Sending a signal

The Effect of Raising Non-attendance Fees in Swedish Health Care

Sofia Frögelius & Sara Lundin
DEPARTMENT OF ECONOMICS
NEKP01
SUPERVISOR: JAN BIETENBECK
2016-06-02

LUNDS UNIVERSITET
Ekonomihögskolan

Acknowledgements

We thank our supervisor Jan Bietenbeck for guidance and support. We would also like to thank Christina Jonsson and Pia Landgren , *Region Skåne*, and Gunnar Ljunggren, *Stockholm Läns Landsting*, for assistance with data collection, insight in patient fee regulations, and the organizational structure in the regional governments. We are grateful for your time and effort.

Abstract

In this paper, we study the effect of raising the fee for non-attendance on non-attendance rates in health care from a Swedish perspective. We do this by using a difference in difference design and exploiting a natural experiment in Sweden, where one county increased non-attendance fee to the double and the other county had a constant fee. Data on individual level allows us to investigate a general effect as well as heterogeneity in treatment depending on age and tendency to miss appointments in the past. The theoretical framework departs in neoclassical theory and the deterrence hypothesis, which predict that patients will decrease their rate of non-attendance. Meanwhile, behavioral research shows that an external incentive can crowd out intrinsic motivation. The theoretical predictions of the effect of a fee are therefore ambiguous and previous research on the field is scarce. Our study can contribute with generalizable evidence from both primary and secondary care in Sweden. The results are a decrease in non-attendance rate by 39% when doubling the non-attendance fee. Skåne County implemented the reform and using back of the envelope calculations indicate cost reductions estimated to at least 69.2 million SEK annually. We draw the conclusion that monetary sanctions have a large effect on non-attendance rates and signaling, through media and in the health care organization, might have enhanced the effect.

Keywords: Health economics, Non-attendance, Deterrence hypothesis, Intrinsic motivation

Table of contents

List of Figures

List of Tables

1. Introduction

Costs in health care are high and they are rising year by year in Sweden. The total health care expenditures in Sweden were about 9.6 percent of GDP in 2012; this is 0.3 percentage points higher than the OECD average of 9.3 percent of GDP. Meanwhile, health care expenditures have been rising by between 2-3 % annually in the past decade in Sweden (OECD, 2014). Because health care accounts for a large part of public spending and expenditures are growing at a fast pace, it is of importance to aim for efficiency in the health care sector and look for methods to reduce spending. Rising costs in health care are an issue that has been widely debated among researchers and the effects of increased spending on health outcomes are ambiguous (Skinner et al, 2006). In Sweden, health care is to a large share funded and organized by the regional governments, which in turn collect revenues from income taxation (Hälso- och sjukvårdslag, 1982:763). Problems with moral hazard and overconsumption may arise in a government funded health care system since patients do not bear the costs of their health care consumption (Aron-Dine, Einav, & Finkelstein, 2013; Bhattacharya, Tu & Hyde, 2014). In the economic field, extensive research has been made on patient co-payment schemes and the possibility to steer patients into socially optimal behavior using monetary incentives (see e.g. Aron-Dine, Einav, & Finkelstein, 2013; Zweifel & Manning, 2000).

Non-attendance to appointments is one problem to efficiency and increased costs in health care for which monetary incentives are used to steer patients into socially optimal behavior (Bech, 2005). Non-attendance fees have long been used in Sweden in an attempt to reduce non-attendance rates. Non-attendance is defined as not attending a health care appointment without cancelling at least 24 hours prior to the appointment or having a legitimate cause for not attending (Rydmarker & Deppert, 2015). A common policy for non-attendance fees in Sweden is that patients that do not attend an appointment must pay the corresponding patient fee for the planned visit (1177 Vårdguiden, 2016). In contrast to actual health care consumption, where a trade-off has to be made between costs and utility of care, non-attendance clearly causes unnecessary costs because of poorer allocation of resources, longer waiting lines, and loss in production. In addition, non-attendance causes poorer health outcomes as a result of insufficient care (Deyo & Inui, 1980; Bech, 2005). Costs of non-attendance are high. At Karolinska university hospital in Stockholm, the losses due to non-

attendance were 12.7 million SEK[1] in 2007, also Akademiska sjukhuset in Uppsala and Sahlgrenska in Gothenburg reported high numbers of missed appointments (Nilsson, 2008).

In this paper, we evaluate the theoretical and empirical effect of increasing the fee for non-attendance in health care from a Swedish perspective. Different groups are used to investigate heterogeneity in the treatment and to understand how individuals react to increasing the fee in order to better target groups that are especially prone to miss appointments. The expected effect of raising the fee for non-attendance in health care is not straightforward. The economic theory on sanctions for unwanted behavior has been formalized by Becker (1968). In this neoclassical economic theory of utility maximizing individuals, patients decide to not attend their appointments until the marginal utility of not attending equals the marginal cost. Hence, if an individual does not face the cost of his or her actions, there is an incentive to utilize too much non-attendance. When raising punishment, the net benefit of non-attendance becomes negative and the rate of non-attendance will therefore decline (Bech, 2005). Meanwhile, in modern research, which has integrated research in social psychology and economic theory, there is evidence for that external incentives, such as a fee for non-attendance, under some circumstances can crowd out intrinsic motivation (Gneezy, Meier, & Rey-Biel, 2011). As a consequence, increasing the fee for non-attendance can lead to higher non-attendance rates. Along with ambiguous theoretical predictions there is little previous research on fees for non-attendance in health care. A study at an ophthalmological care center in Finland in 1992 did not find any statistically significant effects of imposing a fee on non-attendance at their health care center (Mäntyjörvy, 1994). Another study at a community health care center in the US imposed a fee of 30 USD on non-attendance. Patients who frequently did not attend their appointments had a lower rate of non-attendance after imposing the fee (Lesaca, 1995).

Because of this ambiguity in the theoretical and empirical predictions of raising the fee for non-attendance, we estimate the effect of increasing the fee for non-attendance empirically. We do this by exploiting a natural experiment in Sweden where one county, Skåne County (*Region Skåne*), reformed their policy for non-attendance in 2012. The policy requires patients to pay a fee corresponding to twice the patient fee for non-attendance, compared to previously having to pay a single patient fee (Regionfullmäktige Skåne, 2011). Meanwhile, Stockholm County (*Stockholms Läns Landsting*), has had a constant non-attendance fee corresponding to a single patient fee for non-attendance. Using two counties allows for a

[1] 1 SEK=0.10715 EUR (Bloomberg.com, 2016)

difference in difference design and the possibility to control for state specific time invariant factors as well as common time variant factors (Angrist & Pischke, 2009). This makes our results unique and allows for the possibility to draw causal inferences of increasing a fee for non-attendance on non-attendance rates in health care, which to our knowledge has not been done before. We do this using an unique panel data set with administrative data on attendance and non-attendance to patients' health care visits in primary and secondary care between 2008 and 2015 from Skåne County and Stockholm County.

Our empirical results show that a doubling of the fee for non-attendance has a large effect on non-attendance rates. The non-attendance rates decrease by 39 % in the main model specification. These results have large policy implications, using back of the envelope calculations indicate that decreasing non-attendance rates by 39 % decreases costs for Skåne County by at least 69.2 million SEK per year. Different age groups have different treatment effects, but the intervention is effective for all groups. The intervention is especially effective for those who have missed at least one appointment prior to the reform. Increasing the fee for non-attendance is therefore an intervention with large net social benefits. These results indicate that the deterrence hypothesis dominates any crowding out effects. However, heterogeneity in the treatment effect implies that other tools may be needed to further decrease non-attendance. Our results have high external validity since the data set covers both primary and secondary care for all publicly provided health care in Sweden. Hence, the policy implications can be generalized to other regions in Sweden.

The essay is structured as follows: in section 2, the theoretical aspects of sanctions for unwanted behavior are outlined along with previous research. Section 3 presents the data set, institutional setting of the Swedish health care system, and the econometric approach. Section 4 presents the results. Section 5 confirms the robustness of the results and finally, a concluding discussion takes place in section 6.

2. Previous research and conceptual framework

The theoretical viewpoints outlined in this section take departure from standard economic theory through the deterrence hypothesis formalized by Becker (1968). However, modern economic theory on economic behavior questions some of the assumptions made in the traditional economic theory. Theories on economic behavior argue that the assumption that preferences remain stable when an external incentive is introduced does not hold (Frey & Jegen, 2001). Alternative theories concerning crowding out effects and intrinsic motivation are discussed. Moreover, theories on experience learning are presented since researchers question the assumption that behavior changes equally with new information regardless of how information is presented (Haselhuhn et al., 2012; Agerval et al., 2014). Gneezy, Meier, and Rey-Biel, (2011) conclude in an overview of when and why monetary incentives work that economists must have a broader focus on the design and form of incentives. Hence, the introduction of monetary sanctions on non-attendance concerns traditional neoclassical economic theory as well as models inspired by the physiological and sociological research field.

2.1 Standard economic theory and the deterrence hypothesis

A basic assumption in neoclassical economic theory is that individuals are utility maximizing and consume a good until the marginal utility of that good equals the marginal cost. Hence, if an individual does not face the cost of his or her actions, there is an incentive to utilize too much of that good. In health care, patients are often covered by insurances and therefore have incentives to over utilize health care resources (Söderström 2008; Kruse & Ståhlberg 2013). Thus, patients do not face the full costs of their non-attendance and equilibrium non-attendance is therefore too high. Increasing the fee for non-attendance should therefore result in a lower non-attendance rate (Bech, 2005).

A famous paper by Becker (1968), *Crime and punishment: An economic approach*, with the aim to answer how punishment should be imposed to enforce legislation has been used to describe sanctions for unwanted behavior. The result is the deterrence hypothesis, described in equation 1, where the expected number of offences is a function of utility of offences (u), the probability of conviction (p), and the magnitude of punishment (f). Number of offences is related to the input variables as described in equation 2. Punishment (f), i.e. fee of non-

attendance, has a negative marginal effect on number of offences. The effect of increasing the fee for non-attendance can therefore be expected to be negative. Probability of conviction (p) also has a negative marginal effect on number of offences. In the case of non-attendance, the probability to be reported is probably close to one. Utility of offence (u) has a positive marginal effect on number of offences (Becker, 1968). For non-attendance, patients have a positive marginal utility as a result of the utility of the alternative action instead of attending or in the case of forgetfulness, less effort cost of remembering.

$$O = f(p, f, u) \tag{1}$$

$$\frac{\partial O}{\partial f} < 0,\ \frac{\partial O}{\partial p} < 0,\ \frac{\partial O}{\partial u} > 0 \tag{2}$$

Since this paper will review the effects of raising the fee for non-attendance, the input variable punishment will be investigated further. The optimal level of punishment (f) is the value that minimizes the social loss from the criminal behavior, see equation 3. This means that when deciding on the level of punishment, it should be a value that covers the marginal damage by the offence, $D'(O)$, and the cost of prevention and administration, $C'(O)$ (Becker, 1968).

$$D'(O) + C'(O) = f \tag{3}$$

2.2 Intrinsic motivation

It has been argued that the assumption in the *Deterrence hypothesis* that preferences remain stable when introducing an extrinsic incentive, may not hold (Frey & Jegen, 2011). The effectiveness of introducing an external incentive to change behavior has therefore in recent years been challenged. In modern research, which has integrated research in social psychology and economic theory, there is evidence that extrinsic motivation under some circumstances can crowd out intrinsic motivation (Gneezy, Meier, & Rey-Biel, 2011). Frey (2012) makes the distinction between extrinsic and intrinsic motivation, where the extrinsic motivation consist of factors from the outside, whereas the intrinsic motivation comes from within a person. The crowding out theory explains that an extrinsic incentive, such as a monetary sanction, ascribes a price for the behavior. A monetary sanction therefore crowds out intrinsic motivation because perception of the event changes (Gneezy, Meier, & Rey-Biel, 2011).

Introducing a monetary incentive has two consequences: the effect of the own material payoff by the deterrence hypothesis and the psychological effect on intrinsic motivation. The psychological effect is both the effect on preferences of prosocial behavior and the effect of the price signal on how a situation is perceived (Gneezy, Meier, & Rey-Biel, 2011). Prosocial behavior is explained as a mix of altruism, greed, concerns about social reputation, and self-respect (Bénanou & Tirole, 2006). In health care, patients can be expected to have internal motivation to do good and attend appointments and shame towards not attending. An external incentive such as paying a fee reduces prosocial behavior since motive behind the act is affected. The price signal on the other hand ascribes a price and therefore affects perception of the event by providing information regarding the task (Gneezy, Meier, & Rey-Biel, 2011). If the price signal does not support the perceived cost of non-attendance, patients will reevaluate their effort to remember appointments. It can be expected that knowledge regarding the actual cost of non-attendance is low. Patients do not face the full cost for their health care since a substantial part of all health care supply is financed publicly. If there is a high level of uncertainty about the cost of a service the price signal is strong (Frey & Oberholzer-Gee, 1997).

The psychological effect sometimes works in the same direction as the deterrence theory, this is called crowding in, and under some circumstances, it works against the deterrence theory, this is called crowding out. Crowding out occurs when external interventions are perceived as controlling and thus affect self-determination and self-esteem (Frey & Jegen, 2001), or if sanctions are regarded as unfair (Fehr & Rockebach, 2003). Hence, if patients perceive that the fee is unfair or that they are not trusted, intrinsic motivation is reduced. Crowding in, on the other hand, occurs when external interventions are perceived as supportive; this strengthens self-esteem and thereby also self-determination (Frey and Jegen, 2001). Sunstein (2003) exemplifies that crowding in can occur when there is a community effect. If the punishment signals that a large number of the other community members have stopped a behavior, this is more effective than stressing that the act is morally or legally wrong. Hence, if an increased fee signals that other individuals do not miss their appointments; intrinsic motivation to not miss appointments is crowded in. What effects dominate in the case of increasing the fee for non-attendance cannot be determined. However, remembering an appointment is costly in terms of effort to remember, it is therefore a task that is sensitive towards affecting intrinsic motivation (Gneezy & Rustichini, 2000B; Frey & Goette, 1999).

2.3 Information and experience

Models on intrinsic motivation relax the assumption that agents' perceptions of the situation remain constant in standard economic theory. Other researchers have also questioned the capability of processing of information. In the simple deterrence model by Becker (1968), individuals make decisions based on information regarding expected costs and benefits of behaving in a certain way. More recent research has shown that *how* individuals receive information of costs and benefits is of importance for the outcome. Most importantly, it matters whether the individual has knowledge of a fee or has experienced paying the fee (Haselhuhn et al., 2012; Agerval et al., 2014).

Studies by Haselhuhn et al. (2012) on video rental late return fees and Agarwal et al. (2013) on credit card fees, show that consumers act differently if they have knowledge of a fee by information compared to if they have recently experienced the fee. Additionally, there is a degree of forgetting, i.e. that the behavior changes drastically if consumers have experienced a fee recently and that the effect then levels off over time. The authors in both articles suggest that there is a high cost of paying attention and that paying attention is a scarce resource. When the memory of paying a fee is strong the benefit of paying attention feels high. As memory of that event becomes more distant, the effect of the fee on behavior fades (Haselhuhn et al., 2012; Agerval et al., 2014). The implication of this research on non-attendance fees is that whether patients have experienced paying the fee or simply have knowledge of the fee affects non-attendance rates.

2.4 Previous research on non-attendance fees in health care

Theoretical predictions of increasing the fee for non-attendance in health care are ambiguous. Meanwhile, the empirical literature on non-attendance fees in health care in particular is scarce. An experiment by Gneezy and Rustichini (2000A) at the daycare center shows that introducing a fee for late coming parents crowds out intrinsic motivation in a similar setting. Moreover, there are a several studies confirming the effectiveness of phone or text reminders to reduce non-attendance (Gurol-Urganci et al 2013; Reti 2003). However, only two studies are found regarding non-attendance fees in health care.

The first is a study on ophthalmological (eye diseases) patients in Finland. The study compares two months, one in 1992 and the second in 1993. In the latter year a fee for non-

attendance was implemented. The magnitude of the fee was not reported and the results of the study did not show any significant effects of imposing a fee on non-attendance (Mäntyjörvy, 1994). The results in this study should be interpreted with caution. The statistical method of finding causality is not robust since there can be other changes to the clinic between the measurement points. The second study is a comparison between the months before and after an implementation of the fee for non-attendance of 30 dollar at a community health care center in the US. The authors decided to only include clients who frequently missed appointments. Therefore, the sample of interest became very small and the group that was studied very specific. The result is a significantly lower non-attendance rate. The point estimates are 10.83 percentage points lower non-attendance rate (Lesaca, 1995). However, similarly to the study by Mäntyjörvy (1994), the statistical method does not account for time varying factors and the results are therefore not reliable.

Our study has two main advantages compared to the studies by Mäntyjörvy and Lesaca. Firstly, our data set covers not only one clinic but covers a large share of the total non-attendance for two counties and during a long time period. Secondly, the previous studies do not include a control group and hence, other general time trends are not controlled for. This unique data set gives us the opportunity to control for general time trend and compare the effect of raising the fee for non-attendance for different subgroups.

2.5 Theoretical implications

This section has outlined some of the previous findings and theoretical implications of monetary incentives on individual behavior. The theoretical predictions are ambiguous, traditional economic theory predicts that rational individuals reduce non-attendance rates after raising the fee for non-attendance. However, alternative theories of intrinsic motivation predict that raising the fee may work in another direction. Meanwhile, empirical evidence on non-attendance fees is scarce and only two studies have been found that investigate the subject. This leaves an open empirical question and it is therefore of importance to investigate the effect of raising the fee for non-attendance and thereby evaluate its cost effectiveness. The next section will describe the institutional setting and method that is used to test the effect of increasing the fee for non-attendance on rate of non-attendance.

3. Data and econometric approach

3.1 Institutional setting

The vast majority of physician visits in Sweden are financed by the public health care scheme and typically also performed by public providers (SKL 2015A; Regeringen, 2015). There are three levels of government in Sweden – central government, regional government and local government, where the regional governments, i.e. counties, are responsible for public health care organization. The provision of health care is to some extent regulated by the central government, but the counties have a high degree of freedom in the organization (SKL, 2015A). There are in total 21 counties in Sweden with a mean population size of 467,000 inhabitants. The two counties in our sample, Stockholm and Skåne, are the largest and third largest in terms of population size with 2.2 million and 1.3 million inhabitants respectively (SCB, 2016).

The regional governments, governed by elected politicians, are free to decide on fees for medical care visits. However, patient fees do not vary much between counties and are around 100-200 SEK for primary physician care and 200-350 SEK for specialist physician care (SKL, 2015B). The national government has decided that each patient pays a maximum of 1100 SEK in outpatient fees over a one year period. This rule does however not apply for non-attendance fees (Lag om läkarvårdsersättning, 1993:1651). Counties in Sweden are permitted to charge a fee for non-attendance according to national law (Lag om läkarvårdsersättning, 1993:1651). A common policy has been to charge a fee corresponding to the fee of attending an appointment (1177 Vårdguiden, 2016). This policy was employed for both Stockholm and Skåne County until January 2012. The regional government in Skåne County decided to increase the non-attendance fee to twice the patient fee for patients above 18 years old on January 1st 2012 (Regionfullmäktige Skåne, 2011). There is no indication that the fee increase in Skåne is a result of more extensive problems with non-attendance in Skåne than in other counties or that the intervention was part of a larger intervention to control non-attendance rates. Hence, we have no reason to suspect endogeneity in treatment. Skåne Region did however inform about the reform on their website, in hospitals and mass media[2]. The fees for

[2] Personal communication with health care strategist Pia Landgren, who has been involved in the work with patient fees in Skåne county during the study period.

children under the age of 18 follow other regulations. Therefore, this study only concerns individuals above the age 18.

The resulting relationship between non-attendance fees and patient fees before and after 2012 for Stockholm and Skåne County are presented in Table 1 below. Patient fees have changed slightly upwards as well as downwards in both Stockholm and Skåne County during the study period. In primary care, patient fees are 40 and 50 SEK higher after the reform compared to before in Stockholm and Skåne respectively. In specialist care, fees were not changed in Skåne during the study period, whereas Stockholm increased fees by between 20 and 40 SEK at three different times. Although we cannot exclude that this affects our result, we expect the effect to be small due to the small change in price. In addition, the largest change was in both counties simultaneously in 2012.[3] A full list of patient fees over the study period is presented in Table 12 in the appendix.

Table 1 Fees for non-attendance

	Stockholm	Skåne
Pre treatment 2008-2011	1×Patient fee	1×Patient fee
Post treatment 2012-2015	1×Patient fee	2×Patient fee

3.2 Data and variables

We use a combined panel data set with register data from Skåne and Stockholm County covering the period 2008-2015. The register data includes observations on attendance and non-attendance to health care appointments. Each observation is a booked appointment, which the individual attends or not attends to. Emergency care is not included and nor are telephone appointments. As mentioned in the introduction, non-attendance is defined as not attending the appointment without cancelling at least 24 hours before or having a legitimate cause for missing the appointment. Consequently, only missed appointments that correspond

[3] The full list of patient fees covering the study period was received from Pia Landgren, Health care strategist in Skåne County and Gunnar Ljunggren, Chief physican and medical advisor in Stockholm County. For email conversation, please ask the authors. The contact persons in Skåne and Stockholm have moreover confirmed the rules for non-attendance fees described in this section.

to this definition are registered as non-attendance in the system. Individuals are identified with a unique ID which allows us to follow patients over time. The data set includes information on patient ID, gender, age, if patients attended the appointment, and time of appointment. This unique data set is collected separately from both counties and an analysis on non-attendance between counties on individual level has to our knowledge not been made before.

The data sets from the two counties are not completely overlapping. In Stockholm County, the data set contains outpatient primary and secondary care outside the hospital for all care providers i.e. both physicians and nurses. In Skåne County, the data set contains all primary and secondary care, but only covers physician care. It is not possible to exclude non-physician appointments in Stockholm since some of the physician appointments are registered as non-physician in the event of non-attendance. Inpatient care, where patients are enrolled in the hospital, is not included in any of the samples since it is covered by other fee systems. Additionally, since the counties have some freedom in the health care organization, register routines and computer systems differ. A direct mean comparison between Stockholm and Skåne County is difficult since the data sets are not identical. However, since the differences between Skåne County and Stockholm County are generated by differences in the register systems and to some extent different care providers, we argue that the differences in registering and behavior when receiving care from different providers can be expected to be constant over time. The two counties can therefore be used as counterfactuals in a difference in difference analysis where we control for county fixed effects.

The data set from Stockholm contains 172 million appointments and from Skåne 39 million appointments. For our main sample, 6 million observations (appointments) from each county are randomly drawn due to capacity constraints in the statistical software. The dependent variable is an indicator of non-attendance which is 0 if patients attended their appointment and 100 if patients do not attend their appointment. The choice to scale the dependent variable by 100 is to ease the interpretation. Woman is also a binary variable that is 1 for women and 0 for men. The mean and standard deviation for the dependent variable, non-attendance, and control variables are presented in Table 2. Mean age and fraction of women are similar in both counties, which provides further evidence to the assumption that the data sets are comparable.

Table 2 Descriptive statistics

Variables	Skåne			Stockholm		
	Obs	Mean	Std	Obs	Mean	Std
Non-attendance	6000000	1.677	0.128	6000 000	1.511	0.122
Age	6000000	57.001	20.151	5998587	57.566	20.645
Women	6000000	0.585	0.493	6000000	0.603	0.489

3.3 Difference-in-difference estimation

The difference in difference (DiD) method is used to estimate the treatment effect of raising the fee for non-attendance. This method is used when the regressor of interest is on group level rather than on individual level. If there is a policy change on group level with two groups where one is treated while the other is not, the DiD approach can estimate the effect of the policy change by controlling for unobserved time-invariant group level characteristics using group fixed effects and the common time trend using time fixed effects (Angrist & Pischke, 2009). Decisions regarding health care are made both by regional county governments and the national government (SKL, 2015A). Furthermore, counties are affected by the same national specific macroeconomic shocks and other changes in national laws and regulations. Hence, this method makes it possible to control for the general time trend that arises from decisions and shocks on national level. The DiD is therefore used to evaluate the effect of raising the fee for non-attendance by using Skåne County as the treatment group and Stockholm County as control group.

To estimate the DiD model, we use the following standard DiD equation:

$$Non-attendance = \alpha + \beta Skåne + \gamma Skåne * Post\ Reform + \boldsymbol{Year'\delta} + \varepsilon \qquad (4)$$

The dependent variable is an indicator of non-attendance and the treatment variable is a binary variable that is equal to 1 for Skåne County after 2011, i.e. when the fee was doubled. A vector of dummy variables for years capture a common time trend and the binary variable that is 1 for Skåne County captures time-invariant differences between the two counties. Equation 4 is also estimated including the two available control variables: gender and age, where age is included both linearly and in age groups. We choose to include age groups

instead of higher orders of the age variable since it gives a more intuitive interpretation of the result.

Ordinary least squares, OLS, is used to estimate all models. Some difficulties arise when using OLS with limited dependent variables because the estimated regressions do not account for limits. It can therefore be argued that a non-linear probit model should be used when the dependent variable is binary. However, the advantage with OLS compared to a probit model is on the other hand that the interpretation of the coefficients is easier and fewer assumptions regarding the function are needed (Angrist & Pischke, 2009). OLS will therefore be used to estimate the treatment effect. As a robustness test, the preferred model is estimated using a probit model, presented in section 5.

The identifying assumption in the use of DiD is the common trend assumption. It states that although Skåne County and Stockholm County may be different in many aspects, these differences are constant over time and hence captured in the county fixed effects. The time trend on the hand is common and is captured by time fixed effects. This assumption cannot be tested statistically and a threat to the common trend assumption is if the counties have worked differently over time with other tools to decrease non-attendance. After being in contact with public officials in both Skåne and Stockholm we are convinced that no structural changes were made in either county regarding interventions to reduce non-attendance. Health care providers can remind patients by text but the patient's consent is needed (Socialstyrelsen, 2012). Moreover, in both counties health care providers themselves decide on how to organize the booking and whether patients should be reminded. Since this is not decided on regional level, no structural differences can be expected between counties. The common trend assumption is therefore expected to hold. The four years before the reform provide an opportunity to compare the pre-reform trends. The non-attendance rates for Stockholm and Skåne over the time period can be viewed in Figure 1. The Figure confirms a common increasing trend for both Stockholm and Skåne the four years prior to the reform in 2012, but a persistent drop in non-attendance rates for Skåne after the reform. Hence, there is no reason to believe that the common trend assumption is violated based on the sample time period.

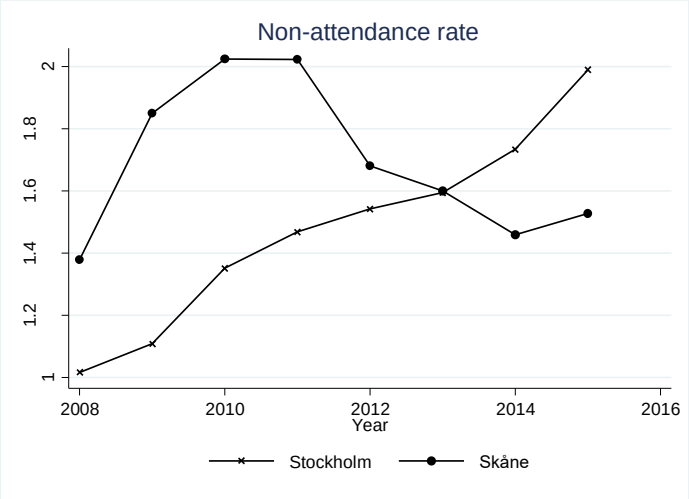

Figure 1 Common trend assumption

3.4 Treatment heterogeneity

An analysis on subsamples will help in understanding heterogeneity in the treatment effect and thereby improve policy design. Firstly, we explore the difference in treatment effect between age groups. This is further explained in section 3.4.1. Secondly, individuals that missed appointment in the past and their probability to miss the next appointment are investigated separately. This proceeding requires new sample selections and is described further in section 3.4.2 and 3.4.3. Table 3 below presents average non-attendance rates in Skåne before and after the change of non-attendance fee for the main sample and different subsamples of the population. The age intervals are decided according to commonly used divisions and give us the possibility to divide between working age and pensioners. Average non-attendance is lower after the reform for all samples; this will be investigated further in the statistical analysis. The choice of subsamples will be discussed in the section below.

Table 3 Mean non-attendance rates Skåne

Sample groups	Pre treatment average non-attendance	Post treatment average non-attendance
Main sample	1.829	1.559
18-24	5.724	4.988
25-44	3.502	3.127
45-64	1.515	1.277
65-84	0.442	0.375
85+	0.275	0.233
Non-attending before reform	8.197	4.389
Missed last appointment	24.433	24.277

3.4.1 Subsample with age groups

The differences in non-attendance rates between age groups, which can be seen in Figure 2 below, make it interesting to look at heterogeneity in the treatment effect. Figure 2 presents non-attendance by age for Skåne and Stockholm. The pattern is similar in both counties; a relatively high non-attendance rate for young adults and decreasing non-attendance by age. The same pattern can be seen in Table 3 above. Due to the difference in baseline non-attendance, these groups may react differently to raising the fee for non-attendance due to factors in their characteristics such as medical state, payment ability, and time availability. Subsamples of the five age groups will therefore be used in the statistical analysis.

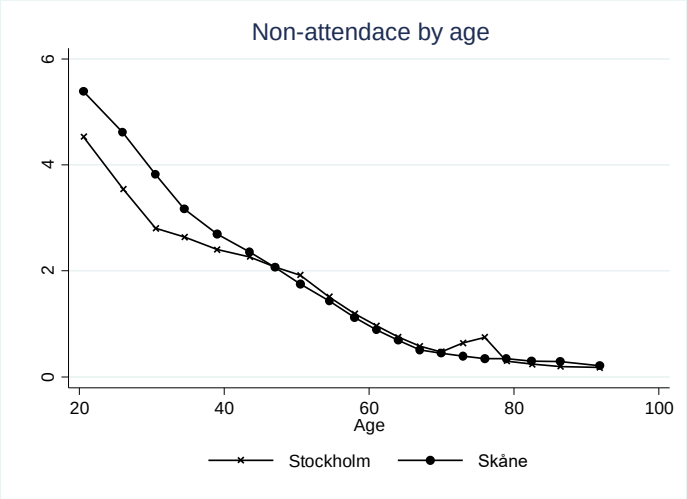

Figure 2 Age trends

3.4.2 Subsamples with non-attending individuals

Heterogeneity in the treatment effect is investigated in a subsample including only individuals who missed at least one appointment before the reform. This sample is used since the problem with non-attendance is possibly a result of some individuals who account for a large share of non-attendance. To know the effectiveness of the reform for this group is therefore important. Differences in characteristics between individuals that have missed at least one appointment and individuals that have never missed an appointment may affect reactions to increasing the non-attendance fee. Additionally, a random sample on individuals who have missed appointments in the past makes our study more comparable with the study by Lesaca (1995), which only included individuals with previously missed appointments.

We created this subsample from the original data by including only individuals that missed at least one appointment before the policy change, i.e. 2008-2011. In this sample construction, the same individuals in each county are followed before and after the reform. Similar to the data set for the main sample, we draw 6 million observations from each county. The non-attendance rate, presented in Table 3, is on average 8.18 % in Skåne before the reform and 4.38 % after the reform. The high non-attendance rate before the reform is partly due to the sample selection, since only individuals with at least one missed appointment before the reform are included. However, non-attendance rates are still higher than the main sample after the reform which confirms that this is a group that is important to reach when designing policies for non-attendance. The choice to only select on non-attendance before the reform is because non-attendance after the reform is a part of the outcome. Hence, making the selection on non-attendance behavior after the reform changes the composition of the treatment and control group. A change in composition as a result of the treatment of interest is one of the pitfalls in choosing treatment and control groups in a DiD estimation (Angrist & Piscke 2009). An additional advantage of this setting is that the same individuals are followed before and after treatment, which means that the difference in composition of individuals between counties is kept constant.

3.4.3 Subsample investigating experience learning

Heterogeneity in treatment is further investigated by selecting a new sample including all observations where individuals missed the appointment before. This sample is used to investigate whether the effect of paying a non-attendance fee affects the probability of attending the next appointment differently with the double fee compared to the single fee. The

non-attendance rate in Skåne for these observations is 24.43 % before treatment and 24.27 % after treatment. This indicates that individuals that missed one appointment have a high probability of missing the next. It is therefore of interest to reduce non-attendance for this group. Research on learning by experience provides some evidence that there is a learning mechanism of paying a fee that is stronger than just knowledge of a fee. Using a DiD approach allows us to estimate whether this effect is stronger with a higher fee.

Due to a smaller sample, all observations meeting the criteria are included; it is therefore not necessary to draw a random sample for this group. 1 669 582 observations in Stockholm and 604 433 observations in Skåne fit the criteria that the individuals missed the appointment before. An issue for this sample is the problem discussed in the previous section regarding choosing composition of the sample on the outcome variable. If individuals react on the higher fee, the composition of treatment and control group changes a consequence of the reform. Hence, individuals that missed their last appointment before and after the reform in Skåne and Stockholm are not perfect counterfactuals. Individual characteristics that differ between the groups might therefore be correlated with treatment. We argue that this result is still useful since it is of interest to understand if patients react differently when experiencing the fee before and after the reform, regardless if the group composition changes. By using this sample, more knowledge is thereby provided on patient behavior and how to communicate and design policies for non-attendance.

4. Results

This section presents the results from an OLS regression using a DiD design with Skåne as treatment group and Stockholm as control group. The treatment effect is first tested for the main sample that includes the whole population to estimate the general policy effects of doubling the non-attendance fee. The main results are discussed along with back of the envelope calculations to gain a better understanding of the possible cost reductions of the reform. In the following section, treatment heterogeneity is estimated using the different subsamples that were presented in section 3 since the initial analysis indicated heterogeneity in non-attendance.

4.1 Main sample

Table 4 below presents the difference in differences results estimated with equation 4. The model is estimated without control variables in column 1 and with control variables for age and gender in column 2. The model is lastly estimated controlling for age groups, as presented in column 3, to allow for a non-linear correlation between non-attendance and age. In the third model specification, age is included to control for age effects within each age group. The treatment effect coefficient, *Skåne*Post Reform*, is significant in all models. The effect size is -0.752 in the model in column 1, but decreases slightly in absolute value to -0.711 in column 2, when controlling for gender and age. The coefficient is almost identical, -0.705 when allowing for non-linearity in age. Since the estimated treatment effect does not change noticeably, the treatment effect is robust towards changes in model specification. This means that increasing the non-attendance fee to the double lowers non-attendance rates with around 0.7 percentage points in primary and secondary care. For further analysis on cost reductions and treatment heterogeneity, we use the model in column 2 due to the small coefficient change when including age groups. Since the coefficient size for the treatment effect represents the percentage point change in non-attendance rate, a transformation to relative change allows for a more intuitive interpretation of the magnitude of the effect. Table 5 below presents the values transformed to percentage changes in relation to mean non-attendance rate during the years 2008-2011 before the reform. The reaction to the policy change of increasing the non-attendance fee to double the patient fee is a change in non-attendance rate of between 41.12% and 38.77% depending on the model specification.

Table 4 Main sample: DiD results

VARIABLES	(1)	(2)	(3)
Skåne*Post Reform	-0.752***	-0.711***	-0.705***
	(0.015)	(0.014)	(0.014)
Skåne	0.585***	0.518***	0.529***
	(0.011)	(0.011)	(0.011)
Women		-0.383***	-0.429***
		(0.007)	(0.008)
Age		-0.065***	-0.060***
		(0.000)	(0.001)
Age groups			
25-34			-1.082***
			(0.028)
45-64			-1.461***
			(0.034)
65-84			-1.301***
			(0.042)
85+			-0.568***
			(0.050)
Year fixed effects	Yes	Yes	Yes
Observations	12,000,000	11,998,587	11,998,587
R-squared	0.000	0.012	0.013

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 Relative change in non-attendance

Model	(1)	(2)	(3)
Mean non-attendance in Skåne (2008-2011)	1.8287	1.8287	1.8287
Coefficient	-0.752***	-0.711***	-0.705***
Percentage change compared to mean	41.12%	38.88%	38.55%

The results are in line with the predictions from deterrence theory, where the number of offences depends on utility from offence, magnitude of punishment, and probability of getting caught. An increased punishment must hence result in lower non-attendance rates. Meanwhile, patients might still react to increasing the fee for non-attendance regarding their perception of the event both in terms of changed prosocial preferences and regarding how the

situation is perceived. However, the large negative effect of increasing the fee indicates that the effects of material payoff, in accordance to the deterrence theory, the signaling value, and any crowding in effects, dominate potential crowding out effects. The reason for why crowding out intrinsic values are dominated by the deterrence hypothesis is possibly that patients do not consider raising the fee for non-attendance as a new punishment but rather an increased punishment. The difference between the two is whether an external incentive is introduced or merely raised. If an external incentive already exists, the price signal has already set a market value for the service and patients no longer feel obligation, shame, and guilt, among other feelings of prosocial behavior, for non-attendance. Gneezy and Rustichini (2000B) investigate this using an experimental setting in the paper *Pay enough or don't pay at all,* where they show that introducing a monetary incentive crowds out intrinsic motivation to perform a task, but that increasing the payment leads to better performance compared to the low payment. As the authors argue, when introducing a fee it should be large enough or it serves in contrast with its purpose. Hence, if intrinsic motivation is already crowded by the single fee, doubling the fee should have the effect that is predicted by deterrence theory. If instead the value above the usual patient fee is perceived as the punishment, a new external incentive is introduced, which might cause larger crowding out effects.

The uncertainty regarding the cost of non-attendance is expected to be high. When the fee is doubled it gives a price signal that affects patients' perception of waste and cost of non-attendance and supports a higher intrinsic motivation since increasing the fee for non-attendance to the double fee is more coherent to the actual cost of non-attendance. Frey and Jegen (2011) conclude that monetary incentives can be expected to crowd in intrinsic motivation if they are perceived as supportive. If a fee is regarded as fair, a monetary incentive can crowd in intrinsic motivation. According to previous research by Sunstein (2003), the community effect is often more effective than moral incentives if it can be made credible that everyone else makes effort to attend their appointments. By doubling the fee, it is more likely that others attend their appointments, which in turn influences more patients to attend their appointments. In summary, the large effect of doubling the non-attendance fee on non-attendance rates is a consequence of that intrinsic motivation is already crowded out by a single fee and the external incentive therefore works in the same way as predicted by the deterrence theory. In addition, if increasing the fee crowds in intrinsic motivation, the double fee can be expected to result in even lower non-attendance rates than expected from the deterrence hypothesis.

Another perspective that has been neglected in previous work on intrinsic motivation is the reaction from health care providers. Increasing the fee for non-attendance carries a price signal in the health care organization similarly to the one for patients. Increasing the fee leads to more focus and more information regarding the costs of missed appointments in the health care organization. This affects motivation by care providers to give more information to patients and to help remind them. Hence, the change might not only be due to behavioral changes among patients, but also behavioral changes among care providers.

An additional effect from increasing the fee for non-attendance is the effect of the reform on media coverage by local newspapers. The new rules for non-attendance fees were presented in all larger local newspapers in Skåne. Thus, the large decline in non-attendance may also be a result of increased information and that it draws attention to waste of tax money from non-attendance among both patients and health care employees. The reform brought up the issue and helped signaling that non-attendance is wasteful and costly by gaining media coverage. It cannot be expected that the extent of media coverage can be generalizable to all interventions that increase non-attendance fees in Sweden, especially since Skåne is one of the largest counties in Sweden. However, as we see in Figure 1, the drop in non-attendance rate is persistent and declines further three years after the reform, which indicates that any effect of the initial media coverage is relatively small.

4.1.1 Cost reductions from treatment

The cost reductions that can be expected from doubling the fee for non-attendance in Skåne County are based on the treatment effect in Table 4 from equation 4 controlling for gender and age, the cost of non-attendance, number of appointments in 2014, and the new non-attendance fees. The number of appointments is reported in the annual report for Skåne in 2015 (Region Skåne, 2016). The cost of non-attendance depends on whether capital resources and human resources can be utilized for something else instead of the appointment and whether another patient could come instead. To estimate the cost of non-attendance the average costs for one appointment in primary care and secondary care in Skåne are used, these were 1409 SEK and 3168 SEK respectively in 2014 (Kolada 2016a, 2016b). Meanwhile, the new non-attendance fees are 320 SEK in primary care and 600 SEK in secondary care (1177.se, 2016). The resulting cost savings from the reform are estimated in Table 6.

To what extent the cost of non-attendance corresponds to the actual costs of attending the appointment is unknown using this data set. Hence, cost savings are estimated for three scenarios: cost of non-attendance equals to 100 %, 75%, or 50% of the total cost of the appointment. The net cost reductions account for a reduction in cost of non-attendance and differences in earnings from doubling the fee for non-attendance. As presented in Table 6, the net cost reductions, accounting for increased earnings, are 155.8, 121.1, and 69.2 MSEK in the three cases[4].

Table 6 Cost savings

Type of care	Average cost per visit	Number of appointments	Treatment effect	100 %	75 %	50 %
Primary care	1 409	7 567 196	0.711	75 808 094	56 856 070	28 428 035
Secondary care	3 168	2 780 500	0.711	62 629 317	46 971 987	23 485 994
Net earnings				17 320 315	17 320 315	17 320 315
Total net cost reductions including net earnings				155 757 726	121 148 372	69 234 344

Note- The treatment effect is given in percentage points. 0.0071 is hence used in the calculation. Number of appointments includes both physician care and other care providers. Net earnings regard the increased revenue from higher patient fees.

The calculated cost reductions are merely back of the envelope calculations. There are several factors that are not accounted for. Among these are social costs of increasing the non-attendance fee, such as increased stress, and administrative costs, such as rescheduling appointments, handling complaints, and invoices. The cost of the possibility to use resources more efficiently that results in shorter waiting times, less prolonged illness, and less production losses, is also not accounted for. On the income side, we lack knowledge regarding to what extent the fees are actually paid. Moreover, the optimal fee for non-attendance according to the deterrence theory, is the one that covers the administrative costs of non-attendance as well as the social costs of non-attendance (Becker, 1968). The new fees cannot be expected to cover the full costs of non-attendance in Skåne. Since patients do not bear the full burden of their actions, the non-attendance rate is therefore still not in

[4] The resulting cost reductions are calculated by using the treatment effect, -0.711, and estimating the cost reductions by using the cost of an appointment, 1409 SEK and 3168 SEK, and number of appointments, 7 567 196 and 2 780 500, in primary and secondary care respectively. It is not possible to distinguish any differences in treatment effect between primary and secondary care in this sample, the same rate is therefore used. The increased earnings as a result of increasing the fee for non-attendance are estimated using the mean non-attendance rate in 2012-2015 in Skåne which is 1.558 % and the difference in non-attendance fee compared to before which is 160 SEK for primary care and 300 SEK for secondary care. Similarly to the treatment effect, data in this sample cannot be used to determine any differences in non-attendance rate between primary and secondary care, the same rate is therefore assumed. The resulting gained earnings are 18 863 506 SEK in primary care and 12 996 057 SEK in secondary care. Meanwhile, earnings are reduced by 0.711 percentage points because of the effect of treatment. This is a reduction of earnings 10 255 064 SEK in primary care and 5 930 807 SEK in secondary care. The resulting net earnings effect is 17 320 315 SEK. The net cost reductions, accounting for increased earnings, are 155 757 726, 121 148 372, or 69 234 344 SEK in the three cases of costs of non-attendance.

equilibrium. Nevertheless, these calculations indicate large cost savings as a result of doubling the fee for non-attendance in Skåne. The next section will examine heterogeneity in the treatment effect.

4.2 Treatment heterogeneity

4.2.1 Subsample with age groups

Age groups will be investigated for heterogeneity in the treatment effect since this was suggested by the descriptive statistics in Table 3. Table 7 presents the results of the DiD regression analysis for different age groups. The model used is equation 4, including age and gender as control variables. Age is included to pick up the age effect within the age groups. The treatment effect is the largest for patients between 18 and 24 of -1.922 percentage points. The treatment effect is still large for those between 25 and 44, about -1.061 percentage points, and decreases slightly to -0.867 for those between 45 and 64. For the patient group 65-84, the treatment effect is -0.270 and -0.074 for those above 85 years old. The relative changes in non-attendance rates are presented in Table 8. Although the coefficient size is larger for the age group 18-24, the percentage change is the largest for the age group 65-84. Note that both mean non-attendance and relative treatment effect is the lowest for those above 85 years old.

Table 7 Age groups: DiD results

	(1)	(2)	(3)	(4)	(5)
VARIABLES	18-24	25-44	45-64	65-84	85+
Skåne*Post Reform	-1.920***	-1.033***	-0.866***	-0.269***	-0.075***
	(0.098)	(0.041)	(0.026)	(0.014)	(0.018)
Skåne	1.917***	1.158***	0.402***	0.048***	0.107***
	(0.072)	(0.030)	(0.018)	(0.010)	(0.014)
Women	-0.859***	-0.930***	-0.510***	-0.093***	-0.035***
	(0.052)	(0.022)	(0.013)	(0.007)	(0.010)
Age	-0.184***	-0.099***	-0.083***	-0.018***	-0.008***
	(0.012)	(0.002)	(0.001)	(0.001)	(0.001)
Year fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	778,183	2,813,828	3,445,115	3,862,475	1,098,986
R-squared	0.002	0.002	0.003	0.001	0.000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8 Percentage change in non-attendance by age group

1. Age group	18-24	25-44	45-64	65-84	85+
Mean non-attendance in Skåne (2008-2011)	5.724	3.502	1.515	0.442	0.275
Coefficient	-1.920***	-1.033***	-0.866***	-0.269***	-0.075***
Percentage change compared to mean	33.5%	29.5%	57.1%	60.9%	27.3%

The treatment effect varies but is strong and significant for all age groups. Hence, although there is heterogeneity in treatment, the intervention is effective for all age groups. This is of importance since young adults, on one hand, have a high non-attendance rate and means to reduce it are needed. Elderly on the other hand, have low non-attendance rates but utilize health care to a larger extent than working age. An additional decrease in non-attendance rate for this group can therefore have a large impact in absolute numbers. The reasons for heterogeneity in the treatment effect can only be speculated on. The group that react the strongest, those between 65 and 84 years old, possibly have higher price elasticity since they are pensioners and therefore receive less money each month. The group that react the least to the change in non-attendance fee are those above 85 years old, they might already have non-attendance at a minimum level, where benefits of non-attendance are larger than most fees since it is so low. The remaining non-attendance might be a result of acute illness, dementia, postal problems, or other reasons to miss appointments; hence it cannot be reduced by increasing the non-attendance fee further. The first age groups, those between 18 and 44, decrease their non-attendance rates less in relative values than those between 45 and 84. Generally, non-attendance is a decreasing function of age, and it may be so that those between 18 and 44 are not particularly sensitive towards neither intrinsic nor extrinsic motivation and a busy schedule should not be neglected to overrule financial incentives. Since some age groups do not react as strongly towards using financial incentives to decrease non-attendance, additional policies are advised. For example, some groups may be particularly sensitive towards decreasing effort cost of remembering by using text reminders.

4.2.2 Subsample of non-attending individuals

We study individuals who miss appointments often since it is suspected that there is substantial heterogeneity in number of missed appointments and that some individuals

account for a large proportion of non-attendance. This is apparent in this sample of individuals who have missed at least one appointment during 2008-2011. Non-attendance for this group compared to the main sample is considerably higher, as have been shown in Table 3. The results from regression output of equation 4 with control variables for age and gender are presented in Table 9. The treatment effect is -3.234 percentage points difference in non-attendance rate as a result of doubling the non-attendance fee.

Table 9 Non-attending individuals: DiD results

VARIABLES	(1)
Skåne * Post Reform	-3.332***
	(0.025)
Skåne	3.625***
	(0.020)
Women	-0.957***
	(0.013)
Age	-0.135***
	(0.000)
Time fixed effects	Yes
Observations	11,999,769
R-squared	0.022

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The absolute percentage point change is large, and considering that the mean non-attendance rate before the reform is higher in this sample due to sample selection, see Table 3, there is a large treatment effect relative to non-attendance rate as well. This subsample is more relevant for a comparison to the previous study by Lesaca (1995) than the main sample, since his study only includes individuals who are especially prone to non-attendance. A 30 USD increase in non-attendance fee decreased non-attendance by 10.83 percentage points. The point estimates can still not be directly compared but we note that there is a large effect of increasing the non-attendance fee in this sample as well, and the percentage point increase in this group is larger than in the main sample. According to neoclassical theories, patients will fail to attend appointments if costs of non-attendance are higher than benefits. For this group, there seems to be a substantial number of individuals where costs of non-attendance following the reform are higher than benefits. A larger share of this group of patients may have an effort cost of remembering just above the single patient fee and when the non-attendance fee is doubled

their net benefit of non-attendance turns negative. Additionally, since this group had non-attendance prior to the reform, this group may naturally have little intrinsic motivation to attend appointments, which is why the treatment effect is larger for this group than the main sample where crowding out can be expected to be larger.

4.2.3 Subsample investigating experience learning

By using the fact that we can follow specific individuals over time, the effect of the reform on probability of non-attendance is estimated given that the individual missed the appointment before. Hence, this sample makes it possible to estimate the effect of experiencing a higher fee, rather than information about a higher fee, in relation to a lower. The regression output for equation 4 including control variables for age and gender is presented in Table 10. The treatment effect is -0.925 percentage points. It cannot be distinguished if it is a causal effect of doubling the fee or if it is patient characteristics that correlate with treatment that cause the effect. It is however interesting to conclude that given the group that missed their last appointment, the change in probability of non-attendance only changes by -0.952 percentage points as a result of the reform. This can be compared to the mean non-attendance rate for those who missed their last appointment in Skåne prior to the reform, in Table 3, which was 24.43 %.

Table 10 Experience learning: DiD results

VARIABLES	(1)
Skåne * Post Reform	-0.925***
	(0.131)
Skåne	-7.976***
	(0.099)
Women	-0.257***
	(0.002)
Age	-2.452***
	(0.060)
Time fixed effects	Yes
Observations	2,273,410
R-squared	0.016

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

These results indicate that those who still choose to not attend do not seem to react differently to experiencing a single fee compared to double fee. The results from the main sample indicate that information regarding the increased fee for non-attendance has a large deterrence effect on non-attendance rates. Yet, the effect of experiencing a fee does not seem to differ between a double and a single patient fee. The findings by Haselhuhn (2008) and Agarwal et al. (2013) regarding late fees in the video rental store and credit card overdraft fees, find that there is a learning dynamics from experiencing a fee. We do not find evidence for a stronger experience learning effect with a higher fee. From a policy perspective, this indicates that individuals that still miss appointments after the policy do not have a strong learning pattern from paying a fee. Therefore, other measures to lower non-attendance are probably needed for this group. However, as discussed in section 3, the sample selection for this model is somewhat problematic and we therefore interpret the result with caution.

5. Robustness testing

The treatment effect from the DiD estimation using the main sample is tested for robustness of the results. It has already been proven that the results are not sensitive to using different subsamples of age groups. Additionally, Table 4 proves that the treatment effect does not change substantially when including control variables and including non-linear age control groups. Assuming a valid research design, where the conditional independence assumption is not violated, including control variables should only reduce standard errors and not change the coefficient for the treatment effect (Currie & Walker, 2011). Further robustness tests are presented in Table 11 below. For the following robustness tests, equation 4 will be used and control variables for age and gender are included.

A placebo test is performed to test if anything questionable is going on before treatment took place (Blackwell, 2013). The results are presented in column 1 in Table 11. We use the four years in Stockholm and Skåne before the policy reform, 2008-2011. The placebo test assumes that treatment takes place in the beginning of year 2011 instead of 2012. The placebo coefficient, *Skåne* Post Placebo,* is not significant which verifies that nothing strange is going on before treatment. Column 2 presents the results from estimating a probit model. As discussed in section 3, oppositely to the OLS, the probit model accounts for limits. The estimated treatment effect is significant and negative, which provides evidence for that the treatment effect is robust towards using a nonlinear model. The coefficient for treatment is not the marginal effect of treatment and the coefficient size can therefore not be directly compared to the OLS estimates. In column 3, the model allows for different time trends for the two counties. This is a relaxation of the common trend assumption. The treatment effect becomes considerably smaller, -0.367 percentage points, when allowing for different trends. This effect may arise from that the slope in Skåne before 2012 is slightly steeper than in Stockholm. The treatment effect is still significant with 20 % lower non-attendance compared to before the reform.

Table 11 Robustness tests

VARIABLES	(1) Placebo	(2) Probit	(3) Different slopes
Skåne * Post Reform		-0.194***	-0.362***
		(0.004)	(0.030)
Skåne* Post Placebo	-0.024		
	(0.025)		
Skåne	0.528***	0.144***	174.445***
	(0.012)	(0.003)	(12.919)
Woman	-0.291***	-0.119***	-0.383***
	(0.011)	(0.002)	(0.007)
Age	-0.063***	-0.018***	-0.065***
	(0.000)	(0.000)	(0.000)
Skåne*Year			-0.087***
			(0.006)
Year fixed effect	Yes	Yes	Yes
Observations	5,313,786	11,998,587	11,998,587
R-squared/Pseudo R-squared	0.012	0.075	0.012

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The robustness tests in this section and the regression output in the results section confirm that the treatment effect is robust to choice of estimation model, relaxation to the common trend assumption, and different subgroups. The placebo tests provides further evidence for that the treatment effect is an actual consequence of the reform and not a coincidence. Additional robustness tests are performed and presented in appendix. The model is estimated controlling for age with a different non-linear specification and subsamples of women and men in Table 13, and with standard errors clustered on individual level in Table 14. These changes in model specification do not change the significant effect of treatment. In summary, the estimated treatment effect in the main model has been exposed to numerous robustness tests and the coefficient does not change substantially. We therefore conclude that there is robust evidence for an effect of increasing the fee for non-attendance on non-attendance rates in primary and secondary care.

6. Concluding discussion

This paper has studied the effectiveness of increasing the fee for non-attendance in Swedish health care. Health care expenditures have been growing in Sweden the past years and non-attendance is one issue that causes unnecessary costs and inefficiencies. The theoretical predictions of raising the fee are not straightforward since a fee can be expected to affect both extrinsic and intrinsic motivation. Empirical research on monetary sanctions gives ambiguous results and empirical evidence for non-attendance is scarce. Hence, the effect of raising non-attendance fees is an open empirical question. By using a unique panel data set with administrative data on patients' visits in health care from two counties over the years 2008 to 2015, we exploit the natural experiment that one county doubled the non-attendance fee in 2012, and the other had a constant fee. It allows us to draw causal inferences of doubling the fee for non-attendance on non-attendance rates in health care. The results are a drastic decrease in non-attendance rate, around 39 % lower than the average non-attendance rate before treatment. By using back of the envelope calculations, the resulting cost reductions in Skåne as a result of the reform are 155.8 million, 121.1 million, or 69.2 million SEK annually depending on if the cost of non-attendance is 100 %, 75 %, or 50 % of the cost of attending an appointment in medical care. We find heterogeneity in the treatment effect but the intervention is effective for all subsamples. The intervention is especially effective for those who have missed at least one appointment before the fee change.

These findings suggest that the deterrence theory along with crowding in effects dominate crowding out intrinsic motivation when the fee for non-attendance is doubled. That the deterrence hypothesis dominates crowding out intrinsic motivation is a result of several factors. Crowding out can be expected to be smaller since the fee is raised rather than introduced. We also suspect that crowding in intrinsic motivation, as a result of that it can be made credible that others attend their appointments, along with signaling within the health care production regarding costs of non-attendance, and increased media coverage, are important factors for why increasing non-attendance fees is so effective. However, it is important to be careful when designing the policy. Heterogeneity in the treatment effect implies that additional policies to lower effort cost of remembering are needed to further lower non-attendance rates.

We argue that the external and internal validity is high. However, we are aware of the threats to internal validity due to small changes in fees, other regional reforms towards non-attendance, and differences in registration patterns. As has been argued, we do not expect that these threats affect our results considerably. In addition, the internal validity has been assured through various robustness tests. The external validity is strong since the data set covers both primary and secondary outpatient care for all public health care providers in each of the two counties and that the county regions in Sweden have similar policies and health care organizations. We conclude that our result provides policy implications for both primary and secondary care to other counties in Sweden.

The findings in this paper provide new causal evidence for that doubling the fee for non-attendance in health care reduces non-attendance rates substantially. This allows for large cost reductions for counties in Sweden. To further build knowledge on non-attendance in health care in Sweden, some questions remain unanswered. Firstly, the learning dynamics of a fee for non-attendance is interesting since this sample indicates that health care consumers do not have the same learning by experience pattern that has been found on consumers in other sectors. Secondly, the cost reductions from decreasing the non-attendance rate depend on the possibility to use resources and personnel for other appointments. If an individual calls 24 hours before, it is plausible that there are difficulties in reallocating resources in a useful way with short notice. By distinguishing between whether patients call and cancel more or attend more as a result of increasing the fee for non-attendance, it is possible to design policies for non-attendance more efficiently. Lastly, this paper provides evidence for that monetary incentives are effective to reduce non-attendance rates. Reminders and information have been proven to be effective in previous research. However, health care providers' incentives to reduce non-attendance through reimbursement schemes have never been studied as an alternative method to reduce non-attendance. Its effectiveness is a question for further research. In summary, new ideas on further research have been raised in the wake of empirical evidence presented in this paper. This paper has contributed to research on monetary sanctions in general and non-attendance fees in particular by providing evidence for the effectiveness of increasing non-attendance fees in Swedish health care.

References

1177 Vårdguiden. (2016). *1177 Vårdguiden - sjukdomar, undersökningar, hitta vård, e-tjänster*. [online] Available at: http://1177.se [Accessed 23 May 2016].

Agarwal, S., Chomsisengphet, S., Mahoney, N. and Stroebel, J. (2014). Regulating Consumer Financial Products: Evidence from Credit Cards. *The Quarterly Journal of Economics*, 130(1), pp.111-164.

Angrist, J. and Pischke, J. (2009). *Mostly harmless econometrics*. Princeton: Princeton University Press.

Aron-Dine, A., Einav, L. and Finkelstein, A. (2013). The RAND Health Insurance Experiment, Three Decades Later. *Journal of Economic Perspectives*, 27(1), pp.197-222.

Bech, M. (2005). The economics of non-attendance and the expected effect of charging a fine on non-attendees. *Health Policy*, 74(2), pp.181-191.

Becker, G. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2), pp.169-217.

Bénabou, R. and Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5), pp.1652-1678.

Bhattacharya, J., Tu, P. and Hyde, T. (2014). *Health Economics*. Houndmills: Palgrave Macmillan.

Blackwell, M. (2013). *PSC: Differences-in-differeces estimators* [PDF]. Available at: http://www.mattblackwell.org/files/teaching/s11-diff-in-diff.pdf [Accessed 10 May 2016].

Bloomberg.com. (2016). *SEK to EUR Exchange Rate*. [online] Available at: http://www.bloomberg.com/quote/SEKEUR:CUR [Accessed 22 May 2016].

Currie, J. and Walker, R. (2011). Traffic Congestion and Infant Health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1), pp.65-90.

Deyo, R. and Inui, T. (1980). Dropouts and Broken Appointments. *Medical Care*, 18(11), pp.1146-1157.

Fehr, E. and Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422(6928), pp.137-140.

Frey, B. (2012). Crowding effects on intrinsic motivation. *Renewal*, 20.2-3, pp. 91-98.

Frey, B. and Jegen, R. (2001). Motivation Crowding Theory. *Journal of Economic Surveys*, 15(5), pp.589-611.

Frey, B. and Goette, L. (1999). *Does pay motivate volunteers?* (Vol. 7). University of Zurich, Institute for Empirical Research in Economics.

Frey, B. and Oberholzer-Gee, F. (1997). The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding- Out. *The American Economic Review*, (4). pp. 746.

Gneezy, U. and Rustichini, A. (2000A). A Fine is a Price. *The Journal of Legal Studies*, 29(1), pp.1-17.

Gneezy, U. and Rustichini, A. (2000B). Pay Enough or Don't Pay at All. *Quarterly Journal of Economics*, 115(3), pp.791-810.

Gneezy, U., Meier, S. and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25(4), pp.191-210.

Gurol-Urganci, I., de Jongh, T., Vodopivec-Jamsek, V., Atun, R. and Car, J. (2013). Mobile phone messaging reminders for attendance at health care appointments. *Cochrane Database Syst Rev*, 12.

Haselhuhn, M., Pope, D., Schweitzer, M. and Fishman, P. (2012). The Impact of Personal Experience on Behavior: Evidence from Video-Rental Fines. *Management Science*, 58(1), pp.52-61.

Hälso- och sjukvårdslag 1982:763. 1 §

Kolada.se. (2016A). *Kolada – Kvalitet och resurser i specialiserad somatisk vård*. [online] Available at: http://www.kolada.se/?_p=jamforelse&unit_id=27502 [Accessed 23 May 2016].

Kolada.se. (2016B). *Kolada – Kvalitet och resurseri primärvård*. [online] Available at: http://www.kolada.se/?_p=jamforelse&unit_id=27502 [Accessed 23 May 2016].

Kruse, A. (2013). *Welfare economics*. Lund: Studentlitteratur.

Lagen om läkevårdsersättning 1993:1651. 23 § 25 §

Lesaca, T. (1995). Assessing the influence of a no-show fee on patient compliance at a CMHC. *Administration and Policy in Mental Health*, 22(6), pp.629-631.

Mäntyjärvi, M. (2009). No-show patients in an ophthalmological out-patient department. *Acta Ophthalmologica*, 72(3), pp.284-289.

Nilsson, J. (2008). Skolkande patienter kostar miljarder. *Dagens Medicin*, pp.6-7.

OECD. (2014). *Health Statistics 2014 How does Sweden compare?*. [online] Available at: www.oecd.org/.../Briefing-Note-SWEDEN-2014.pdf [Accessed 23 May 2016].

Regeringen. (2015). *Kommittedirektiv.* [Online] Available at: http://www.regeringen.se/contentassets/0b3de375299446d7830d401739f9f094/offentlighetsprincipen-i-offentligt-finansierad-privat-utford-vard-och-omsorg-samt-utbildning-inom-vissa-skol--och-utbildningsformer-201592.pdf [Accessed 24 May 2016]

Regionfullmäktige Skåne. (2011). *Protokoll från sammanträde 2011-11-28-29.* [PDF].

Region Skåne. (2016). *Årsredovisning 2015.* [PDF].

Reti, S. (2003). Improving outpatient department efficiency: a randomized controlled trial comparing hospital and general-practice telephone reminders.*The New Zealand Medical Journal,* pp. 116(1175).

Rydmarker, N. and Deppert, U. (2015). *Uteblivna besök-förstudie*. [online] Available at: https://www.skane.se/siteassets/organisation_politik/revision/pm15/forstudie-uteblivna-besok.pdf [Accessed 23 May 2016].

SCB. (2016). *Medelfolkmängd.* [Online] available at: http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101D/MedelfolkFodelsear/table/tableViewLayout1/?rxid=d26b29dc-a2b5-4b6c-beee-9967f5de093a [Accessed 24 May 2016]

Skinner, J., Staiger, D. and Fisher, E. (2006). Is Technological Change In Medicine Always Worth It? The Case Of Acute Myocardial Infarction. *Health Affairs,* 25(2), pp.w34-w47.

SKL. (2015A). *Så styrs landsting och regioner*. [online] Available at: http://skl.se/demokratiledningstyrning/politiskstyrningfortroendevalda/kommunaltsjalvstyresastyrskommunenochlandstinget/sastyrslandstingen.1790.html [Accessed 23 May 2016].

SKL. (2015B). *Patientavgifter i hälso- och sjukvården 2016*. [online] Available at: http://skl.se/halsasjukvard/patientinflytande/patientavgifter.246.html [Accessed 23 May 2016].

Socialstyrelsen (2012). *Klartecken till sms-påminnelser i vården*. [online] Available at: http://www.socialstyrelsen.se/nyheter/2011maj/klarteckentillsms-paminnelserivarden [Accessed 10 May 2016].

Söderström, L. (2008). *The economics of social protection: the welfare state and its alternative*. Cheltenham: Edward Elgar.

Sunstein, C.R. (2003). *Why societies need dissent*. Cambridge, MA: Harvard University Press

Zweifel, P., & Manning, W. G. (2000). Moral hazard and consumer incentives in health care. *Handbook of health economics*, *1*, pp. 409-459.

Appendix

Table 12 Patient fees Skåne & Stockholm 2008-2015

	Primary Care		Secondary Care	
Year	Stockholm	Skåne	Stockholm	Skåne
2008	140 SEK	150/200 SEK	260 SEK	300 SEK
2009	140 SEK	120/200 SEK	300 SEK	300 SEK
2010	150 SEK	120/200 SEK	320 SEK	300 SEK
2011	150 SEK	120/200 SEK	320 SEK	300 SEK
2012	200 SEK	160/220 SEK	350 SEK	300 SEK
2013	200 SEK	160/200 SEK	350 SEK	300 SEK
2014	200 SEK	160/200 SEK	350 SEK	300 SEK
2015	200 SEK	160/200 SEK	350 SEK	300 SEK

Table 12 presents the patient fees in primary and secondary care for Stockholm and Skåne respectively. Skåne charge different fees depending on whether the patient is enrolled in the primary care unit. The lower fee is for enrolled patients. The fee in secondary care does not include referrals since the non-attendance fee both after than before the reform was the same whether there was a referral appointment or not.

Table 13 Regression results with non-linear age & different gender

	(1)	(2)	(3)	(4)
VARIABLES			Women	Men
Skåne* Post Reform	-0.701***	-0.701***	-0.663***	-0.799***
	(0.014)	(0.014)	(0.018)	(0.024)
Skåne	0.535***	0.533***	0.564***	0.463***
	(0.011)	(0.011)	(0.013)	(0.017)
Women	-0.450***	-0.448***		
	(0.008)	(0.008)		
Age	-0.181***	-0.201***	-0.058***	-0.077***
	(0.001)	(0.005)	(0.000)	(0.000)
Age^2	0.001***	0.001***		
	(0.000)	(0.000)		
Age^3		-0.000***		
		(0.000)		
Year fixed effects	Yes	Yes	Yes	Yes
Observations	11,998,587	11,998,587	7,124,961	4,873,626
R-squared	0.013	0.013	0.011	0.013

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14 Regression result with standard errors clustered on Individual ID

VARIABLES	(1)	(2)	(3)
Skåne* Post Reform	-0.752***	-0.711***	-0.705***
	(0.017)	(0.017)	(0.017)
Skåne	0.585***	0.518***	0.529***
	(0.013)	(0.013)	(0.013)
Woman		-0.383***	-0.429***
		(0.010)	(0.010)
Age		-0.065***	-0.060***
		(0.000)	(0.001)
Age groups			
25-34			-1.082***
			(0.035)
45-64			-1.461***
			(0.042)
65-84			-1.301***
			(0.051)
85+			-0.568***
			(0.061)
Observations	12,000,000	11,998,587	11,998,587
R-squared	0.000	0.012	0.013

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1