is the fact that changes in DRG-weights are lagged. If the reimbursement for a specific DRG was increased compared to the previous year, this demonstrates that the reimbursement was too low for at least part of the previous year. Similarly, a decrease in a DRG-weight identified DRGs that were too high for some time until they were revised. The change does not identify profitable and unprofitable DRGs, but it identifies DRGs that were marginally better or worse financially in one year relative to the next year. These changes in the DRG reimbursements are observed in the dataset and can be used to test whether hospitals adjust priorities based on changes in reimbursements as represented by the changes in the DRG-weights. If hospitals do not use reimbursements to prioritize between DRGs, changes in activity should not be related to the changes in DRG-weights in the DRGs. On the other hand, if they consider changes in reimbursements, one would expect the hospitals to have larger activity increases in the DRGs where the reimbursement increased, than in the DRGs where the reimbursement decreased.
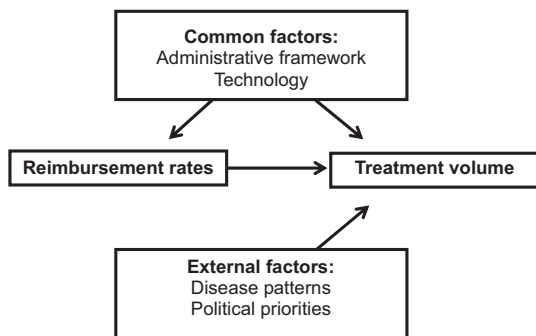
Fig. 1. Illustrating the problem of isolating causal effects. The arrows indicate causal relationships. Many factors influence treatment volume, but the key problem for causal analysis is when there is a common factor that affects both reimbursement rates and treatment volume.

A potential problem is that since cost is unknown a DRG that was unprofitable before a reimbursement change may still be unprofitable after the change. The method outlined above focuses on changes as opposed to levels of activity. An increase in reimbursement at the national level makes the DRG more favorable and this marginal change may affect hospital behavior even if it is still unprofitable. Such an effect assumes that hospitals care about more than profitability because if they did not they would only focus on the most profitable DRG. It does not assume that hospital costs are known, stable, or homogeneous across hospitals, only that an increase in the reimbursement will make it marginally more favorable than it was.

2.2. Isolating the causal effect of financial incentives

The method described above allows identification of financially favorable and unfavorable changes in diagnosis-related groups, but the causal relationship between financial incentives and prioritization between treatments is confounded by many factors. Fig. 1 illustrates the problem of causal inferences when estimating the effect of changes in reimbursements on treatment volume. First, treatment levels in different DRGs are affected by several factors external to the payment system, such as disease patterns in the population and political priorities like the introduction of treatment guarantees for certain diseases. Although these factors are important, they are related to the number of treatments and not to reimbursement. This means that they do not confound the relationship between reimbursement rates and volume (Morgan and Winship, 2007). For instance, sudden events such as disease outbreaks may cause a large number of interventions to be registered in a particular DRG in one year, but as long as there is no systematic relationship between the outbreak and the change in the reimbursement, these factors do not create any bias in the estimated relationship between financial incentives and treatments. They increase the noise and variance, but since they are sometimes positive and sometimes negative, they do not create systematic bias. The key challenge for causal identification is the second group of variables; those which affect both the reimbursement rate and the volume.

There are two main types of changes that affect both reimbursement rates and volume of a DRG: Changes in the administrative framework and the introduction of new technology. Administrative changes include the implementation of new DRGs, splitting of existing DRGs into new categories, and altering regulations for when a DRG is to be used. For instance, after splitting a DRG, both the reimbursement and the treatment volume in the original DRG will change, but in this case the change in volume is caused by the new regulation and should not be interpreted as being the outcome of changing reimbursement rates. Second, the introduction of new technologies is likely to influence both the average treatment cost (which is used to calculate the reimbursement) and the activity level, for instance by making the treatment more effective. In order to isolate the relationship between financial incentives and volume, it is important to reduce the influence of these types of variables and explore the type of bias they may create.

To reduce the confounding effect of administrative changes we eliminated observations that were affected by these factors. For instance, we excluded all observations where the DRG reimbursement or the number of treatments was zero. A large increase from zero, or a large decrease to zero, are usually caused by administrative rules about coding practice and contain no information about the effect of financial incentives.

To further eliminate the effect of irrelevant changes, we eliminated DRGs where special reimbursements rules apply, for instance DRGs related to rehabilitation. In these cases, the DRG reimbursements did not represent the true payment because the hospitals also received extra payment for each day the patients were under their care.

During 2009 and 2010 a major revision of the reimbursement system took place, which included the introduction of several new DRGs in both years, and a comprehensive recalibration of all DRGs and reimbursements in 2010. This means that the system in 2010 was very different from the system that was in place in 2009. In order to avoid the bias introduced by the recalibration and reorganization of the entire set of DRGs we excluded all changes between 2009 and 2010 (see Fig. 2).

There were some cases where the reimbursements or treatment volume was almost zero, but not quite zero. In this case, small absolute changes will cause very large relative changes, which could dominate the results. We used two strategies to reduce this problem. The first strategy was to eliminate the most extreme changes since they most likely reflect administrative changes and not responses to economic incentives. We did this in the main results, but we also report the results where these changes were not eliminated. Second, we examined whether weighting the results by the budget share of the disease (relative to the hospital budget) made a difference. Weighting the percentage change in a category by the budget share of the category reduced the problem of results being dominated by many large changes in small DRGs.

To examine the stability of the results, we compared differences in means between DRGs with increasing or decreasing DRG-weights using many different samples and sub-groups. The first sub-group contained the entire
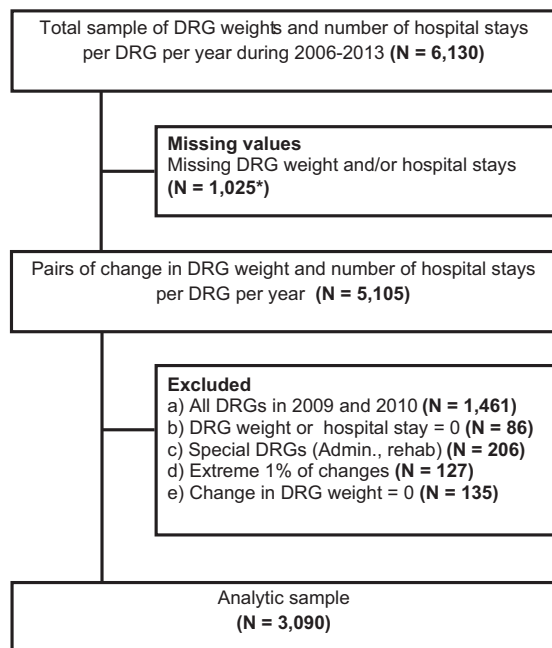
Total sample of DRG weights and number of hospital stays per DRG per year during 2006-2013 **(N = 6,130)**

Missing values
Missing DRG weight and/or hospital stays
(N = 1,025*)

Pairs of change in DRG weight and number of hospital stays per DRG per year **(N = 5,105)**

Excluded
a) All DRGs in 2009 and 2010 **(N = 1,461)**
b) DRG weight or hospital stay = 0 **(N = 86)**
c) Special DRGs (Admin., rehab) **(N = 206)**
d) Extreme 1% of changes **(N = 127)**
e) Change in DRG weight = 0 **(N = 135)**

Analytic sample
(N = 3,090)

Fig. 2. Included and excluded observations. The figure shows the number of observations that were excluded, either because there was some missing information or because the information was not useful to answer to the causal question.

analytic sample. The results for this group are compared to sub-samples containing only the years prior to and after the major administrative revisions in 2009 and 2010, and the results for only inpatient care and non-surgical DRGs. Moreover, we calculated whether the average effect of changing DRG-weights differed depending on which main diagnostic category the DRG belonged to and whether weighting the results according to the size of the DRG mattered (as measured by the budget share of the DRG).

To learn more about the importance of reporting change vs. prioritization, we provide results using two sub-sets of DRGs. First, we present results after excluding 248 DRGs that are likely to be affected by upcoding. The excluded DRGs are the so-called "twin-DRGs" that describe the same underlying condition, but one is reserved for uncomplicated cases and the other (and more rewarded) category is supposed to be used for complicated cases. Because the underlying description is similar, these DRGs are likely to be affected by upcoding and excluding these reduce the influence of changes in reporting behavior. Second, although excluding twin DRGs gives some insight into the relative importance of upcoding, the approach is limited by the possibility that there might also be upcoding between DRGs that are not twins. To reduce this problem, we exploit a change from 2010 that split DRGs related to inpatient vs. outpatient surgery. Excluding 17 outliers (for instance, reimbursement ratios that were reduced by 100%) there are 97 such surgical DRG pairs in the dataset (between 2010 and 2013). These DRGs can be used to analyse whether changes in the reimbursement ratio also affected the ratio of surgeries performed as inpatient or outpatient. The sub-group is interesting because it is administratively difficult

to misreport an outpatient surgery as an inpatient surgery. This means that if changes in the price ratio affect the balance of inpatient to outpatient surgery, reimbursements are likely to have real effects and not just reflect a change in reporting behavior.

Finally, to estimate the overall average elasticity of the level of reported activity in response to changes in DRG-weights, and not just the difference between increasing and decreasing the DRG-weight, we estimated three regression models. In all the models the dependent variable was the changes in the level of activity (for all DRGs in all years). In the first and simple model the independent variable was only the percentage change in the DRG-weight. Next a more complex model was estimated to examine if the elasticity was constant after adjusting for other factors that might affect changes in the level of activity; whether the DRG was inpatient or outpatient, surgical or non-surgical, the size of the DRG-weight, the number of events in the DRG, changes in reimbursements in previous years and events specific to the each year. Finally, a model with interactions and the change in the DRG weight was estimated to examine if changes in DRG-weights had different effects depending on whether the change occurred in different types of DRGs.

3. Results

3.1. Descriptive statistics

Across all DRGs for all periods, the DRG reimbursement (weight) ranged from 0.01 to 39.13, with a mean value of 1.49 and a median of 0.73. The average annual increase in the reimbursement was 3.3%, ranging from −52.0% to 102.9% for the different DRGs. The number of hospital stays in a DRG varied between 4 and 299,229 per year, with a mean value of 6050 treatments. Due to laws protecting privacy, diagnostic categories with less than four treatments are set to missing in the central registry. The mean annual growth in the number of treatments in the DRGs was 2.0%, ranging from −44.1% to 72.7%.

3.2. Reimbursement and activity

For inpatient and outpatient care across all years, treatment activity increased by 3.16% for diagnostic groups with increased reimbursement, while a reduction in the reimbursement was associated with a 0.74% average increase in treatment activity (see Table 1). The same pattern of a higher increase in activity for DRGs with reimbursement increases was found in different sub-samples and time periods (see Fig. 3). During the years 2006–2008 the increase in activity was 4.44% for diagnostic groups with an increase in reimbursement, and there was a decrease of 0.10% in the number of treatments provided for DRGs with a decrease in reimbursement. In the time period 2011–13 an increase in reimbursement was associated with an increase in activity of 2.72%, while the DRGs with a decreased reimbursement increase by less than half of that (1.14%). For inpatient care across all years the number of treatments increased by 3.49% for procedures with increased reimbursement, while a reduced reimbursement corresponded to an increase in activity of 0.36%. Both

Table 1

Comparing annual changes in activity levels for DRGs with a reduction and an increase in reimbursement in different sub-groups (unweighted and weighted by budget share).

	Unweighted DRG weight			Weighted by budget share DRG weight		
	Reduction	Increase	Difference	Reduction	Increase	Difference
All years	0.74 (0.36)	3.16 (0.33)	2.42[*]	1.74 (0.46)	3.67 (0.47)	1.94[*]
Subsample 2006–08	−0.10 (0.60)	4.44 (0.69)	4.54[*]	1.81 (0.70)	5.12 (0.88)	3.31[*]
Subsample 2011–13	1.15 (0.45)	2.72 (0.37)	1.58[*]	1.69 (0.62)	2.81 (0.50)	1.12[***]
Only inpatient care DRGs	0.36 (0.39)	3.49 (0.37)	3.13[*]	1.53 (0.51)	3.74 (0.52)	2.21[*]

Standard error in parenthesis below the mean.

[*] Significant at a 1% level.

[***] Significant at 10% level (two-sample *t*-test assuming unequal variances).

surgical and non-surgical DRGs had larger increases in activity when the DRG-weight increased than when the DRG-weight decreased, but the difference was smaller for surgical DRGs compared to medical DRGs (1.6 vs. 3.5 percentage point differences in activity level changes). After excluding the DRGs most likely to be affected by upcoding ("twin DRGs"), DRGs with an increase still increased the activity significantly more than DRGs with a decreased reimbursement (3.15% vs. 1.23%), but the difference was smaller than in the original sample.

In order to avoid the results being dominated by a few large DRGs the above numbers are unweighted by the number of events in the DRGs. Unweighted results avoid one problem, but create another: The conclusions may be influenced by changes in many small categories. To examine the extent to which weighting affected the results, we calculated the weighted averages using the budget share of each DRG as weight. The budget share is calculated

using the number of events in the DRG and the size of the reimbursement (to avoid the results being overly influenced by many small "cheap" events). After weighting using the budget share, both the size and the pattern of the differences remained similar to the unweighted results, which indicate that the results were not dominated by some large or many small changes (Table 1). For instance, using weighted calculations led to the result that increased reimbursements raised activity levels by 3.67% and decreases in reimbursement were associated with a 1.74% increase in activity.

To summarize the information, it is useful to focus on the difference between the DRGs with an increase in reimbursement vs. the DRGs with decreased reimbursement. If the difference is positive, DRGs with increased reimbursement had a higher increase in activity than DRGs that had their reimbursement decreased. As seen in Table 1, the differences were all positive in different sub-samples (2.42,
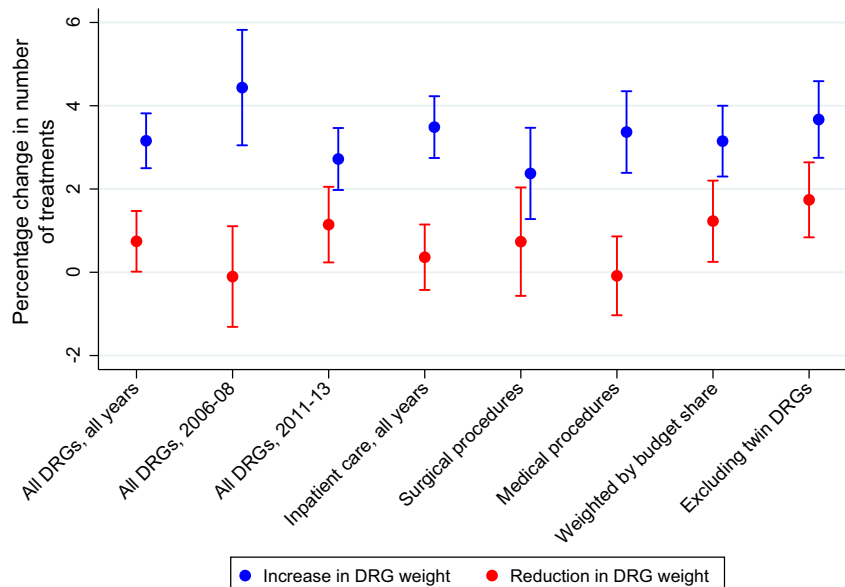

Fig. 3. Annual percentage change in treatment levels for DRGs with increased and decreased reimbursement in different sub-groups.

Percentage point difference in activity levels between DRGs with increases vs. decreases in reimburcements
(High values: Most responsive to changes in reimbursements)
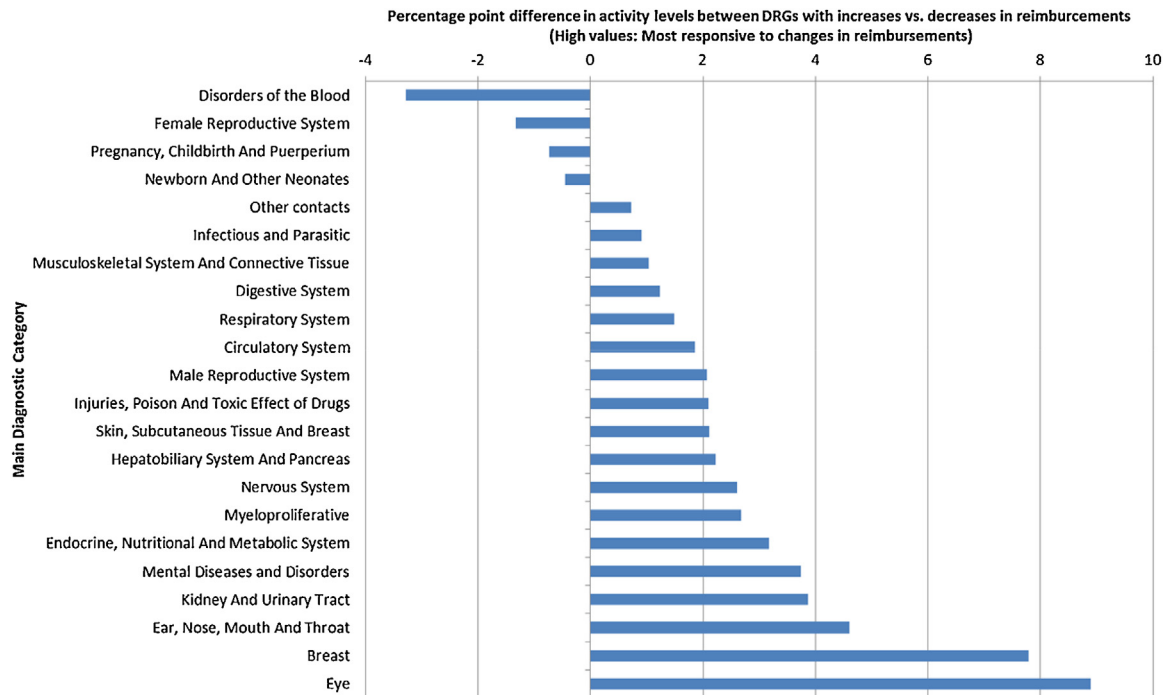

Fig. 4. Difference between DRG's with increased and decreased reimbursement in different main diagnostic categories.

4.54 and 1.58 and 3.13) and they were all significant at the 5% level.

Fig. 4 shows the average difference between increasing and decreasing DRG-weights for DRGs depending on which main diagnostic category the DRG belongs to. A value of zero indicates that there was no difference in the change in activity between the DRGs with an increase or a decrease in the DRG-weight. Negative values, as observed in the diagnostic categories related to diseases of the blood, female reproductive system, pregnancy and newborns, indicate that the level of activity is not affected by changes in the DRG-weight (i.e. the level of activity increased most for those DRGs that experienced a decrease in the DRG-weight). Negative results for some categories are to be expected since incentives is not the only factor affecting levels of activity and activity levels in some DRGs are dominated by factors outside the control of hospitals.

Information on the ratio of surgical DRGs that are done inpatient vs outpatient is presented in Fig. 5. The DRGs are split into two groups: Those for which the reimbursement ratio between inpatient and outpatient surgery moved in favor of outpatient treatment from one year to the next (increase), and those for which the reimbursement ratio moved in disfavor of outpatient treatment (decrease). In 2011 and 2012 the proportion of all surgery that was done outpatient decreased, but the reduction was less for surgical DRG pairs with increased reimbursement ratios. In 2013 the share of outpatient surgery increased for both groups, but the increase was largest for DRG-pairs that had experienced an increase in the reimbursement ratio compared to the previous year. The overall average for the whole period was that for surgical DRG pairs with increased reimbursement ratios, the proportion of DRGs that was

done outpatient increased by 0.57%. For surgical DRG pairs with a decrease, the ratio of surgeries done outpatient was reduced by 0.62%.

The results from the regressions indicated that a ten percentage increase in the DRG-weight is associated with an increase of between 0.76 and 1.10 percent in the level of activity (Table 2). The DRG-weight and its lagged value were statistically significant in all models. In addition to changes in the DRG-weight and the dummy for the year 2012, the only other variable that was significantly associated with variations in activity levels in the model was the absolute level of the DRG-weight: DRG's with large absolute weights tended to have large increases in activity levels. In the model with interactions, the effect of changing DRG-weights was smaller for inpatient DRGs relative to other DRGs in the model, but none of the other interactions were statistically significant.

4. Discussion and policy implications

Our analysis indicates that reimbursements most likely have a significant impact on hospital behavior. The pattern was consistent across different time periods, sub-groups and sample selections. On average, DRGs with increased reimbursements increased the reported level of activity by one to five percentage points more than DRGs with decreased reimbursements. The difference between DRGs with increased and decreased reimbursement rates may appear small when expressed in terms of absolute percentages, but in relative terms the increase in volume for the DRGs with increases in DRG-weights (3.16%) was more than four times larger than the increase for DRGs that experienced reduced DRG-weights (0.76%). To put these
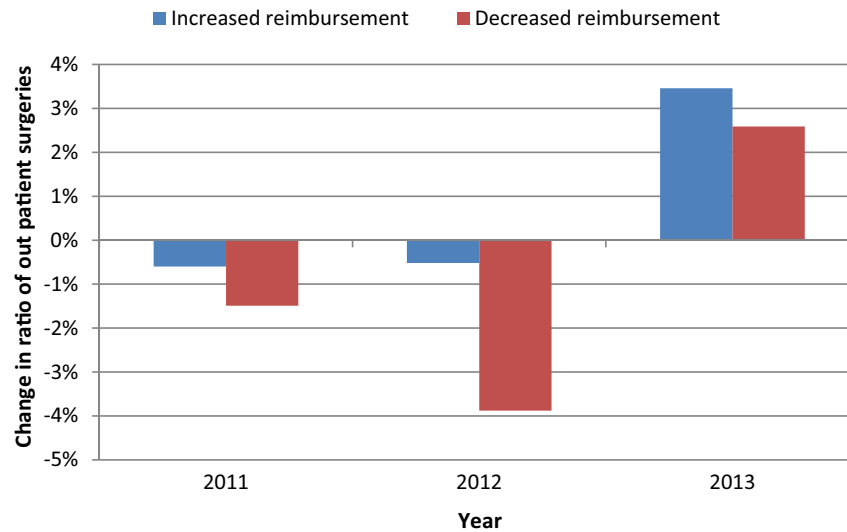
Fig. 5. Changes in proportion of surgeries done outpatient for surgical DRG pairs with increasing and decreasing reimbursement ratios in favor of outpatient surgery.

numbers into perspective, it is useful to recall that the overall annual growth in reported DRGs in the same period was 2%. Another useful background figure is that a difference of 2% amounts to about 36,000 hospital inpatient days or about 9000 patients.

After excluding the DRGs most likely to be affected by reporting practice as opposed to changes in activity levels, the importance of reimbursement on reported activity levels remained significant, but the size of the difference decreased. In the large sample, DRGs with increased reimbursements experienced four times larger increase in activity levels, but in the restricted sample the increases in reimbursements were associated with changes in activity levels that were twice as large as those for DRGs with decreases (3.15% increase in activity levels for DRGs with increased reimbursement and 1.23% for DRGs with decreased reimbursements). The difference demonstrates both the effect of reimbursements for reporting practice, but also that the difference remained large and significant even after eliminating the DRGs that are most susceptible to upcoding. The same conclusion emerged from the analysis of the proportion of surgeries done outpatient in

DRG pair where both inpatient and outpatient treatment was possible: Increases in the reimbursement ratio for outpatient surgery tended to correspond to a higher share of outpatient surgeries.

The results complement the general finding that financial incentives matter for hospital behavior. For instance, previous studies in different countries have shown that a prospective DRG payment system tends to reduce length of stay [13], to reduce waiting times [14] and to increase the treatment volume [15]. This study complements these and other studies by showing that the system, in contrast to the aims of the policy makers, creates changes in behavior that could partly be due to prioritization between different DRGs. Furthermore, our results complement the findings of another study that investigated the effect of reimbursement changes in Norway [16]. They estimated a price elasticity of 0.094, i.e., a ten percent increase in the price for a treatment would yield an increase in treatment volume of almost one percent.

The result that financial incentives influence hospital prioritization should not be confused with the naïve conclusion that hospitals care more about money than

Table 2
Regression results, changes in DRG activity levels ($n = 3090$).

Variable	Simple model	Full model	With interactions
DRG-weight (% change)	0.075[*]	0.087[*]	0.061[*]
Constant	1.786[*]	0.831	1.358
Outpatient DRG		−0.406	−0.374
Non-surgical DRG		0.989	0.811
DRG weight (level)		0.420[*]	0.252[*]
Number of treatments in DRG (1000)		0.040	0.00004
Previous period DRG-weight (% change)		0.034[*]	0.037[*]
Dummy for 2011		−0.527	−0.430
Dummy for 2012		−1.889[*]	−1.915[*]
Dummy for 2013		−1.711	−1.618
DRG-weight change × outpatient DRG			−0.158[*]
DRG-weight change × non-surgical DRG			0.0179
DRG-weight change × DRG-weight (level)			0.0196

[*] Significant at 5% level.

patients' health. Instead, it may be interpreted as a consequence of hospitals trying to maximize the provision of care to as many patients as possible in a given payment system. Faced with a budget constraint and externally set reimbursement rates, they must take financial incentives into account if they want to maximize the number of treatments across different diseases. If they fail to do so they could end up spending a large share of their budget on relatively few and expensive treatments. In this sense, being responsive to financial incentives is not necessarily negative. However, a strong focus on imperfect and misaligned financial incentives may cause an undesirable composition of health production in terms of over-treatment of patients with the most profitable diagnoses and under-treatments for patients who are in strong need but whose procedure is too expensive for the hospital.

Given that the core of the problem is the misaligned incentives, one important policy implication of the finding is the importance of accurate and frequently updated DRG-weights to reduce the unintended consequences of a DRG based payment system. For instance, in some systems the DRG-weights are only updated once every year, and often using data on costs that are two years old. New technology and costs may change significantly within this period which, in turn, creates undesirable priority differences between the DRGs.

Secondly, the result that hospitals are affected by DRG reimbursements also has the policy implication that the reimbursement could be used as a tool to set health priorities. For instance, Poland has done this by changing a tariff to give special emphasis to the treatment for stroke patients [17], and recently Norway increased all DRG reimbursements for surgical outpatient treatments in order to change hospital behavior. The results suggest that these changes in reimbursements could be an effective way of affecting hospital behavior. The results also suggest which diagnostic groups that are most sensitive to incentives, and those that for medical and other reasons are less sensitive (Fig. 4).

A potential limitation of the analysis is that the original cause of the change in the DRG reimbursement is unobserved. For instance, if the cause is the introduction of a new technology this may lead not only to changes in costs, but also to improved treatments. This is a problem because one should not attribute all the change in volume to the change in financial incentive, when in fact the causal mechanism was that new technology changed the utility of the treatment. This is an important argument, but in order to determine the size and direction of the potential bias, it is useful to distinguish between the different scenarios tied to the introduction of new technology: (i) better and cheaper, (ii) better, but more expensive, (iii) no change in medical outcome, but reduced costs.

In the first scenario a cheaper and better treatment technology appears. The reduced costs will be reflected in reduced reimbursements. This will not necessarily reduce the level of activity since the effect depends on the difference between the reduction in cost and the reimbursement. If the system works as intended, the new reimbursement will be closer to the true cost than the old reimbursement, which was too high for part of the previous period. In this case the reduction creates an incentive for reducing the number of treatments. However, the negative price effect is balanced by the increased utility from the treatment, which provides incentives for increasing the number of treatments. The net effect in the first scenario is that the pure price effect is dampened by the change in the utility of the treatment. In this case the danger is that we underestimate the effect of financial incentives, not that we exaggerate its effects.

In the second scenario, a better and more costly technology appears. When it appears, the net price will be higher, for a given reimbursement, than the old technology. For this reason, the hospital will not use it as much as they would like to while the old reimbursement is in place. However, because the technology is better, the difference between the volumes during the period of unrevised reimbursement and after the reimbursement has been updated, will not be as large as the cost difference would indicate. Once again, the utility effect of the new technology dampens the effect of financial incentives. The data cannot isolate the effect of each mechanism, but since they point in different directions we know that the observed effect of financial incentives in this case is a conservative estimate of the true effect.

In the last scenario reduced costs will impact treatment volume only through reduced reimbursement. In this case the observed change in volume reflects the change in financial incentives, but the size depends on the length of the period the reimbursement was wrong. Once again this leads to an underestimation since the misaligned rate will not last for the whole time period.

In sum, although technological advancements are unobservable, the different logical possibilities indicate that the results are conservative and if anything the various limitations indicate that the effect should be larger than our estimates.

In addition to the limitations caused by unobservable technology, there are also some other limitations that suggest further studies. One such key limitation is the use of aggregate data on the national level. More detailed data, with information about waiting times and activity at the hospital/ward level as opposed to only national level data, could give more information on how increases in some DRGs are associated with decreases in other DRGs. For instance, more detailed data could be used to examine whether emergency care is less sensitive to reimbursement changes than elective care. A previous study using similar data has shown, surprisingly, that this was not the case [16]. This was partly explained by the fact that the definition of emergency care was more fluid than one might expect, for instance whether a heart problem was considered serious enough for the hospital to admit the person as a patient.

Another potentially important limitation is that hospital behavior is affected by changes in budgets.[1] This means that changes in the activity levels are not only influenced by reimbursements, but also by the income

elasticity of different treatments. As budgets increase, some treatments may increase the activity levels faster than other treatments, and the effect may also differ between hospitals. Partly for this reason the analysis was limited to the time period where the reimbursement level as a percentage of the whole budget was held constant by the authorities. However, overall changes in income could still influence the activity levels and it is important to consider how this affects the results.

The main problem of income effects is not the effect of income changes in itself, but the extent to which these are correlated with the variable of interest i.e. reimbursement changes. If income levels affect the level of activity for some DRGs positively and some negatively, and this is not correlated with the reimbursement levels, the results still hold. If, however, there is a systematic relationship between changes in DRG reimbursement and income, the results are weakened. For instance, if DRGs for the treatment of heart problems experience frequent and systematic changes in reimbursements during the time period, and income changes lead hospitals to focus more on heart problems during the same time period, there will be a correlation between reimbursement changes and activity levels. In this case the observed effects will be caused by income and not only a substitution effect driven by relative differences in reimbursements. The results of the paper should be interpreted with this limitation in mind.

The same argument that applies to income, also applies to other factors. The activity levels for different DRGs are clearly not only influenced by reimbursements, but by many factors including health trends, prioritization, guidelines and so on. Often it is not physically possible to quickly increase or decrease treatment levels since it takes time and space to increase activity and specialists in different departments are not perfect substitutes. The finding that reported activity levels increase significantly more for DRGs with increased reimbursement should not be interpreted as conclusive evidence that only money governs prioritization, but it suggests that reimbursement is one of many factors that affects hospital behavior in a public system using DRGs.

5. Conclusion

Understanding the mechanisms of the payment system is crucial to achieve the desired allocation of health care resources. Our analysis utilizing recent data from Norwegian hospitals provide evidence that hospitals change both reporting and actual behavior in response to variations in the reimbursement for different diagnosis-related groups. In this sense the system is not neutral with respect to priority setting and decision makers must face the difficulty that a policy aimed at increasing effectiveness also tend to affect prioritization. This tradeoff can be done better when they have knowledge of how much payments affect prioritization between different patient groups. Finally, the results underline the importance of accurate DRG-weights to avoid undesired health priorities in a DRG-based payment system.

References

[1] Busse R, Quentin GA, Wiley WM, editors. Diagnosis related groups in Europe: moving towards transparency, efficiency and quality in hospitals. Open University Press; 2011.
[2] Fourie C, Biller-Andorno N, Wild V. Systematically evaluating the impact of diagnosis-related groups (DRGs) on health care delivery: a matrix of ethical implications. Health Policy 2014;115:157–64.
[3] Steinbusch PJ, Oostenbrink JB, Zuurbier JJ, Schaepkens FJ. The risk of upcoding in casemix systems: a comparative study. Health Policy 2007;81:289–99.
[4] Martinussen PE, Hagen TP. Reimbursement systems, organisational forms and patient selection: evidence from day surgery in Norway. Health Economics, Policy and Law 2009;4:139–58.
[5] Ellis RP, McGuire TG. Provider behavior under prospective reimbursement. Cost sharing and supply. Journal of Health Economics 1986;5:129–51.
[6] Dafny LS. How do hospitals respond to price changes? American Economic Review 2005;95:1525–47.
[7] Ellis RP. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. Journal of Health Economics 1998;17:537–55.
[8] The Norwegian Directorate of Health. Activity-based funding of health services in Norway. An assessment and suggested measures, Oslo, Norway; 2007.
[9] Kvaerner KJ. Thought-provoking adaptation of otorhinolaryngologic surgery. Tidsskrift for Den Norske Laegeforening 2004;124:2743–5.
[10] Kuwabara H, Fushimi K. The impact of a new payment system with case-mix measurement on hospital practices for breast cancer patients in Japan. Health Policy 2009;92:65–72.
[11] Silverman E, Skinner J. Medicare upcoding and hospital ownership. Journal of Health Economics 2004;23:369–89.
[12] Carter GM, Newhouse JP, Relles DA. How much change in the case mix index is DRG creep? Journal of Health Economics 1990;9: 411–28.
[13] Hamada H, Sekimoto M, Imanaka Y. Effects of the per diem prospective payment system with DRG-like grouping system (DPC/PDPS) on resource usage and healthcare quality in Japan. Health Policy 2012;107:194–201.
[14] Siciliani L, Hurst J. Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries. Health Policy 2005;72:201–15.
[15] Or Z. Implementation of DRG Payment in France: issues and recent developments. Health Policy 2014;117:146–50.
[16] Januleviciute J, Askildsen JE, Kaarboe O, Siciliani L, Sutton M. How do hospitals respond to price changes? Evidence from Norway. Health Economics 2016;25(5):620–36.
[17] Bystrov V, Staszewska-Bystrova A, Rutkowski D, Hermanowski T. Effects of DRG-based hospital payment in Poland on treatment of patients with stroke. Health Policy 2015;119:1119–25.