Profiling, Auditing and Public Policy: Applications in Labor and Health Economics

Vriend, S.

CHAPTER 3

Audit Rates and Compliance: A Field Experiment in Care Provision*

3.1 Introduction

The design of audit policies, using instruments such as the audit rate and sanctions, is concerned with a trade-off between the size of the budget for auditing and incentives for agents. This trade-off is relevant in many settings, including taxation, social insurance, pollution and health care. This chapter studies the effects of audit policies on behavior of care providers. Since the operational budget for auditing is limited and care expenditures are rapidly increasing, the design of audit policies is essential. In our field experiment, we consider both unconditional and conditional changes in audit rates. In the conditional case, audit rates depend on the provider's recent performance.

The effectiveness of audit policies has not been studied yet in health-care markets. These markets are often characterized by a mixture of public funding and private provision. There is not only the usual trade-off between strictness of audits and compliance, but also between efficient spending of resources and providing care when needed. We focus on types of care that should be applied immediately and have a median duration of three months. The need for quick provision has made the Dutch gatekeeper, who manages the long-term care budget, to introduce ex-post auditing.

*This chapter is based on Lindeboom et al. (2015a).

This allows care provision to start before the result of an audit is known. However, no direct sanction is possible in case of noncompliance. The audit regime thus mainly relies on trusting care providers, but uses auditing to provide feedback.[1] Receiving feedback is often argued to be important (e.g., Bandiera et al., 2009; Jamtvedt et al., 2006).

The gatekeeper traditionally applied a higher audit rate to so-called high-risk applications (applications associated with higher costs) and found that these have higher approval rates than low-risk applications. This was used to justify ex-post auditing without direct sanctions. However, care providers are involved in a long-run relationship with the gatekeeper giving them access to funding for different types of long-term care. Poorly performing care providers can lose their contract with the gatekeeper and thereby their opportunity to apply for public funding. Therefore, the long-run costs of noncompliance can be substantial.

Our study relates to theoretical work on random auditing (Allingham and Sandmo, 1972) and conditional audit rules (Landsberger and Meilijson, 1982; Greenberg, 1984; Harrington, 1988; Friesen, 2003). This literature has focused on tax compliance and environmental regulation.[2] Empirical evidence on the performance of audit policies primarily comes from laboratory experiments.[3] The use of field data has proven to be more difficult because of the typical hidden nature of noncompliance and issues of endogeneity of the audit rate (e.g. subsection 6.3 in Andreoni et al. (1998)).

Recently, there have been some studies on the effects of auditing on compliance behavior using field experiments. Iyer et al. (2010) consider construction firms and find a significant effect of increasing the awareness of detection risk on the reported tax base. Slemrod et al. (2001) find a significant positive effect of informing individuals

[1]The Dutch long-term care system is much more extensive than in most other countries. Most countries have a decentralized system where at the regional or municipality level needs assessment takes place, often implying a 100% audit rate. Like the Netherlands, Belgium also performs random checks ex-post to evaluate needs assessments. In the US sometimes ex-post auditing (”post-claims underwriting”) by insurance companies happens, but there sanctions will be imposed in the case of noncompliance.

[2]Compliance behavior of firms has been studied in particular by Crocker and Slemrod (2005), Bayer and Cowell (2009) and Hoopes et al. (2012). For extensive reviews of the taxation literature see Andreoni et al. (1998), Alm and McKee (1998), Slemrod and Yitzhaki (2002), Slemrod (2007), Kirchler et al. (2007) and Alm (2012). Empirical studies focusing on environmental regulation are e.g., Gray and Deily (1996), Laplante and Rilstone (1996) and Helland (1998).

[3]Compliance has been found to increase (slightly and non-linearly) with audit rates and modestly with sanction rates (Alm and McKee, 1998). Clark et al. (2004) compare random auditing, a conditional audit rule, and the optimal audit rule proposed by Friesen (2003). They find the latter two to be associated with fewer audits, whereas the compliance rate is maximized under random auditing. Cason and Gangadharan (2006) analyze the conditional audit rule suggested by Harrington (1988) and find participants to behave broadly in line with theoretical predictions.

that they will be audited closely on tax payments of low and middle-income taxpayers. Finally, Kleven et al. (2011) randomly sent ”threat-of-audit” letters to individuals in Denmark. They find that prior audits and ”threat-of-audit” letters have a significant positive effect on self-reported income, as opposed to third-party reported income. However, all these studies generate variations in beliefs rather than true variations in audit rates.

The aforementioned studies are concerned with ‘carrot-and-stick’ type of policies including sanctions for noncompliance. External interventions, like sanctions, may crowd out intrinsic motivation (e.g., Bénabou and Tirole, 2003). Some studies looked at the effect of warnings as an alternative (Nyborg and Telle, 2004; Eckert, 2004). Another alternative to deterrence is the incorporation of trust into enforcement policy.[4] Mendoza and Wielhouwer (2015) build a theoretical model in which a trusted agent faces a lower audit rate and remains being trusted until found not complying. They find feasibility of trust-based regulation to hinge on agents’ discount rates being sufficiently low and the existence of some costs of screening to the agent.

Our study contributes to the literature in three ways. First, while there are quite some laboratory experiments studying auditing policies, the evidence from field experiments is scarce. Exceptions are Slemrod et al. (2001) and Kleven et al. (2011) who study the effects of announcing certain audits. In contrast, our field experiment introduces true variation in audit rates, including also conditional variation. Second, audits in the market for immediate care are much more frequent than in tax auditing, due to both higher application rates and higher audit rates. Conditional audit rate updates occur, therefore, more often and are based on more information. Third, direct monetary sanctions are absent in our setting, instead we rely on providing feedback. This contrasts other studies, in particular laboratory experiments, which study audit regimes with direct financial consequences of noncompliance.

We have detailed administrative data on all immediate care applications filed by the providers participating in our field experiment. This includes the type, amount and duration of assessed care services and some basic patient characteristics. Furthermore, we observe whether an application was selected for audit and, if so, the audit decision and motivation. We can thus study effects on both application quantity and quality (i.e., compliance rate). Our results do not show significant effects of an exogenous change in the unconditional audit rate on the number of

[4] Rousseau et al. (1998) give a cross-discipline overview of how the concept of trust is defined. A huge (experimental) literature on trust and reciprocal behavior in (repeated) interactions has developed, starting with the work on trust games by Berg et al. (1995) (see Fehr and Gächter (2000) for a review and Fehr and List (2004) for more recent evidence).

applications nor on the audit approval rate. We also do not find significant effects on these two outcome variables of switching to a conditional audit regime. Even though we do observe divergence in audit rates among care providers in the conditional audit regime, we do not find much evidence for behavioral responses of care providers. The resulting audit rate changes can be caused by sorting based on pre-experiment average approval rates of care providers. Finally, we do not find evidence for the presence of heterogeneous treatment effects.

The remainder of this chapter is organized as follows. The next section provides details on the system covering long-term care in the Netherlands. Section 3.3 describes the experiment and discusses hypotheses on the effects of the various treatments (audit regime changes). Section 3.4 describes the data. The empirical results are presented in section 3.5. Finally, section 3.6 concludes.

3.2 Institutional background

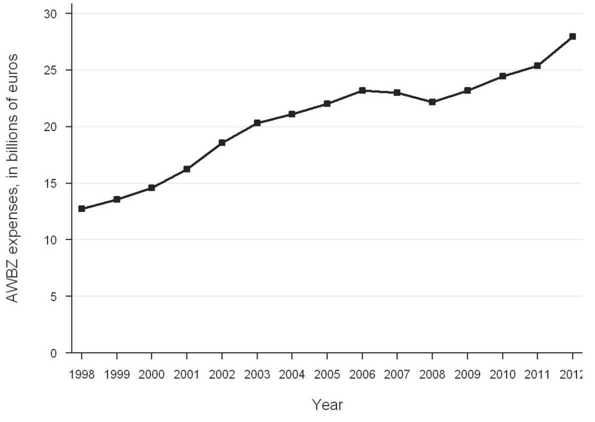
Each inhabitant of the Netherlands is publicly insured for long-term care by means of the Exceptional Medical Expenses Act (AWBZ).[5] The immediate (home and nursing) care we consider in this chapter is part of this long-term care arrangement. On July 1, 2013, 4.8% of the Dutch population qualified for receiving any type of long-term care. Of these eligible individuals, 24% were older than 85 years. Long-term care expenses are financed by means of general taxation and co-payments. In most developed countries, long-term care expenditures are rapidly increasing. Figure 3.1 shows that the Netherlands is no exception. Expenditures from the long-term care budget amounted to 28 billion euros (i.e., 4.7% of GDP) in 2012, a 51% increase compared to the 2002 level. Correcting for price level changes, the increase was about 60% between 1998 and 2012, which can only partly be explained by demographic trends.

Annually, around one million applications for long-term care are submitted. These are categorized in several types. We focus on one type, the so-called 'standard assessment guideline' application, which accounts for approximately 34% of all applications. We refer to this type of care as immediate care. Immediate care concerns relatively cheap care that needs to be provided quickly and typically lasts for a relatively short period of time after which a follow-up application for prolonging care services can be filed if care services are still needed.[6] In the remainder of this

[5] This section draws on Mot (2010), Nederlandse Zorgautoriteit (2012a,b) and information from the Center for Care Assessment (CIZ).

[6] The median duration of immediate care assessments in our data is 92 days. Other, more expensive types of care, are audited ex-ante. The median duration for those types of care is 5478 days.

Figure 3.1: AWBZ expenses in the Netherlands, 1998 - 2012



Note: Data are from CBS StatLine; figures for 2011 and 2012 are preliminary.

chapter, we refer to this type of care as immediate care. This contains, for example, wound care and/or personal care after hospitalization, but also help in putting on stockings and administering injections. The immediate care often substitutes for more expensive care in hospitals or institutionalized care such as nursing homes.

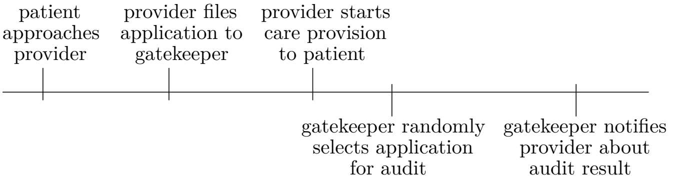
Figure 3.2 shows the process of making an application for receiving immediate care. When a patient wants to receive immediate care services, (s)he contacts his/her preferred care provider. The patient is allowed to switch provider during care provision, which gives care providers an incentive to provide good-quality care and to provide it quickly. The provider then files a request to the gatekeeper for funding for care provision.[7] Since immediate care is quickly needed, care provision can start immediately after filing the application. However, ex-post, a random sample of the applications is checked by the gatekeeper.

Applications selected for audit are randomly divided among specialized assessors employed by the gatekeeper. During an audit, the assessor makes an assessment of the needs, stating the type and amount of care services.[8] An audit always takes place within two weeks after making the application, but usually within one week. The audit can result in an approval or disapproval. Several reasons for disapproval exist.

[7]The patient can also file an application directly to the gatekeeper, but usually (around 85% of cases) the application is filed by the care provider.

[8]Types of care services distinguished are: personal care (e.g., help with showering), nursing (e.g., wound care), assistance (help in organizing practical matters in daily life), treatment (e.g., rehabilitation) and short-term stay or stay for an extended period of time in an institution (e.g., nursing home). For each service the amount of care (usually in hours per week) the individual qualifies for is determined.

Figure 3.2: Timeline for the application process



For instance, the type of limitations making the patient eligible for care services may differ, different services may be needed or the necessary amount of care may differ. After the audit, the assessor contacts the care provider to give feedback and in case of a disapproval explains how applications should be improved in the future.

A disapproval decision resulting from the ex-post audit does not have direct implications, the care provider does not have to repay anything. The goal of this so-called ex-post auditing is to provide feedback to the care providers, which should increase the quality of application. Furthermore, it allows the gatekeeper to obtain insight in the application behavior of care providers. For their funding care providers are fully dependent on the gatekeeper and are, therefore, involved in a long-term relationship also covering other types of care. So while direct costs of noncompliance are absent, long-run indirect costs can be high.

Already since the gatekeeper became responsible for managing immediate care, they distinguish between low-risk and high-risk applications. They defined high-risk applications as applications for relatively expensive types of immediate care, for which they also expected higher disapproval rates. The definition of high-risk applications has not changed since then and also does not vary between care providers. High-risk applications are often related to the transfer from a hospital to care at home. The gatekeeper usually audits 6% of the low-risk applications and 16% of the high-risk applications.

In 2011, there were almost 1,400 extramural care providers and 800 intramural care providers active (a provider may provide both types of care). Intramural care providers are required to be non-profit institutions, while extramural care providers can also be for-profit organizations. Since immediate care replaces more expensive hospital care at the end of a hospitalization, a substantial share of the immediate care applications are filed by hospitals in the process of discharging a patient from the hospital.

3.3 The experiment

The goal of the field experiment is to investigate the consequences of being exposed to different audit regimes. In the next subsection we discuss how care providers are randomized over the control group and the three treatment groups. In subsection 3.3.2 we present the hypotheses concerning the effects of being assigned to the different treatment groups.

Our field experiment started September 17, 2012 and ran until April 7, 2013. In the *control group*, audit rates remained at the pre-experiment levels of 6% for low-risk applications and 16% for high-risk applications. Next, we defined three treatment groups. The gatekeeper required that the total number of audits should remain relatively constant such that all audits could be done by the existing (experienced) assessors. Furthermore, the gatekeeper imposed that audit rates should remain within the range that is considered feasible. In practice this meant that audit rates should be at least 2% and at most 26%.

Given the restrictions of the gatekeeper we defined two treatment groups with unconditional audit rates and one treatment group with conditional audit rates. The first unconditional treatment group is the *low-rate group*, for which audit rates are set to 2% for low-risk applications and 10% for high-risk applications. The *high-rate group* is the other unconditional treatment group, and providers in this group face audit rates of 10% and 26% for low-risk and high-risk applications, respectively.

For care providers in the *conditional audit-rate group*, the audit rate depends on previous performance. Care providers started at the pre-experiment audit rate levels of 6% and 16%. For each approval the audit rate was reduced by 0.2 percentage point. On the other hand, if an application was disapproved, the audit rate was increased by three percentage points.[9] These adjustments imply a constant audit rate when 93.75% of audited applications are approved, which is about the target of the gatekeeper. The audit rates are subject to a minimum of 2% and a maximum of 26%. We updated audit probabilities in this group once every two weeks during the experiment.[10] For updates implemented in week t, data for weeks $t-2$ and $t-1$ were used. With each series of adjustments, care providers were informed about the outcomes of conducted audits and the resulting audit rate.

[9] Our conditional audit regime is a variant of the rule proposed by Greenberg (1984) and Harrington (1988). In their setting agents switch between a small number of groups with varying audit intensity. We implement a continuous audit rate adjustment scheme.

[10] In the start-up phase of the experiment the procedures surrounding these updates took some more time. The first (second) adjustments were in place four (three) weeks after the start of the experiment (first update).

3.3.1 Implementation

We selected all 226 care providers that filed at least four immediate care applications in April 2012. These care providers account for 78% of all, approximately 22,000, immediate care applications filed in this month. We randomly assigned the care providers to one of the four groups.[11] The managing boards of the participating care providers were informed, by letter, early September 2012 about the purpose and set-up of the experiment and the assignment of the audit policy applicable to them. Furthermore, case managers of the care providers were informed via e-mail. Finally, a short notice of the experiment was posted in the online information bulletin at the end of October.[12] More details on the information given to participants in the experiment are provided in Appendix 4.A.

To show that the groups are balanced, Table 3.1 provides descriptive statistics on pre-experiment outcomes and some other characteristics. The table shows no significant differences across groups in the number of applications, the number of audits and the approval rate. On average, care providers filed 18 applications per week in the pre-experiment period of which about 80% are low-risk applications. Of the audited applications, 83% are approved. Around 38% of care providers are hospitals, which is balanced across groups.

3.3.2 Hypothesized effects

Predictions on behavioral responses to audit regime changes are frequently derived from models building on Allingham and Sandmo (1972). In these models incentives for agents being audited result from the presence of (financial) sticks. Our setting lacks direct financial consequences of disapprovals of audits. However, care providers may still dislike disapprovals, because the gatekeeper will contact them to discuss the application and the reason for disapproval. This takes time of employees of the care providers. A straightforward prediction is then that an exogenous increase in the audit rate should increase compliance. This yields predictions for the behavior of care providers in both unconditional treatment groups. In the low-rate treatment group disapproval rates should be higher than in the control group and these should be lower in the high-rate treatment group.

[11] The original proposal for the field experiment including power analysis is available at http://personal.vu.nl/b.vander.klaauw/OpzetCIZOnderzoek.pdf [in Dutch].

[12] At the start of the experiment, care providers were only provided a rough indication about the minimum duration of the field experiment. Eventually, the experiment lasted somewhat longer and care providers were never provided additional information about the end date of the experiment. This should minimize anticipation behavior on the end of the experiment period.

Table 3.1: Descriptive statistics on outcomes in the pre-experiment period (Jan 2012 - Sept 2012) and care provider characteristics, by treatment group.

	control	**low**	**high**	**conditional**	**p-value**
Outcomes					
applications (per week)	18.61	18.21	18.73	17.13	0.969
	(2.57)	(2.43)	(2.64)	(2.11)	
audits (per week)	1.39	1.39	1.44	1.33	0.998
	(0.19)	(0.19)	(0.19)	(0.17)	
approval rate	0.84	0.83	0.80	0.88	0.102
	(0.03)	(0.02)	(0.03)	(0.02)	
Care provider characteristics					
hospital	0.32	0.44	0.38	0.39	0.669
	(0.06)	(0.07)	(0.06)	(0.07)	
frac. low risk	0.82	0.84	0.80	0.79	0.528
	(0.03)	(0.03)	(0.03)	(0.03)	
observations	1001	987	1026	1017	
# two-week periods	18	18	18	18	
# care providers	56	55	58	57	

Notes: Standard errors of the means are in brackets. Reported p-values are for Kruskal-Wallis rank tests for equality of populations.

If disapprovals are related to invalid applications and higher audit rates reduce disapprovals, then an exogenous increase in the audit rates should also reduce the number of applications. Then there should be more applications in the low-rate treatment group and fewer applications in the high-rate group than in the control group. So based on the traditional model we can formulate two hypotheses concerning the behavior of care providers in the low-rate and high-rate treatment group.

Hypothesis 1: *If care providers consider audits and in particular disapprovals as costly, then in the low-rate treatment group the disapproval rate should be higher than in the control group and in the high-rate treatment group lower than in the control group.*

Hypothesis 2: *If care providers consider audits and in particular disapprovals as costly and disapprovals are related to invalid applications, then in the low-rate treatment group more applications should be made than in the control group and in* *the high-rate treatment group fewer than in the control group.*

For care providers in the conditional audit rate treatment group there is a direct effect of the change in audit rates, but also an incentive effect. The direct effect implies that for poorly performing care providers audit rates will increase, while these decrease for the well performing care providers. Most disapprovals concern applications made by poorly performing care providers. These care providers will experience higher audits rates having the same effects as hypothesized for the high-rate treatment group. Additionally, there is an incentive effect due to the fact that each disapproval will increase audit rates. If again care providers dislike being audited, then they prefer lower audit rates. Both the direct effect and the incentive effect predict that the disapproval rate of a care provider in the conditional audit-rate treatment group should be lower. And again if disapprovals are associated with invalid applications, also the number of applications should decline.

Hypothesis 3: *If care providers consider disapprovals as costly and/or dislike being audited, then in the conditional audit-rate treatment group the disapproval rate should be lower than in the control group.*

Hypothesis 4: *If care providers consider disapprovals as costly and/or dislike being audited and disapprovals are related to invalid applications, then in the conditional audit-rate treatment group fewer applications should be made than in the control group.*

However, recall that in our setting direct financial incentives are absent, i.e., disapprovals do not have immediate consequences. The predictions based on the Allingham-Sandmo model, therefore, depend on care providers disliking audits and disapprovals. There are other mechanisms through which audit regimes affect behavior. Audits provide feedback, which care providers can use to learn about the appropriate application behavior. A higher audit rate should then improve behavior if care providers are intrinsically motivated to perform well in audits, for example, because they care about their reputation with the gatekeeper. The presence of feedback and motivation generate the same predictions on behavioral responses to the audit regime as in the Allingham-Sandmo model (i.e., Hypotheses 1 and 2). Also for care providers in the conditional audit rate group predictions remain the same as described in Hypotheses 3 and 4. Intrinsically motivated care providers in this treatment group aim at obtaining a low audit rate, which they can achieve by improving their applications to reduce the disapproval rate.

Alternatively, auditing may affect how much care providers feel trusted by the gatekeeper. According to Mendoza and Wielhouwer (2015) trust-based enforcement policies, that do not include sticks, may work when there is a cost associated with audits and when care providers sufficiently value the future (e.g., for reputational reasons). Such costs exist when auditors contact care providers for more information, in particular in case of disapprovals. The audit rate signals the degree of trust. If trust-based enforcement works, unconditional high audit rates and thus less trust can have adverse effects on compliance. The conditional audit-rate regime includes trust dynamics, trust can be built and dissolved based on observed performance. To the extent that providers care about the extent of trust, this regime may provide additional incentives and lead to higher approval rates. A model based on trust thus generates predictions opposite to those described in Hypotheses 1 to 4.

The hypotheses describe behavioral responses to the audit regime to which a care provider is exposed. In our field experiment care providers are randomized in the different audit regimes. Therefore, we can test the validity of the hypotheses by comparing the application behavior (disapproval rate and number of applications) between the different treatment groups. The audit regime to which a care provider is exposed will be captured by a simple dummy variable. It is important to stress that we are not focusing on the effects of the size of the audit rates in the conditional regime. These are endogenous and, therefore, much more difficult to analyze.

3.4 Data

We use administrative data on all 269,142 applications for immediate care filed by the 226 participating care providers between January 1, 2012 and April 7, 2013. For each application we observe the care provider filing the application, the amount and type of care services, the application date and whether or not the application has been audited. If there was an audit, we observe the date of the audit, the assessor who performed the audit, the result of the audit (approved or disapproved) and the reasons for this audit decision. In total, 8.3% of all applications (22,279 applications) have been subject to an audit. The overall approval rate was 0.88.

For most analyses we aggregate the data at the period and care provider level, with a period length of two weeks.[13,14] We normalize the number of applications to

[13]The approval rate is constructed as the aggregated number of approved audits divided by the number of audits, so that a missing value results whenever no audits have been undertaken for a particular period-provider observation.

[14]Table 3.A.1 in Appendix 3.A shows that the results are robust to aggregating into one week or four week periods.

weekly averages for ease of interpretation. The panel data set is unbalanced, because some care providers became inactive.[15]

Table 3.2 shows descriptive statistics on outcome variables during the experiment, by treatment group. On average, 17 to 20 applications are filed by each care provider on a weekly basis. Most of these are low-risk applications. A Kruskal-Wallis test shows that there are no significant differences in the number of applications across groups. The significant difference across groups in the number of audits is a direct result of the imposed audit rate variation. This is illustrated in Figure 3.3 showing the average number of audits over time. We see a clear jump at the start of the experiment for the low and high-rate groups, and a gradual increase in the conditional audit-rate group. We return to the dynamics in the conditional audit rate group in subsection 3.5.1. Finally, Table 3.2 shows the overall approval rate to be similar across groups (around 0.84). The approval rate is larger for high-risk applications. This was already the case before the start of the experiment. Although before the experiment the difference across treatment groups in the approval rate of high-risk applications was insignificant, we observe a significant difference during the experiment.

We are interested in the effects of the audit regime on application behavior of care providers. We consider effects on both quantity (number of applications) and quality (audit approval rate). Figure 3.4 shows the trend in the average number of applications per week, by treatment group. The pattern in the number of applications appears similar across groups, both before and during the experiment, with an increasing trend emerging from around August 2012 onwards.[16] Finally, Figure 3.5 shows the trend in the average approval rate of audits. The figure illustrates that approval rates fluctuate considerably between 0.76 and 0.96. However, no systematic patterns emerge across groups.

[15]Inactivity arises, for example, due to closing down or merging. Alternative approaches could be to set the number of applications to zero in case of inactivity, or to exclude all care providers from the data that became inactive during the observation period. Table 3.A.2 in Appendix 3.A shows that our results are robust against these choices.

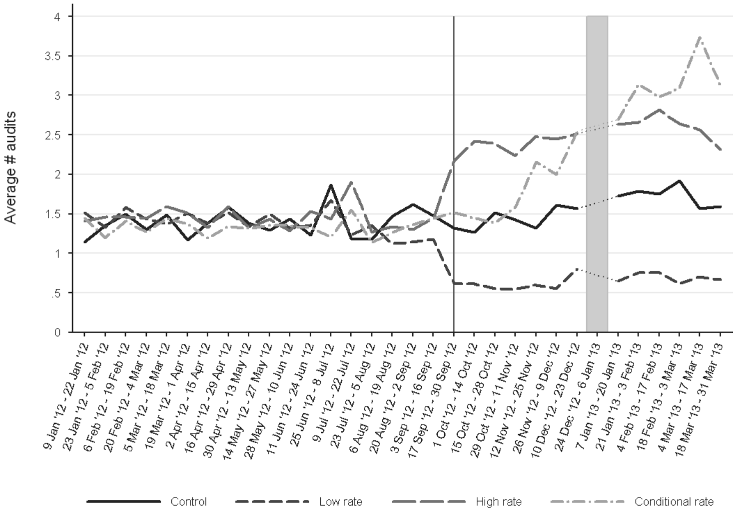
[16]The number of audits exhibits a small spike during July 2012, so before the experiment started. Since the number of applications did not increase and the composition of applications did not change, there is no clear explanation. So it is either randomness or the effect of assessors delaying or speeding up audits due to their summer break.

Table 3.2: Descriptive statistics on outcomes during the experiment.

	control	**low**	**high**	**conditional**	**p-value**
applications (per week)	20.34	17.17	19.04	18.74	0.931
	(3.01)	(2.24)	(2.41)	(2.22)	
low risk	16.81	13.30	15.25	13.89	0.928
	(2.99)	(2.00)	(2.35)	(2.09)	
high risk	3.48	3.85	3.76	4.80	0.661
	(0.84)	(0.88)	(0.79)	(1.06)	
audits (per week)	1.52	0.64	2.43	2.34	0.000
	(0.20)	(0.10)	(0.29)	(0.31)	
approval rate	0.83	0.79	0.84	0.87	0.876
	(0.02)	(0.04)	(0.02)	(0.01)	
approval rate low risk	0.83	0.79	0.83	0.87	0.657
	(0.02)	(0.04)	(0.02)	(0.02)	
approval rate high risk	0.96	0.89	0.87	0.89	0.013
	(0.01)	(0.03)	(0.04)	(0.04)	
observations	754	756	774	752	
# two-week periods	14	14	14	14	
# care providers	54	54	56	54	

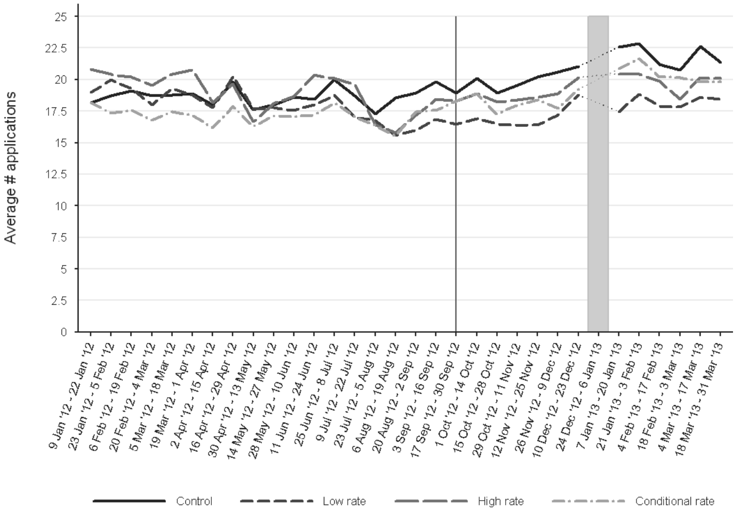
Notes: Standard error of the mean in brackets. Reported p-values are for Kruskal-Wallis rank tests for equality of populations. Recall that fraction of audits approved is missing in case of either zero applications filed or zero audits performed. Fraction of audits approved is defined in 4,857 out of 7,067 observations. For low risk and high risk fraction approved is defined in 3,949 and 2,321 cases, respectively.

Figure 3.3: Average number of audits.



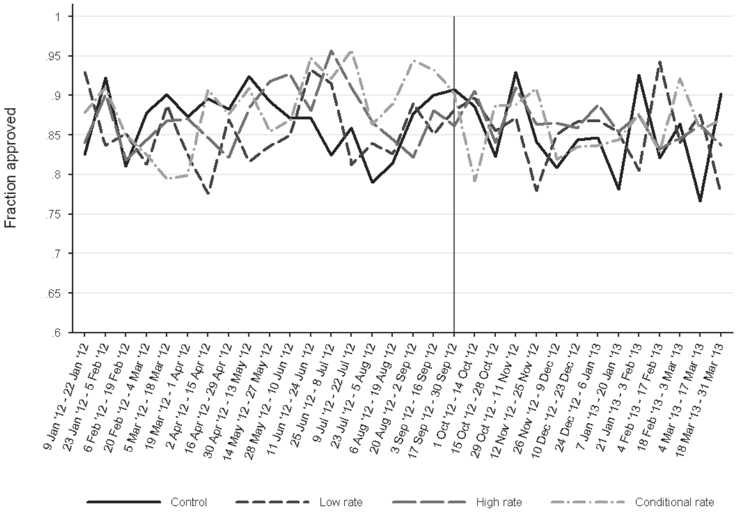
Notes: The vertical line indicates the start of the experiment. The grey band shows an interpolated value as around Christmas much fewer applications have been filed.

Figure 3.4: Average number of applications.



Notes: The vertical line indicates the start of the experiment. The grey band shows an interpolated value as around Christmas much fewer applications have been filed.

Figure 3.5: Average approval rate across care providers.



3.5 Results

We estimate a linear panel data model to obtain the causal effects of the audit regime.

$$Y_{i,t} = \alpha_i + \gamma_t + \sum_{\substack{g \in \{low, high, \\ conditional\}}} \delta^g T^g_{i,t} + \varepsilon_{i,t} \tag{3.1}$$

The outcome variables $Y_{i,t}$ are the number of applications and the approval rate of provider i in time period t and $T_{i,t}$ describes the audit regime to which the provider was exposed.[17] Recall that there are one control group and three treatment groups, for which we have dummy variables. When estimating effects on the number of applications, we normalize the outcome by dividing through the pre-treatment provider-specific average. Therefore, we leave out care provider fixed effects α_i from this specification.[18] The treatment effects δ^g's (multiplied by 100) represent percentage changes. For regressions having the approval rate as the outcome, the

[17] As outcome we also considered other application types, but we did not find any evidence for spillover effects due to changes in the audit regime. These results are available on request.

[18] Since care providers are randomly assigned to the treatment groups, the treatment variables are uncorrelated with the care provider specific effects. Therefore, also random effects estimation can be used. This gives very similar results with only modestly smaller standard errors.

causal effect estimates represent percentage points changes in the approval rate.[19] Finally, we include time fixed effects γ_t to account for common time trends.[20] We report standard errors clustered by care provider. Bertrand et al. (2004) discuss the case of serially correlated standard errors. We have followed their suggestions to collapse the time series data into one pre-experiment and one post-treatment period. These results are presented in Table 3.A.4 in Appendix 3.A, and are virtually the same as the results discussed below.

The baseline estimation results in the first column of Table 3.3 show a 3.6% and 5% decrease in the number of applications for the low audit rate group and the conditional audit rate group, respectively. For the high audit rate group we find a small increase of 2.5%. The signs for the low audit-rate and high audit-rate groups do not agree with our predictions in Hypothesis 2. For the conditional audit-rate group the sign confirms Hypothesis 4. But, none of these treatment effect estimates are significantly different from zero. Results for the effects on the approval rate are shown in the fifth column. We do not see any effect on the quality of applications. The estimated coefficients are almost zero for each treatment group and insignificant.[21,22] Therefore, the results do not confirm our Hypotheses 1 and 3.

The results contradict results from laboratory experiments, where higher audit rates are found to induce slightly more compliance (Alm and McKee, 1998). However, audit regimes studied in the laboratory differ from the audit regimes we study. In particular, in laboratory experiments there are often strong direct incentives associated with disapprovals, which is absent in our setting. The sign of the estimates for the unconditional audit-rate changes is consistent with what we would expect when less trust (i.e., more frequent auditing) crowds out intrinsic motivation. But again we do not find significant effects. The lack of effects of switching to a conditional audit regime is not in line with previous studies that found positive effects on performance when using a conditional instead of a random auditing rule (Clark et al., 2004).

[19] Approval rates are bounded between zero and one, which is not taken into account in our model specification. A Tobit model, which takes account of this limited support gives very similar results.

[20] We investigate whether results are robust to different time trend specifications, such as a polynomial in time or quarter dummies. Results do not change considerably, as shown in Table 3.A.3 in Appendix 3.A.

[21] If there are learning effects, it may take some time for providers to change behavior. Therefore, we estimated separate effects for the first 14 weeks of the experiments and the final 14 weeks, but we did not find evidence for different effects. These results are available on request.

[22] Since the approval rate for low-risk and high-risk applications differs, we studied effects on the approval rates for both types of applications separately. The results are presented in Table 3.A.5 in Appendix 3.A and do not show differences between both types of applications.

Our baseline results show insignificant changes in the number of applications. However, there may be compositional differences. In particular, the audit regimes may affect the composition of high-risk and low-risk applications. We find a small but insignificant decrease in the fraction of low-risk applications. The changes in audit regimes do not seem to influence the composition in terms of low-risk and high-risk applications. The estimation results can be found in Table 3.A.6 in Appendix 3.A.

Next, we study the presence of heterogeneous effects across care providers. The key sources of heterogeneity are whether the care provider is a hospital (38.5% of care providers) or a non-hospital (e.g., nursing homes and home care institutions), and the size of the provider measured in terms of number of immediate care applications.[23] We estimate heterogeneous treatment effects by interacting the time trend and treatment indicators in equation (3.1) with a dummy variable for care provider type.[24]

Treatment effects on the number of applications for small compared to large care providers are shown in column (2) of Table 3.3. We see a slight decrease for all but small care providers in the high audit-rate group, but none of the treatment effects is significantly different from zero. Furthermore, F-tests indicate that effects for small and large care providers are similar for all treatment groups. Effects on the approval rates are positive for large care providers in all treatment groups, whereas effects are negative for small care providers. However, at a 5% significance level, we do not find significant heterogeneity in treatment effects nor is any of the treatment effects estimates significantly different from zero.

Finally, we also do not find evidence for the presence of heterogeneity in effects on the number of applications for hospitals and non-hospital. For the approval rate we see that hospitals in the conditional audit rate group tend to perform worse during the experiment, in comparison to the control group, whereas non-hospitals perform somewhat better on average. However, this difference in effects is only significant at a 10% level.

While we obviously need aggregated data when having the number of applications as the outcome, causal effects on the approval rate can also be estimated using individual application-level data. In such a model, where the outcome measure is an approval dummy variable, we can add additional control variables for type and

[23] Other measures of size strongly correlate with this and, therefore, give very similar results.

[24] The model specification we estimate is $Y_{i,t} = \alpha_i + \gamma_{h_i,t} + \sum_{g \in \{low, high, conditional\}} \delta^g_{h_i} T^g_{i,t} + \varepsilon_{i,t}$, where h_i indicates the care provider type. This is equivalent to estimating the baseline model separately for different types of care providers.

amount of care and basic patient characteristics. Estimation results are presented in Table 3.A.7 in Appendix 3.A. In each of the specifications we find small and insignificant effects of the treatments on the probability of approval during an audit. Compared to our baseline results, the magnitude of the effect in the conditional audit rate group is somewhat larger.

3.5.1 Sorting effects in the conditional audit rate group

This subsection describes the dynamics in the conditional audit-rate group caused by audit rate updates during the experiment. In total, twelve updates of the audit probability have been implemented. Table 3.4 provides descriptives on these adjustments. The first column shows a gradual increase in the average low-risk audit rate from 6% to 13.5%. On average, care providers are scoring below the target approval percentage of 93.75%. In each update there are changes in both directions and for 40 to 50% of care providers the audit rate remains constant. During the experiment, the fraction of care providers for which the high-risk audit rate reached the maximum of 26%, increased. After the final update, this was the case for almost two-fifth of the care providers, as illustrated in the penultimate column. For 73% of those, also the low-risk rate reached the maximum. As shown in the final column, only for a few care providers the audit rate hit the floor of 2% at some point.

Even though the audit rate increases on average during the experiment, this does not necessarily imply a decrease in compliance of care providers. Similar dynamics in the audit rate might result when providers, already before the experiment, consistently had compliance levels below the 93.75% target approval percentage. This would cause a gradual increase in the (average) audit rate. In this subsection we investigate whether the audit rate dynamics can be attributed to sorting of providers based on pre-experiment compliance levels.

Table 3.4 clearly shows the presence of dynamics in the conditional audit rate group. Our earlier results indicated the absence of behavioral responses to the change in audit regime. Therefore, these changes should be the consequence of heterogeneity among care providers determining audit rates. To study sorting of care providers we regress the audit rate at the end of the experiment on the average approval rate of each care provider in the pre-experiment period. The estimation results in Table 3.5 show a significant negative relationship between pre-experiment application quality and the final audit rate. That is, the lower the pre-experiment approval rate, the higher the final audit rate.

Table 3.3: Estimation results: baseline and heterogeneous effects.

	applications			**approval rate**		
	baseline (1)	*size* (2)	*hospital* (3)	*baseline* (4)	*size* (5)	*hospital* (6)
low	−0.036			−0.001		
	(0.057)			(0.022)		
low × small		−0.039			−0.085	
		(0.095)			(0.060)	
low × large		−0.051			0.026	
		(0.067)			(0.022)	
low × hospital			−0.046			−0.038
			(0.051)			(0.024)
low × non-hospital			−0.010			0.032
			(0.089)			(0.036)
high	0.025			0.000		
	(0.126)			(0.021)		
high × small		0.062			−0.021	
		(0.254)			(0.065)	
high × large		−0.027			0.005	
		(0.043)			(0.019)	
high × hospital			0.301			−0.018
			(0.290)			(0.025)
high × non-hospital			−0.146*			0.011
			(0.078)			(0.031)
conditional	−0.050			−0.000		
	(0.064)			(0.020)		
conditional × small		−0.115			−0.034	
		(0.123)			(0.053)	
conditional × large		−0.006			0.010	
		(0.039)			(0.018)	
conditional × hospital			−0.014			−0.036*
			(0.047)			(0.019)
conditional × non-hospital			−0.066			0.028
			(0.100)			(0.032)
large		0.028				
		(0.059)				
hospital			−0.055			
			(0.054)			
mean control	1.000			0.835		
mean control, small		1.000			0.756	
mean control, large		1.000			0.897	
mean control, hospital			1.000			0.893
mean control, non-hospital			1.000			0.807
F-test p-value, low		0.918	0.730		0.081	0.105
F-test p-value, high		0.731	0.138		0.702	0.460
F-test p-value, conditional		0.398	0.637		0.432	0.085
observations	7,067	7,067	7,067	4,857	4,857	4,857
# providers	226	226	226	224	224	224

Notes: Pre-experiment means of the outcome in the control group are reported as a reference, separately for different types of providers. Cluster-robust standard errors are in brackets, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. F-test p-values are for testing the null hypothesis of no heterogeneity in treatment effects.

Table 3.4: Descriptive statistics on the periodic adjustments in the conditional audit rate group.

Update round	Mean (st.dev.) audit rate				Direction of change			Thresholds	
	low risk		*high risk*		% ↓	% =	% ↑	*% at max*	*% at min*
0 (17 Sep '12)	6.0%	(0.0)	16.0%	(0.0)	-	-	-	-	-
1 (15 Oct '12)	5.8%	(1.1)	15.8%	(1.1)	36.8%	45.6%	17.5%	0.0%	0.0%
2 (5 Nov '12)	7.6%	(3.4)	17.5%	(3.1)	21.1%	24.6%	54.4%	1.8%	3.5%
3 (19 Nov '12)	8.2%	(4.3)	17.9%	(3.5)	28.1%	43.9%	28.1%	5.3%	1.8%
4 (3 Dec '12)	8.5%	(5.4)	18.0%	(4.3)	33.3%	43.9%	22.8%	12.3%	10.5%
5 (17 Dec '12)	9.0%	(5.8)	18.3%	(4.5)	28.1%	47.4%	24.6%	12.3%	7.0%
6 (31 Dec '12)	9.9%	(7.3)	18.6%	(5.2)	29.8%	36.8%	33.3%	21.1%	8.8%
7 (14 Jan '13)	10.5%	(7.5)	18.9%	(5.1)	17.5%	54.4%	28.1%	22.8%	7.0%
8 (28 Jan '13)	11.4%	(7.9)	19.4%	(5.2)	19.3%	50.9%	29.8%	28.1%	5.3%
9 (11 Feb '13)	12.1%	(8.1)	19.8%	(5.2)	15.8%	52.6%	31.6%	31.6%	5.3%
10 (25 Feb '13)	12.7%	(8.3)	20.1%	(5.4)	24.6%	43.9%	31.6%	36.8%	3.5%
11 (11 Mar '13)	12.9%	(8.8)	20.0%	(5.5)	21.1%	50.9%	28.1%	35.1%	5.3%
12 (25 Mar '13)	13.5%	(9.4)	20.1%	(5.5)	17.5%	54.4%	28.1%	38.6%	5.3%

Notes: Standard deviation in brackets. Averages and percentages are computed over all 57 care providers in the conditional audit rate group. In the first column the start date of the updated audit rates is shown.

Next, we use the control group to simulate audit rate updates under the conditional audit-rate regime. Behavioral responses are, of course, absent in the control group, so simulated adjustments only reflect sorting. We take the realized number of applications and the long-run average approval rate for each care provider in the control group. In the simulations we select applications for audit based on draws from a binomial distribution with success probability equal to the audit rate in that period. Each application selected for audit is approved with a probability equal to the long-run average approval rate of the care provider. Then, we use the resulting number of approved and disapproved audits to compute the updated audit rate according to the rules in the conditional audit rate regime. We updated audit rates in twelve periods, as we did in the experiment.

We simulate the complete adjustment process 10,000 times. For each care provider, we then compute the average final audit rate over all simulations. Table 3.6 compares some distributional statistics for the simulated and the actual audit rate distributions. Although there are some differences in the percentages of care providers arriving at the ends of the distribution, the mean and standard deviation are quite close. Histograms of the simulated and the actual audit rate distributions are shown in Figure 3.6, separately for the audit rates for low-risk applications and the audit rates for high-risk applications. Testing for the equality of the simulated and the actual audit rate distributions by means of a Kolmogorov-Smirnov test leads us to conclude that the null hypothesis of equal distributions cannot be rejected (exact

p-value is 0.215 for the audit rates of low-risk applications and 0.210 for the audit rates of high-risk applications). Pure sorting based on pre-experiment application quality, without behavioral responses to implemented performance incentives, yields the audit rate distribution as realized for the conditional audit rate group. This confirms again the lack of behavioral responses and that divergence of audit rates in the conditional audit-rate regime is solely driven by selection effects.

Table 3.5: Pre-experiment performance and final audit rates in the conditional audit rate group.

	audit rate
pre-experiment performance	-0.356^{**}
	(0.134)
constant	0.172^{***}
	(0.022)
R^2	0.074
observations	57

Notes: The outcome is the uncapped low risk audit rate. Only the constant term differs when using the high risk audit rate as the outcome variable. Robust standard errors are in brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

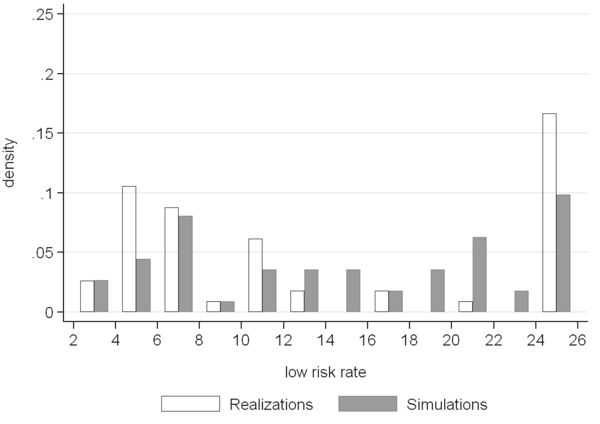
Table 3.6: Actual and simulated audit rate distribution.

	actual	**simulated**
mean low risk rate	13.5%	14.7%
st. dev. low risk rate	9.4	7.8
mean high risk rate	20.1%	20.8%
st. dev. high risk rate	5.5	5.2
% at maximum	38.6%	46.0%
% at minimum	5.3%	10.5%

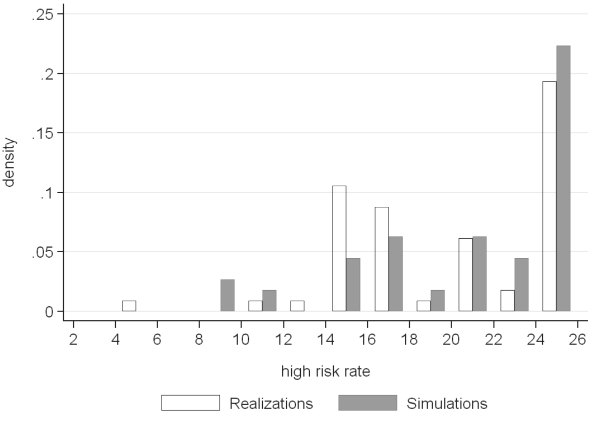
Notes: **Actual** refers to the resulting, end of experiment, distribution in the conditional audit rate group. **Simulated** concerns the audit rates from simulating conditional audit rate updates for the control group.

Figure 3.6: Histograms of realized and simulated end-of-experiment audit rates.

(a) low risk rate



(b) high risk rate



The results in this subsection illustrate that in the conditional audit-rate treatment there was no improvement in quality of applications, nor a change in the number of applications. However, the average audit rate increased, resulting in a larger number of audits. The caseload for assessors increased without generating behavioral responses of care providers. The costs of regulation thus increased due to the conditional audit-rate regime. More generally, because of the lack of behavioral responses, the costs of regulation are lowest when audit rates are low.

3.6 Conclusion

The design of audit policies is concerned with a trade-off between the size of the auditing budget and the incentives for agents. In this chapter, we reported on a field experiment studying the effects of the audit selection rule and the audit rate on compliance behavior of care providers. We considered both unconditional changes to random audit rates as well as conditional, performance-related changes.

We studied this in the Dutch market for immediate care, where an additional trade-off between efficient spending of public resources and quick provision of care exists. The need for quick provision made the gatekeeper introduce the ex-post auditing considered in this chapter. In such a setting possibilities for correction and repayment in case of, for instance, errors in the amount of care provided relative to the care needs, are missing. In case of a disapproved application in an audit, the care provider will be contacted. Because of the lack of a stick, care providers are trusted to some extent. The gatekeeper justified ex-post auditing from the higher approval rate among high-risk applications which traditionally have a higher audit rate than low-risk applications.

Our results do not show evidence that unconditional changes in the audit rate influence the number of applications for immediate care and the approval rate of applications selected for audit. We investigated the presence of heterogeneous effects across various types of care providers, distinguishing them for instance based on size and hospital-status. We did not find evidence for the presence of heterogeneous effects. Finally, even though we did observe divergence in audit rates in the conditional audit regime, this is the result of sorting rather than of behavioral responses of care providers.

The lack of effects on the number of applications and the approval rate suggests that when auditing ex-post, the frequency of audits does not provide additional performance incentives for care providers. Of course, we cannot rule out that the lack of effects is due to the setup of our field experiment with a limited duration and limited sample size, but our preferred explanation is that lack of direct (financial) sanctions in case of disapproval of an audit does not provide incentives to care providers to comply.

The reason for conducting this field experiment was that the Dutch gatekeeper had the idea that trust-based audit policies can be as effective as audit policies using sanctions. The ex-post auditing studied in this chapter fits within the trust-based approach and prior to our study its use has been expanded for cheaper types of long-term care. Based on our results the Dutch gatekeeper decided not to implement such auditing for more extensive types of long-term care with higher audit rates, but to explore alternative audit rules which incorporate sanctions for noncompliance. For example, the gatekeeper is exploring a much larger performance-related variation in audit rates within a control-based audit regime. Our results suggest that gatekeepers or health-care insurers in other countries using trust-based audit rules may want to reconsider this rather than focussing on audit rates to stimulate compliance.

3.A Additional tables

Table 3.A.1: Robustness: level of aggregation.

	# applications			approval rate		
	(1)	(2)	(3)	(4)	(5)	(6)
low	−0.036	−0.037	−0.036	−0.001	−0.000	−0.004
	(0.057)	(0.057)	(0.057)	(0.022)	(0.020)	(0.029)
high	0.025	0.022	0.023	0.000	−0.002	0.007
	(0.126)	(0.126)	(0.126)	(0.021)	(0.017)	(0.027)
conditional	−0.050	−0.054	−0.049	−0.000	−0.002	0.010
	(0.064)	(0.066)	(0.064)	(0.020)	(0.016)	(0.025)
mean control	1.000	1.000	1.000	0.835	0.833	0.835
	(0.013)	(0.012)	(0.014)	(0.027)	(0.026)	(0.027)
period length	2 weeks	1 week	4 weeks	2 weeks	1 week	4 weeks
observations	7,067	14,559	3,538	4,857	8,090	2,792
# providers	226	226	226	224	224	224

Notes: For the control group the mean and standard error of the outcome variable over the pre-experiment period are reported as a reference. Cluster-robust standard errors in brackets, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Columns (1) and (4) repeat the baseline results.

Table 3.A.2: Robustness: various ways of dealing with inactive care providers.

	# applications			**approval rate**	
	(1)	(2)	(3)	(4)	(5)
low	−0.036	−0.013	−0.029	−0.001	0.003
	(0.057)	(0.064)	(0.057)	(0.022)	(0.022)
high	0.025	0.029	0.041	0.000	0.006
	(0.126)	(0.127)	(0.128)	(0.021)	(0.021)
conditional	−0.050	−0.059	−0.033	−0.000	0.003
	(0.064)	(0.075)	(0.063)	(0.020)	(0.019)
mean control	1.000	1.000	1.000	0.835	0.838
	(0.013)	(0.013)	(0.013)	(0.027)	(0.028)
observations	7,067	7,232	6,830	4,857	4,721
# providers	226	226	214	224	212

Notes: For the control group the mean and standard error of the outcome variable over the pre-experiment period are reported as a reference. Cluster-robust standard errors in brackets, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Note that by construction no distinction has to be made between unbalanced and balanced panel data estimates for the approval rate. Columns (1) and (4) repeat the baseline results, column (2) shows balanced panel estimates, and columns (3) and (5) present results when excluding inactive care providers.

Table 3.A.3: Robustness: time trend specification.

	# applications			approval rate		
	(1)	(2)	(3)	(4)	(5)	(6)
low	−0.036	−0.020	−0.004	−0.001	0.005	−0.007
	(0.057)	(0.051)	(0.046)	(0.022)	(0.020)	(0.019)
high	0.025	0.040	0.057	0.000	0.007	−0.005
	(0.126)	(0.122)	(0.121)	(0.021)	(0.018)	(0.017)
conditional	−0.050	−0.035	−0.018	−0.000	0.007	−0.005
	(0.064)	(0.060)	(0.056)	(0.020)	(0.017)	(0.015)
mean control	1.000	1.000	1.000	0.835	0.835	0.835
	(0.013)	(0.013)	(0.013)	(0.027)	(0.027)	(0.027)
time trend	2-week dum.	quarter dum.	2nd order pol.	2-week dum.	quarter dum.	2nd order pol.
observations	7,067	7,067	7,067	4,857	4,857	4,857
# providers	226	226	226	224	224	224

Notes: For the control group the mean and standard error of the outcome variable over the pre-experiment period are reported as a reference. Each specification contains a separate dummy for the time period including Christmas. Cluster-robust standard errors in brackets, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Columns (1) and (4) repeat the baseline results.

Table 3.A.4: Robustness: accounting for serial correlation.

	# applications		**approval rate**	
	(1)	(2)	(3)	(4)
low	−0.036	−0.037	−0.001	−0.033
	(0.057)	(0.057)	(0.022)	(0.054)
high	0.025	0.018	0.000	0.043
	(0.126)	(0.125)	(0.021)	(0.053)
conditional	−0.050	−0.056	−0.000	−0.008
	(0.064)	(0.065)	(0.020)	(0.042)
mean control	1.000	1.000	0.835	0.835
	(0.013)	(0.000)	(0.027)	(0.027)
observations	7,067	444	4,857	428
# providers	226	226	224	224

Notes: For the control group the mean and standard error of the outcome variable over the pre-experiment period are reported as a reference. Baseline results have cluster-robust standard errors in brackets, clustered by care provider; for collapsed estimation heteroskedasticity robust standard errors are reported. * significant at 10%, ** significant at 5% and *** significant at 1%. Columns (1) and (3) repeat the baseline results. Columns (2) and (4) show results when collapsing the data into one pre-treatment and one post-treatment period.

Table 3.A.5: Estimation results: approval rate low-risk and high-risk applications.

	approval rate	
	low-risk	*high-risk*
low	0.022	−0.040
	(0.031)	(0.025)
high	−0.003	−0.025
	(0.024)	(0.024)
conditional	0.037	−0.028
	(0.027)	(0.018)
constant	0.851***	0.934***
	(0.025)	(0.017)
observations	3949	2321
# providers	224	116

Notes: Cluster-robust standard errors are in brackets, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table 3.A.6: Estimation results: composition low/high risk applications.

	fraction low risk applications
low	−0.016
	(0.011)
high	0.006
	(0.015)
conditional	−0.000
	(0.010)
constant	0.812***
	(0.008)
observations	6,836
# providers	226

Notes: Cluster-robust standard errors are in brackets, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table 3.A.7: Estimation results using individual application-level data.

	(1)	(2)	(3)	(4)	(5)
low	0.011	0.001	0.003	0.003	0.004
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
high	−0.001	−0.000	−0.000	−0.000	0.003
	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)
conditional	0.016	0.017	0.018	0.019	0.019
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
mean control	0.835	0.835	0.835	0.835	0.835
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
care provider f.e.	yes	yes	yes	yes	yes
time dummies	yes	yes	yes	yes	yes
care type characteristics	no	yes	yes	yes	yes
care amount characteristics	no	no	yes	yes	yes
patient characteristics	no	no	no	yes	yes
assessor dummies	no	no	no	no	yes
observations	22,279	22,279	22,279	22,279	22,279

Notes: Cluster-robust standard errors in brackets, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Care type characteristics include indicators for high risk category applications, an indicator for personal care, nursing care, assistance, treatment and intramural care. Care amount characteristics include constructed indicator variables for low, moderate and high intensity personal care, and similarly for nursing care. Patient characteristics include a gender dummy and dummies for age categories younger than 50, 50 to 60, 60 to 70, and so on up to older than 90. For almost 50% of observations, the assessor is unknown. We include a separate dummy for this.