Audit regimes in long-term care

Maarten Lindeboom* Bas van der Klaauw† Sandra Vriend‡

April 18, 2020

Abstract

This paper studies the effects of various audit regimes used by a gatekeeper, differing in the degree of enforcement and the presence of performance incentives, on behavior of care providers filing applications for providing long-term care services to patients. We present evidence from a large-scale field experiment in the Dutch market for long-term care. We find that increasing enforcement reduces the number of applications and that introducing performance incentives reduces this even further. Finally, we find detrimental effects on audit approval rates, but we provide some results showing that assessors are less strict when audits do not have direct implications. This implies that an audit regime with performance incentives can support policy makers, who are faced with the trade-off between providing services quickly and efficient spending of public resources.

Keywords: auditing, incentives, field experiment, long-term care, gatekeeper, enforcement
JEL-code: C93, H51, I18

*VU University Amsterdam, and Tinbergen Institute. E-mail: m.lindeboom@vu.nl
†VU University Amsterdam, and Tinbergen Institute. E-mail: b.vander.klaauw@vu.nl
‡Atlas voor gemeenten. E-mail: vriend@atlasvoorgemeenten.nl
Address: Department of Economics, VU University Amsterdam, De Boelelaan 1105, NL–1081 HV Amsterdam, The Netherlands.

Earlier versions of the paper circulated with the titles *Control-based or trust-based auditing: A field experiment in long-term care* and *The effect of audit regimes on applications for long-term care.*

The authors would like to thank the Centrum Indicatiestelling Zorg (CIZ) for facilitating the field experiment and providing the data. We thank the editor, two anonymous reviewers, and conference and seminar participants in Aarhus, Alicante, Amsterdam, Barcelona, Braga, Ljubljana and Rotterdam for useful comments.

1 Introduction

The costs associated with health impairments are seldom born by the individual. Social insurance funds provide sick pay to individuals who report sick from work, insurers pay for the costs of doctor or hospital visits, and long-term care is often financed publicly. To reduce potential moral hazard, many countries have introduced gatekeepers, that have to restrict access to services.[1] Gatekeepers should decide on which (auditing) instruments to use within their limited budget. They could decide on audit frequency, degree of enforcement, (financial) sanctions, and performance incentives. In this paper, we consider the gatekeeper for the public long-term care system in the Netherlands. Using a nationwide field experiment, we study the consequences of changing the audit regime.

In the Netherlands, as in most countries, long-term care expenditures have been increasing steadily over the past decades. Long-term care involves chronic care, either facility-based or home-based, for the elderly and the mentally or physically impaired. The rise in long-term care expenditures introduces a tension between adequate provision of care for those who need it and efficient spending of public resources. Long-term care systems differ across countries both in the types of care covered and in how public expenditures are managed. Roughly two systems exist to control long-term care provision. In the first system, patients apply to a gatekeeper who determines how much care can be provided by a care provider. In the second system, a care provider makes an assessment of needs of the patient that requires approval from a gatekeeper. For detailed cross-country comparisons of the long-term care systems, we refer to Colombo et al. (2011) (chapters 7 and 9 in particular) and Thomson et al. (2013).[2]

When a patient requires long-term care services in the Netherlands, she approaches a care provider. Examples of care providers are nursing homes and companies delivering home-based care. Subsequently, the care provider files an application for (public) funding to the gatekeeper. In the application, the care provider specifies the care needs and asks for a specific type and amount of care services and the way in which this should be delivered to the patient (e.g., facility-based or home-based). The gatekeeper can either approve the application directly or perform an audit. In case of an audit, assessors employed by the gatekeeper study the actual care needs of the patient and compare this to care services that are requested by the care provider.

[1] For example, general practitioners (GPs) act as gatekeepers by giving prescriptions for medication and referring patients to specialized care. Hartman et al. (2013) show that GPs can be effective as gatekeepers, but their costs are high.

[2] Another difference between countries is whether decisions are made at the local (municipal or regional) or at the national level.

Care providers are not allowed to start care provision before the gatekeeper has approved the application. As a consequence, audits delay the moment at which the patient receives care services. Auditing thus affects the speed at which care provision can start. We refer to this as *ex-ante* auditing. Under ex-ante auditing, an audit can result in an approval, denial or adjustment of the application. An adjustment implies that the amount of care for which the provider receives funding, differs from what has been requested. The ex-ante audit regime falls within random auditing as presented by Allingham and Sandmo (1972).

We compare the ex-ante audit regime to two alternative audit regimes. The first is an *ex-post* audit regime, which allows care provision to start immediately after filing the application. Only after the start of care provision, the gatekeeper selects applications for audit. Unlike in ex-ante auditing, the requested services cannot be denied or adjusted in an ex-post audit. Assessors still determine whether the application should have been approved, adjusted or denied and this outcome is communicated to the care providers such that they learn from the feedback. The degree of enforcement of the gatekeeper is thus smaller when applying ex-post auditing compared to ex-ante auditing, but ex-post auditing has the advantage of immediate provision of services to patients. Belgium has a similar system, which relies on frequent (ex-post) random evaluations (Willemé, 2010).[3] In the context of the Netherlands, neither ex-ante nor ex-post auditing allow for (financial) sanctions in case of disapproval. Prior to our field experiment ex-post auditing was already in use for cheaper and less extensive types of long-term care (Lindeboom et al., 2016). The main motivation for the gatekeeper to use ex-post auditing is that it believes that many care providers are intrinsically motivated to file applications correctly and in accordance with actual care needs of patients. Ex-post auditing can then be equally effective as ex-ante auditing, while ensuring immediate access to care services and reducing the bureaucracy experienced by care providers.

In addition to the ex-ante and the ex-post audit regime, we introduce a *conditional* audit regime that relates to the framework described by Harrington (1988). In this regime, the timing of audits (i.e., ex-ante or ex-post auditing) is adjusted regularly during the experiment based on recent approval rates of a care provider. The approval rate, which is the fraction of audited applications that is approved by the gatekeeper, serves as a proxy for application quality and care provider performance. Care providers that perform well, i.e., have high audit approval rates, are exposed to ex-post auditing, while care providers with low audit approval rates are subject to ex-ante auditing. Compared to the ex-ante audit regime, the conditional audit regime has a lower overall degree of enforcement, but enforcement is targeted

[3] In the US ex-post auditing is sometimes applied by insurers and is called "post-claims underwriting". There, sanctions are imposed in case of noncompliance.

towards poorly performing providers. This regime yields fewer delays in care provision than the ex-ante audit regime and introduces incentives for care providers to comply because of the threat of switching between ex-ante and ex-post auditing.

Care providers are randomized into one of the three audit regimes. We study the effects of the audit regime on application behavior of care providers. Our prime outcome of interest is the quantity of applications since this is the main determinant of the long-term care budget. If auditing is efficient, it should reduce moral hazard in demand for long-term care and thus the number of applications. We, therefore, also focus on substitution of applications to other types of long-term care where the strictness of auditing differs. In addition, we focus on the approval rate, but here we are more cautious because the type of auditing (ex-post or ex-ante) may also affect the stringency of the assessor employed by the gatekeeper. Finally, we focus on the number of approved applications. We do not consider health effects of providing patients access to long-term care. These health effects will only become visible in the long run, so beyond our observation period.

Our field experiment is related to another field experiment described by Lindeboom et al. (2016). Both field experiments took place at the same time period and within the same institutional setting. Lindeboom et al. (2016) study the effect of changing audit rates (either conditional or random). Because they consider a setting of ex-post auditing, audit results never have consequences for care provision. Their key findings are that both exogenous and conditional variation in the audit rate do not affect application behavior and compliance rates. Varying the timing of audits as is done in this paper does affect the degree of enforcement and yields direct incentives to comply. Furthermore, the type of long-term care studied in this paper is much more extensive and thus more expensive than the cheap care considered in the field experiment by Lindeboom et al. (2016).

The theoretical literature on auditing and regulatory compliance starts with the traditional (tax) auditing model of Allingham and Sandmo (1972), which interprets the decision of agents to comply with the rules as an economic trade-off of the costs and benefits associated with noncompliance. Whereas, in this traditional model auditing is random, later papers study conditional audit rules (Landsberger and Meilijson, 1982; Greenberg, 1984; Harrington, 1988; Friesen, 2003). Conditional audit rules typically reduce the number of audits and also find a lower compliance rate (e.g., Clark et al., 2004; Cason and Gangadharan, 2006).[4] However, conditional

[4]Blundell (2020) shows that conditional enforcement improves compliance and decreases emissions of harmful pollutants by plants in Florida. A more recent literature considers the performance of audit tournaments, in which the probability of being audited not only depends on one's own behavior, but also on the behavior of others (e.g., Cason et al., 2016; Gilpatric et al., 2011, 2015). The results come from laboratory experiments and show that dynamic audit tournaments have

audit rules usually relate to variations in the audit rate (e.g., Lindeboom et al., 2016). In this paper, the conditional audit rule applies to shifting between ex-ante and ex-post auditing, which may be more visible and has immediate consequences for the applicants.

Using the theoretical literature on random auditing (Allingham and Sandmo, 1972) and conditional auditing (Harrington, 1988), we construct a theoretical model that guides interpretation for the empirical results. The key prediction is that care providers file fewer applications in the ex-ante and the conditional audit regime. The effect on the approval rate is ambiguous. We use a large administrative data set containing information on long-term care applications filed by care providers, and find that when being exposed to the conditional audit regime care providers file 20% fewer applications than in the least strict ex-post audit regime. This can be the result of substitution to other types of long-term care applications. Furthermore, we find an (insignificant) 10% percent decline in the number of applications in the ex-ante audit regime compared to the ex-post audit regime. Our results show some negative effects on the quality of applications (audit approval rate), for care providers in the ex-ante and conditional regimes compared to the ex-post audit regime. This results in about 40% fewer approved applications in both the ex-ante and the conditional audit regime. We provide some evidence that part of this effect can be attributed to behavior of assessors who screen ex-post audits less accurately. Finally, by comparing care providers assigned to participate in our experiment to those who are not part of our experiment, we do not find any evidence for a Hawthorne effect of being assigned to participation in the experiment.

The remainder of this paper is structured as follows. The next section provides institutional details on the Dutch long-term care market. Section 3 describes the experiment and discusses hypotheses on the effects of the audit regimes. The data are described in section 4. The econometric analysis and results are discussed in section 5. Finally, section 6 concludes.

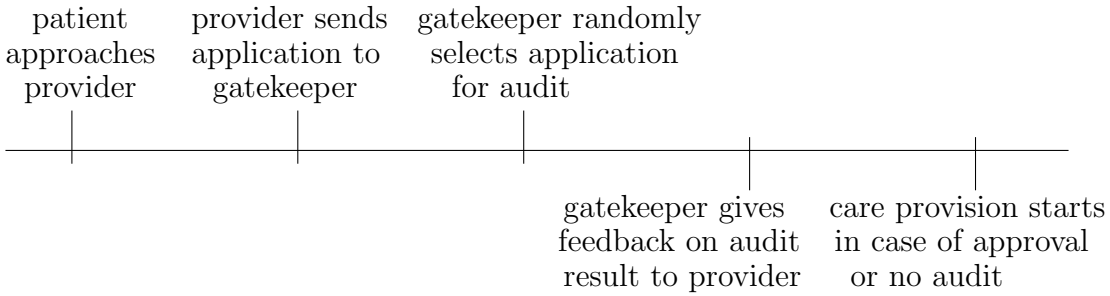
2 Institutional background

Each person living in the Netherlands is publicly insured for long-term care by means of the Exceptional Medical Expenses Act (AWBZ).[5] This covers chronic care, either facility-based or home-based, for the elderly, the mentally and/or physically impaired, and chronic psychiatric patients. On July 1, 2013, around 4.8% of the Dutch

more favorable outcomes that go beyond more truthful reporting.

[5]The description of the institutional details in this section is similar to Lindeboom et al. (2016), and draws on Mot (2010), Nederlandse Zorgautoriteit (2013) and information from the Center for Care Assessment (CIZ).

Figure 1: Process of application for long-term care services (before the experiment).



population qualified for receiving long-term care. Long-term care expenditures are financed by means of general taxation and co-payments. As in many countries, long-term care expenditures in the Netherlands increased substantially in the past decade; real expenditures increased by 60% from 1998 to 2012.

Figure 1 illustrates the process of applying for long-term care services prior to the experiment. When a patient wants to receive long-term care services, she should contact her preferred care provider.[6] The care provider studies the care needs of the patient and files a request to the gatekeeper, which is a national authority that has the task to manage access to long-term care services. The request states which types and amounts of care a patient needs, the way in which this care has to be delivered and for which period.[7] After having received the application, the gatekeeper decides about a so-called *assessment of the needs* stating amongst others the type and amount of care services the patient is eligible for. The gatekeeper can approve the application immediately or initiate an audit. As soon as the decision is made, the provider can start delivering the care to the patient. This implies that a budget is available for providing all approved care. The care provider receives funding from this budget for all care delivered to the patient (within the restrictions set by the assessment of needs).

In 2012, there were around 1,600 care providers in the Netherlands of which almost 1,400 provided home-based and 800 provided facility-based care. Facility-based care providers are required to be non-profit institutions, while home-based care providers can also be for-profit organizations. The care providers differ considerably

[6] A patient can file a request directly to the gatekeeper and indicate the preferred care provider in this request, but around 85% of the applications are filed by care providers. Applications filed by the patient are always thoroughly checked by the gatekeeper before an assessment of needs is provided.

[7] Types of care services 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 stay in an institution (e.g., nursing home).

in size. Some providers are large-scale nursing homes, whereas others consist of a very small number of care givers who visit patients at their homes. Next to goals of profit maximization, for home-based care providers, incentives for the provision of high quality care exist. This is the case since patients are allowed to switch providers during care provision.[8]

A provider can choose between various types of applications. Which application is suitable typically depends on the long-term care history of the patient and the care needs. In case of a first or complex application, the care provider should file an application type that is subject to intensive screening by the gatekeeper. On the contrary, if care is really quickly needed (in emergency cases), an application type for immediate care provision for a short duration could be used. In the field experiment studied in this paper, we focus on the so-called re-assessment applications.[9] Re-assessment applications are used to prolong the provision of existing care services, but often request an adjustment in the intensity or type of services. This can concern facility-based and home-based care. It is not possible to use a re-assessment application in case of a first or a complex application. In 2012, around 70,000 re-assessment applications were filed. Re-assessment applications usually request care provision for a long period of up to 15 years.[10] Hence, the associated expenses are substantial and often much larger than for other types of applications. Therefore, although re-assessment applications are a specific type of application that do not form the majority of applications filed to the gatekeeper in terms of quantity, their importance in the long-term care budget is substantial.

A random sample of around 25% of the re-assessment applications are audited. In practice, every fourth application filed, gets audited. Applications selected for audit are randomly divided among specialized assessors. In an audit, assessors make an assessment of the care needs of the patient in line with eligibility rules set by the government.[11] The outcome of an audit is an 'approval' or 'disapproval'. An approval means that requested care is in line with the needs assessment made by the assessor. Requested care can then be delivered to the patient by the care

[8] In the period from July 2010 to July 2011, about 2.5% of those receiving home-based care through a re-assessment application switched provider and 8% of those receiving facility-based care switched (Nederlandse Zorgautoriteit, 2013).

[9] The gatekeeper selected this application type for the experiment. This choice was amongst others related to the feasibility of implementation of the audit regimes in the software that is used for registration and auditing. The label re-assessment application is used by the gatekeeper.

[10] The median duration of re-assessment applications in our data is 5478 days, i.e. approximately 15 years.

[11] Assessors are instructed to audit according to the rules set by the gatekeeper. They do not have targets on audit approval rates. In line with the goal of the gatekeeper, assessors should aim to audit in such a way that patients receive long-term care in the form and amount that is truly necessary.

provider. A disapproval can either be an adjustment, meaning that requested and adjudged care services can differ, or denial of the application. In the latter case, a new application for long-term care services has to be filed. Disapproval may result for various reasons. For instance, the type of limitations making the patient eligible for long-term care may differ, different types of care services may be needed or the necessary amount of care may be different. After an audit, care providers receive feedback from the assessor, explaining what went wrong and how the quality of future applications can be improved.

Only when an assessment of needs has been provided by the gatekeeper, a care provider is allowed to start care provision. Therefore, when a re-assessment application is selected for audit, care provision is delayed. The care provider has to postpone care provision until approval from the gatekeeper in that case. The outcome of the audit should be made available within two weeks (on average, it is known within a week).

As briefly mentioned earlier, besides the re-assessment applications that are the main focus of our field experiment, providers can choose to file alternative application types. Although there are guidelines for the type of application to be used in a particular situation, the possibility for substitution exists. Alternative types of long-term care applications that providers can file are '*emergency care applications*', '*standard applications*', '*facility-based notifications*', and '*regular applications*'. Emergency care applications make care immediately employable, but only for a period of two weeks after which another application is needed. Standard applications cover a limited set of services concerning cheaper care for a relatively short period. These applications are subject to ex-post auditing.[12] Facility-based notifications concern institutionalized care for patients over 80 years of age. Finally, regular applications are used for first and complex applications. Using this application type, care provision can start within six weeks, after completion of an audit.

3 Experimental design

The goal of our field experiment is to study the effects of being exposed to different audit regimes on application behavior of care providers. In the first subsection we describe the set-up of the field experiment. The randomization of care providers over the three treatment groups is discussed in subsection 3.2. Subsection 3.3 states the hypothesized effects of the audit regimes on application behavior of care providers.

[12]Lindeboom et al. (2016) study the effect of the audit frequency in this ex-post auditing regime for standard applications and found no effect of the audit rate on application behavior of care providers.

Table 1: Consequences of ex-ante and ex-post audits.

	ex-ante	**ex-post**
assessment of needs immediately provided	no	yes
care provisions starts directly after provider files application	no	yes
financial sanctions in case of disapproval	no	no
adjustment in allowed care provision in case of disapproval	yes	no
denial of application is possible	yes	no

3.1 Set-up

Prior to the start of our field experiment on September 17, 2012, re-assessment applications were subject to an ex-ante audit regime. During the experiment, which ran until April 7, 2013, we randomly assigned providers to (1) an ex-ante audit regime, (2) an ex-post audit regime, and (3) a conditional audit regime. Compared to ex-ante auditing, ex-post auditing allows immediate provision of requested care services, but removes the possibility to actually adjust or deny applications. In particular, no (financial) sanctions can be applied in case of disapproval. Moreover, in case of ex-post auditing, care provision can continue without changes to, for example, the type or amount of care, even in case an assessor labels the application as disapproved. The care provider is still, as in ex-ante auditing, informed about the decision of the assessor, but the decision lacks consequences for care provision. Ex-post auditing is thus less strict than ex-ante auditing. Table 1 lists the implications of both ex-ante and ex-post audits.

In the conditional audit regime the timing of audit can switch between ex-ante and ex-post, depending on observed performance in the recent past. In particular, a provider is subject to ex-post auditing if the audit approval rate over the last eight weeks is at least 0.92. When the approval rate falls below 0.92, ex-ante auditing is used.[13] If, by chance, no audits took place, the audit regime remains unaltered. All care providers randomized in the conditional audit regime started at ex-ante auditing, the pre-experiment situation. The timing of audit was updated once every two weeks.[14] These updates were communicated to the care providers supported by a list of results of the relevant audits. All care providers in all treatment groups received the same detailed feedback about the outcomes of their own audits.

In all three treatment groups, the probability that an application is selected for audit is held constant. Assessors were instructed to audit along the same procedures

[13]The threshold of 0.92 is the gatekeeper's target approval rate.

[14] In the first weeks of the experiment, implementation of regime-updating took some more time. As a consequence, the first adjustment was in place four weeks after the start of the experiment, and the second adjustment three weeks after the first update.

Table 2: Balancing of pre-experiment outcomes and provider characteristics across treatment groups (January 1, 2012 - September 16, 2012).

	ex-ante regime	**ex-post regime**	**conditional regime**	**p-value**
Outcome measures				
re-assessment applications (per week)	3.34	3.56	3.26	0.308
	(0.35)	(0.36)	(0.41)	
re-assessment audits (per week)	0.76	0.83	0.70	0.301
	(0.08)	(0.09)	(0.08)	
re-assessment approval rate	0.84	0.83	0.83	0.935
	(0.01)	(0.01)	(0.01)	
Care provider characteristics				
nursing expertise	0.50	0.52	0.61	0.247
	(0.05)	(0.05)	(0.05)	
mental health expertise	0.22	0.25	0.23	0.845
	(0.04)	(0.04)	(0.04)	
care for handicapped expertise	0.26	0.21	0.15	0.181
	(0.04)	(0.04)	(0.04)	
# care providers	101	99	99	

Note: Standard errors of means in parentheses. Reported p-values are for Kruskal-Wallis rank tests for equality of populations.

as before the experiment.

3.2 Implementation

Of the approximately 1,600 long-term care providers in the Netherlands, we selected the 299 providers who filed at least four re-assessment applications in April 2012 for participation in the experiment. These providers are responsible for 80% of all re-assessment applications filed in April 2012. We randomly assigned care providers to one of the three treatment groups discussed in the previous section.[15] Participation was compulsory. Care providers were informed by letter and e-mail about the set-up of the experiment.[16]

[15]The original proposal for the field experiment including power analysis is available at http://personal.vu.nl/b.vander.klaauw/OpzetCIZOnderzoek.pdf [in Dutch]. The pre-experiment data reveal that about 300 care providers, on average, file over four applications per month and the power analysis show that this sample size is sufficient.

[16]An indication of the duration of the experiment was provided to the care providers. However, the experiment continued for some additional months, so that care providers were uninformed about the exact duration and could not anticipate on the end of the experiment.

Table 3: Assessor approval rates by group of assessors.

	before experiment	**during experiment**	
	ex-ante audits	*ex-ante audits*	*ex-post audits*
three assessors doing ex-post audits	0.832	0.777	0.872
	(0.018)	(0.022)	(0.023)
	[N = 3062]	[N = 747]	[N = 4239]
other assessors	0.824	0.811	-
	(0.028)	(0.026)	-
	[N = 4355]	[N = 3185]	-
p-value	0.817	0.364	-

Note: Standard errors of means in parentheses. N denotes the number of audited applications. Only audits by assessors who audited applications both before and during the experiment are included. That is, 12 out of 20 assessors auditing re-assessment applications. P-values are for t-tests for equality of the means (assuming unequal variances).

Table 2 illustrates that the treatment groups are balanced in terms of pre-experiment outcomes and care-provider characteristics. On average, slightly more than three applications are filed per week and of the applications selected for audit, about 83% are approved. Furthermore, the table shows that half of the care providers in each group have an expertise in nursing care, a quarter are specialized in the provision of mental health care and another quarter in the provision of care for the handicapped. These fractions are not significantly different between treatment groups, as shown in the fourth column.

Conditional on ex-ante and ex-post auditing, applications are randomly assigned to assessors. The gatekeeper assigned three assessors to perform the ex-post audits.[17] These assessors also perform ex-ante audits, but they could distinguish whether the audit is ex-ante or ex-post. Table 3 provides descriptives on the pre-experiment fraction of approvals of these three assessors and assessors doing ex-ante audits only. There is no significant difference in the fractions of approval before the experiment and the fraction of approvals of ex-ante audits during the experiment. So the results are not likely to be driven by differences between assessors. However, approval rates of ex-post audits are substantially higher than of ex-ante audits. We return to the role of the assessors in more detail in subsection 5.3.2.

[17]The reason for having three assessors perform ex-post auditing is that this required some changes in the administration of audit results. It was more convenient to train three assessors for this rather than to expose all assessors to the changes in the administration. The three assessors assigned to ex-post auditing are representative for all assessors.

3.3 Theoretical framework

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).

The relationship between gatekeeper and care provider can be characterized as a principal-agent problem. The gatekeeper (principal) has the task to manage the long-term care budget and simultaneously ensure access to long-term care for targeted patients. Care providers (agents) advocate the wishes of their patients, who typically like to receive long-term care as soon as possible. The gatekeeper uses auditing to restrict care providers in their application behavior. We develop a stylized model to illustrate the implications of changing between the three audit regimes used in the field experiment. This model is used to derive predictions that will be tested in our empirical analysis.

Our model is a continuous-time model, where patients arrive at a care provider with a rate λ. The needs of a patient are characterized by θ, which is drawn from a distribution function $F(\cdot)$. A low value of θ means that the patient only has minor limitations that require little care. A high level of θ means serious limitations that require intensive care. The care provider experiences returns $R(\theta)$ from providing care to a patient with needs θ, and we assume $\partial R(\theta)/\partial\theta > 0$.[18] When a patient contacts a care provider, the care provider decides whether or not to file an application to the gatekeeper. Filing an application requires costly effort of the care provider, not only because the care needs of the patient have to be described, but also because of administrative actions associated with the application process. We assume that filing an application requires effort $e \geq e_0$ from the care provider (e_0 is the minimum effort necessary for an application) and the costs of effort $c(e)$ are increasing in e.

When an application is filed, the gatekeeper has to decide whether or not to approve the application. In an audit, an assessor employed by the gatekeeper devotes effort $a \geq a_0$ to learn about the needs θ of the patient. The costs of auditing effort are $k(a)$. When the gatekeeper decides about the audit, it has no information about the quality of the application (both e and θ are unknown). Therefore, the gatekeeper does not differentiate auditing effort between applications filed under the same auditing conditions. An audit results in a (noisy) assessment $\phi(\theta, a, e, \varepsilon)$ of the care needs, where ε is a random audit error. The gatekeeper applies the simple decision rule that an audited application is only approved if $\phi(\theta, a, e, \varepsilon)$ exceeds the

[18] Care providers can be for-profit or non-profit. For-profit care providers earn more when providing more care, while non-profit care providers might have higher returns from helping patients with more serious impairments.

threshold value $\bar{\theta}$, i.e. the gatekeeper believes that needs of the patient are sufficiently high. Auditing effort of the gatekeeper reduces the noise in the assessment $\phi(\theta, a, e, \varepsilon)$, implying $\frac{\partial \Pr(|\phi(\theta,a,e,\varepsilon)-\theta|>\xi)}{\partial a} < 0$ for all values $\xi > 0$. The application effort of the care provider improves the assessment, but with decreasing marginal returns, so $\frac{\partial \phi(\theta,a,e,\varepsilon)}{\partial e} > 0$ and $\frac{\partial^2 \phi(\theta,a,e,\varepsilon)}{\partial e^2} < 0$.

According to the traditional model of Allingham and Sandmo (1972), care providers weigh the benefits of requesting funding for more care and devoting more effort to filing the application against the expected costs of noncompliance (i.e., filing incomplete or incorrect applications) given the degree of enforcement and the level of punishment imposed by the gatekeeper. The profit function of a care provider for a patient with needs θ, is

$$\pi(\theta; A) = \max_{e \geq 0} \left\{ I(e \geq e_0) \Pr(\phi(\theta, a^*, e, \varepsilon) > \bar{\theta})^A R(\theta) - c(e) \right\}$$

where a^* describes the (optimal) audit effort devoted by the gatekeeper and A denotes the auditing regime. In our setting, the gatekeeper can apply ex-ante ($A = 1$) and ex-post ($A = 0$) auditing. Ex-ante auditing implies that care provision cannot start before the application is approved. In case of ex-post auditing funding is granted even if the application is disapproved during an audit. Both ex-ante and ex-post auditing do not yield direct financial sanctions in case of a disapproval.[19]

The gatekeeper aims at minimizing mistakes, so providing funding to patients with care needs θ less than $\bar{\theta}$ or not providing care to patients with need θ above $\bar{\theta}$. This objective of the gatekeeper results in the following criterion function.

$$\max_{a \geq a_0} \Bigg\{ w \int_{\bar{\theta}}^{\infty} \Pr(\phi(\theta, a, e^*(\theta), \varepsilon) > \bar{\theta})^A dF(\theta) - (1-w) \int_{\theta^*}^{\bar{\theta}} \Pr(\phi(\theta, a, e^*(\theta), \varepsilon) > \bar{\theta})^A dF(\theta) \\ -k(a)(1 - F(\theta^*)) \Bigg\}$$

The policy parameter w indicates how much weight the gatekeeper puts on funding care for patients who have needs above $\bar{\theta}$ relative to care for patients with lower needs (below $\bar{\theta}$). The criterion function takes into account that in case of ex-post auditing the gatekeeper cannot decline applications, even if these are disapproved in an audit. θ^* describes the lowest value of θ for which a care provider will file an application; we return to this choice variable below.

In the ex-post regime, care providers do not face the risk of having applications rejected. The optimal behavior of care providers is, therefore, to devote minimal effort e_0 to the application. The costs of making the application are then $c(e_0)$. The care provider only files applications for patients with needs $\theta > \theta_0$, where

[19] The profit function could be extended to incorporate the delay in care provision in case of ex-ante auditing. This pronounces the differences between the audit regimes.

$R(\theta_0) = c(e_0)$. In the ex-post regime $\theta^* = \theta_0$. This means that the rate at which applications are filed to the gatekeeper equals $\lambda(1 - F(\theta_0))$. Also for the gatekeeper there are no incentives to devote effort to auditing. The gatekeeper will thus devote the minimal effort a_0 to auditing and obtains the assessment $\phi(\theta, a_0, e_0, \varepsilon)$. The share of approved applications is given by $1/(1 - F(\theta_0)) \int_{\theta_0}^{\infty} \Pr(\phi(\theta, a_0, e_0, \varepsilon) > \bar{\theta}) dF(\theta)$.

In the ex-ante audit regime, the care provider and the gatekeeper both use a best-response strategy. The care provider makes two decisions: whether or not to file an application and how much effort to devote to the application. Both decisions depend on the needs of the patient θ and the optimal effort a^* devoted by the gatekeeper. The optimal effort of the care provider is described by $e^*(\theta)$, and the care provider only devotes positive effort to applications of patients with needs $\theta > \theta^*$, where $\Pr(\phi(\theta^*, a^*, e^*(\theta^*)) > \bar{\theta}) R(\theta^*) = c(e^*(\theta^*))$. The gatekeeper devotes optimal effort a^*. It is easy to see that $a^* > a_0$, $\theta^* > \theta_0$ and that if this is the case $e^*(\theta) > e_0$ for all $\theta > \theta^*$.

The third regime considered in this paper is the conditional audit regime in the spirit of Harrington (1988). Conditional audit rules have also been studied by Landsberger and Meilijson (1982), Greenberg (1984) and Friesen (2003). In our conditional audit regime, a care provider is audited ex-post if the recent approval rate (Ω_t) of applications exceeds a threshold value $\bar{\Omega}$. An application is audited according to $A = I(\Omega_t < \bar{\Omega})$ and after an audit the recent approval rate is updated as $\Omega_{t^+} = \Omega_t + \Delta\omega$, where $\omega = 1$ implies approval and $\omega = -1$ disapproval during the audit and Δ is a policy parameter describing the speed of updating. The assessors employed by the gatekeeper can distinguish between applications that are audited ex-post and ex-ante, but they do not observe the state variable Ω_t.

In the conditional audit regime the gatekeeper devotes more effort to an ex-post audit than in the ex-post audit regime, where audit effort is minimal. The reason is that the outcome of the audit has longer run consequences, a disapproval may cause that the care provider becomes exposed to ex-ante auditing where the audits have direct consequences. For the same reason the gatekeeper also devotes more effort to ex-ante audits than in the ex-ante regime, i.e. an approval may cause that the care provider gets exposed to ex-post auditing where applications cannot be declined.

The care provider knows the level of its state variable Ω_t. The value function of a care provider who submits an application for care provision to a patient with needs θ is

$$\begin{aligned}\pi(\theta; \Omega_t) = \max_{e \geq 0} \Big\{ I(e \geq e_0) \Big\{ &\Pr(\phi(\theta, a^*_{I(\Omega_t < \bar{\Omega})}, e, \varepsilon) > \bar{\theta})^{I(\Omega_t < \bar{\Omega})} R(\theta) - c(e) \\ &+ \frac{\lambda}{1+\rho} \int_0^{\infty} \Big(\Pr(\phi(\theta, a^*_{I(\Omega_t < \bar{\Omega})}, e, \varepsilon) > \bar{\theta}) \pi(\theta'; \Omega_t + \Delta) \\ &+ \Pr(\phi(\theta, a^*_{I(\Omega_t < \bar{\Omega}, \varepsilon)}, e) < \bar{\theta}) \pi(\theta'; \Omega_t - \Delta) \Big) \, dF(\theta') \Big\} \Big\}\end{aligned}$$

The care provider devotes most effort to applications when the level of the state variable Ω_t is close to $\bar{\Omega}$. Then the outcome of an audit strongly determines future enforcement by the gatekeeper. When Ω_t is very low, the care provider needs many approved audits to become exposed to ex-post auditing. Because the implications of the audit results are limited the care provider behaves quite similar as in the ex-ante regime (with the exception that the gatekeeper devotes more effort to auditing). When Ω_t is very high, then a disapproval does not have direct consequences on how the care provider will be audited in the future. This causes that the care provider is less selective in which applications to file to the gatekeeper and reduces application effort.

We use our theoretical framework to formulate a number of testable implications. We also consider the special case where the gatekeeper applies the same auditing effort a^* to all audits. When designing the experiment it was the intention of the management of the gatekeeper to audit ex-post as thoroughly as ex-ante, but this could not be monitored. We first compare the ex-ante and the ex-post regime. Recall that $a^* > a_0$, $\theta^* > \theta_0$ and $e^*(\theta) > e_0$ for all $\theta > \theta^*$.

Hypothesis 1 *The rate at which applications are filed in the ex-post regime is higher than in the ex-ante regime.*

In the ex-post regime the rate at which applications are made equals $\lambda(1 - F(\theta_0))$, while this is $\lambda(1 - F(\theta^*))$ in the ex-ante regime. Since $\theta^* > \theta_0$, it is the case that $\lambda(1 - F(\theta_0)) > \lambda(1 - F(\theta^*))$. This holds both in case assessors optimize their effort and when they comply to $a = a^*$.

Hypothesis 2 *When all regimes apply $a = a^*$, then the approval rate is higher in the ex-ante regime than in the ex-post regime. Otherwise, the sign of the difference in approval rates is ambiguous.*

The additional applications that will be filed in the ex-post regime compared to the ex-ante regime all have needs θ between θ_0 and θ^*. These are applications that have a low approval rate. So when $a = a^*$ in all regimes, then the ex-post regime has a lower approval rate than the ex-ante regime.

When auditing effort differs between regimes, the approval rate of applications changes from $1/(1 - F(\theta_0)) \int_{\theta_0}^{\infty} \Pr(\phi(\theta, a_0, e_0, \varepsilon) > \bar{\theta}) dF(\theta)$ in the ex-post regime to $1/(1 - F(\theta^*)) \int_{\theta^*}^{\infty} \Pr(\phi(\theta, a^*, e^*(\theta), \varepsilon) > \bar{\theta}) dF(\theta)$ in the ex-ante regime. The audit approval rate increases because care providers are more selective in the applications that they file ($\theta^* > \theta_0$) and they devote more effort to each application ($e^*(\theta) > e_0$). However, also the gatekeeper devotes more effort ($a^* > a_0$), which may reduce the

audit approval rate. Which of these mechanisms dominates depends on the costs of effort of both the care provider and the gatekeeper. If for the care provider the costs of filing an application are high, then θ^* will increase substantially compared to θ_0 and the care provider will not increase the effort for filing applications much. If also the gatekeeper experiences substantial costs of auditing effort, the gatekeeper maintains low effort. Selectivity in filing applications then is the main behavioral response causing an increasing approval rate. If the care provider faces very low costs of effort, then the care provider will continue filing many applications. If also the gatekeeper faces low costs of effort, audits will be much more thorough in the ex-ante regime. This may decrease the approval rate.

Hypothesis 3 *The number of approved applications is higher in the ex-ante regime than in the ex-post regime.*

This follows from the cases described above.

Hypothesis 4 *The rate at which applications are filed in the ex-post regime is higher than in the conditional regime.*

Hypothesis 5 *When all regimes apply $a = a^*$, then the approval rate is lower in the ex-ante regime than in the conditional regime. Otherwise, the sign of the difference in approval rates is ambiguous.*

Hypothesis 6 *The number of approved applications is higher in the ex-ante regime than in the conditional regime.*

To motivate these final three hypotheses, the same arguments apply as for the first three hypotheses.

How the conditional audit regime compares to the ex-ante regime depends on how tight the threshold value $\bar{\Omega}$ is set. If complying to the threshold is easy (low value of $\bar{\Omega}$), then the conditional audit regime is very similar to the ex-post regime. In case $\bar{\Omega}$ is very high, the conditional audit regime becomes similar to the ex-ante regime. The gatekeeper should try to set $\bar{\Omega}$ such that most care providers comply to this threshold with some extra effort. A number of our hypotheses depend on how costly effort is for both the care provider and the gatekeeper and how demanding it is to comply to the threshold $\bar{\Omega}$ in the conditional audit regime. The results of the empirical analysis will thus be informative on these parameters.

The theoretical model makes a few simplifying assumption. First, the model assumes that all applications are audited, while in reality only a random subsample of 25% of the applications is audited. This does not affect the differences between regimes. Next, an additional attractive feature of ex-post auditing is that patients can be served immediately without waiting for the result of the audit. Furthermore, the gatekeeper might have some interest in providing feedback to care providers to help them file proper applications. Finally, care providers may have some intrinsic motivation to comply. A higher degree of enforcement (ex-ante instead of ex-post auditing) may crowd out some intrinsic motivation.[20] Mendoza and Wielhouwer (2015) discuss conditions when more trust based auditing is preferable. Our conditional audit regime shares some of these conditions. The model could be enriched to take account of these issues. This would complicate the model and may make differences between regimes less pronounced, but it does not change the main behavioral responses to audit regime changes.

4 Data

We have access to administrative data containing all applications filed by all care providers between January 1, 2012 and November 24, 2013. This implies that we observe a pre-experiment period (January 1, 2012 - September 16, 2012), the experiment period (September 17, 2012 - April 7, 2013) and a post-experiment period (April 8, 2013 - November 24, 2013). In the empirical analysis we mainly consider the pre-experiment and the experiment period, but in subsection 5.2.1 we also look at behavior after the experiment and long-term consequences of the experimental variation in audit regimes. There are 642,921 applications filed by care providers participating in the experiment.[21]

Re-assessment applications, which are subject to the experimental variation, account for 17.3% of the applications of participating care providers. Of the 111,024 re-assessment applications, 22.2% (24,654 applications) were selected for audit. For each application we observe the application date, the date of audit, the care provider filing the application, the type of application, and the type, requested amount and maximum duration of care services. Furthermore, we observe the assessor who performed the audit, the result of the audit (approved or disapproved) and the reasons

[20] Bohnet et al. (2001) conclude from a laboratory experiment that the longer agents are trusted (low enforcement probabilities), the more likely they are to act trustworthy. Also in a laboratory experiment Falk and Kosfeld (2006) find that most agents reduce performance when the principal chooses to enforce.

[21] Only in subsection 5.2.2 we also consider applications of care providers not participating in the experiment.

Table 4: Application characteristics and patient characteristics (January 1, 2012 - April 7, 2013).

	ex-ante	**ex-post**	**conditional**
Types of re-assessments			
change in home-based care	69%	73%	75%
prolong (unchanged) facility-based care	16%	13%	11%
change in facility-based care	12%	11%	11%
other	3%	3%	3%
Types of care			
personal care	48%	54%	58%
nursing care	17%	19%	20%
assistance	27%	27%	26%
special medical treatment	1%	1%	1%
facility-based care	31%	27%	25%
Patient characteristics			
female	57%	58%	59%
age (in years)	62.3	64.3	66.2
older than 80	31%	33%	36%
married	22%	23%	24%
# applications	23,550	26,859	22,037

for this audit decision. Data on applications provided by the gatekeeper are linked to administrative micro data from Statistics Netherlands, providing information on basic patient characteristics including age, gender and marital status.[22]

Descriptive statistics on application and patient characteristics for whom re-assessment applications are filed, are provided in Table 4. The first panel shows the relative importance of various types of re-assessment applications. By far the most frequently filed re-assessment application is changing an assessment for home-based care. A single application can be concerned with multiple types of care. The most frequent type of care is personal care, followed by assistance and facility-based care. Finally, the table shows that more re-assessment applications are filed for women, and patients are, on average, approximately 64 years old. Around one-third of the patients are over 80 years of age and slightly over 20% are married. This may relate to the presence of informal care options, which is taken into account in determining eligibility for formal long-term care. Additional descriptives on outcomes during the experiment are provided in Table A.3 in the appendix.

[22]This linkage is available for all applications filed by participating care providers before and during the experiment, but not for applications filed after the experiment.

Changes to the timing of audit for care providers in the conditional regime are registered in the data every two weeks. During the experiment, there were twelve opportunities for providers to switch between ex-ante and ex-post auditing. On average a provider switches 1.8 times during the experiment. Appendix A.2 provides details on switching behavior and plots the distribution of switches.

When studying the effects of audit regimes on the number of applications, we aggregate the data in periods of two weeks at the care provider level. For ease of interpretation we normalize the number of applications to weekly averages for each care provider. The resulting panel data set is unbalanced, because providers may become inactive due to closing down activities or merging with another provider.[23]

5 Results

Our key interest is the number of applications. But we also study the quality proxied by the audit approval rate and the number of approved applications. Subsection 5.1 provides descriptive evidence on the trends in these outcomes. In subsection 5.2 we quantify the effects of the audit regime on the number of applications. Subsection 5.3 looks at effects on the quality of applications.

5.1 Descriptive evidence

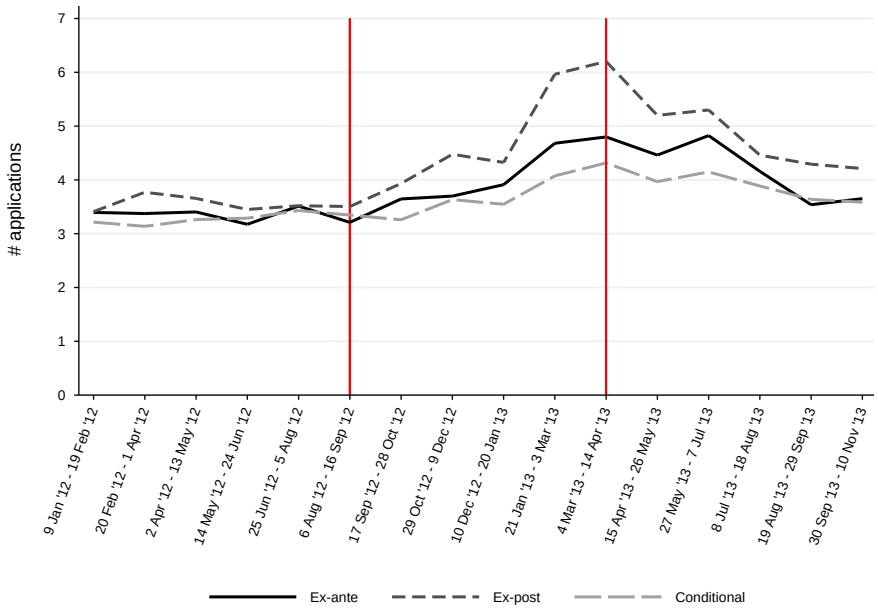
Figure 2 shows the trend in the average number of re-assessment applications in each treatment group before, during and after the experiment. There are no differences between the treatment groups before the experiment, but during the experiment the number of applications diverges. The increase in the number of applications is strongest in the ex-post audit regime and smallest in the conditional audit regime. After the experiment ended, the number of applications in the treatment groups converged again.

Next, we are interested in the audit approval rate as an outcome measure. Figure 3 shows approval rates over time by treatment group. The increase in the number of applications in the ex-post audit regime is accompanied with a somewhat higher approval rate. There are no clear differences in approval rate between the ex-ante and conditional audit regime.

Both the number of applications and the audit approval rate influence the number of approved applications. Figure 4 shows the trend in the average number of approvals. Recall that only 25% of the applications are audited, so the number of

[23]Details on the construction of the panel data set, such as the level of time aggregation and the way to deal with inactivity of care providers, are provided in the appendix. Tables A.4 and A.5 in the appendix show that results are robust with respect to these choices.

Figure 2: Weekly average number of re-assessment applications in the three treatment groups.



Note: The experiment period is between both vertical lines. Before and after the experiment the ex-ante audit regime was applied.

approved applications is more than four times lower than the number of applications. During the experiment the number of approved applications diverges mainly for the ex-post regime, which confirms the higher number of applications and higher approval rate for this regime. After the experiment ended, the number of approved applications in each of the three regimes converged.

5.2 Number of applications

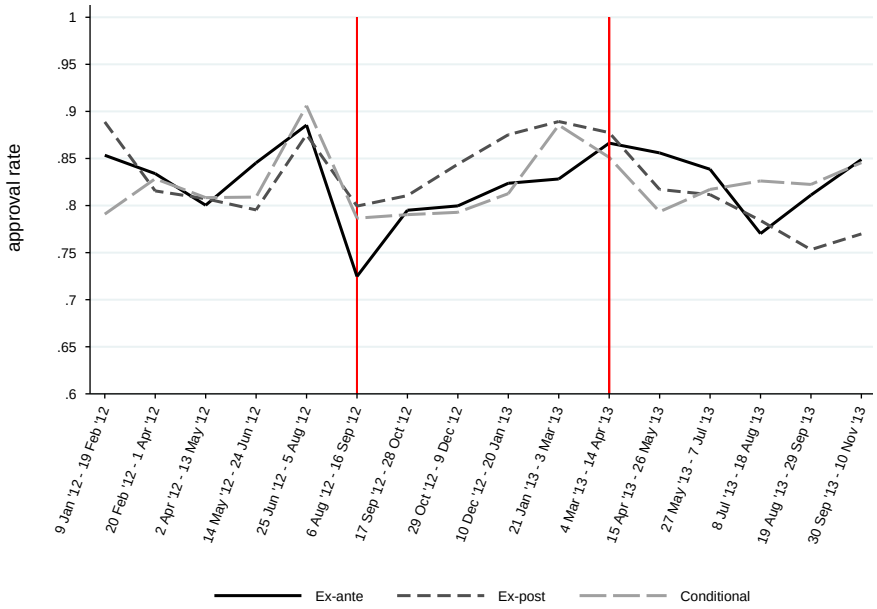
We estimate linear panel data models on data before and during the experiment to obtain the causal effects of the audit regime on the number of applications for each care provider within a period of two weeks,

$$Y_{c,t} = \alpha + \gamma_t + \delta_{\text{ex-ante}} T_{c,t}^{\text{ex-ante}} + \delta_{\text{conditional}} T_{c,t}^{\text{conditional}} + \eta_c + \varepsilon_{c,t} \tag{1}$$

Here c indexes the care provider and t is the time period. We include time fixed effects (γ_t) to account for common time trends.[24] Because data are from a randomized experiment, the treatment status is independent from the care provider

[24]We also estimated models with alternative specifications of the common time trend (i.e., polynomials in time and quarter dummies combined with an indicator for the experiment period). Results are very similar to our baseline estimates and available on request. Furthermore, recall that in the beginning of the experiment regime-updating took some more time (see footnote 14). Results are robust to excluding the first two, four or six weeks of the experiment from the analysis.

Figure 3: Average approval rates in the three treatment groups.



Note: The experiment period is between both vertical lines. Before and after the experiment the ex-ante audit regime was applied.

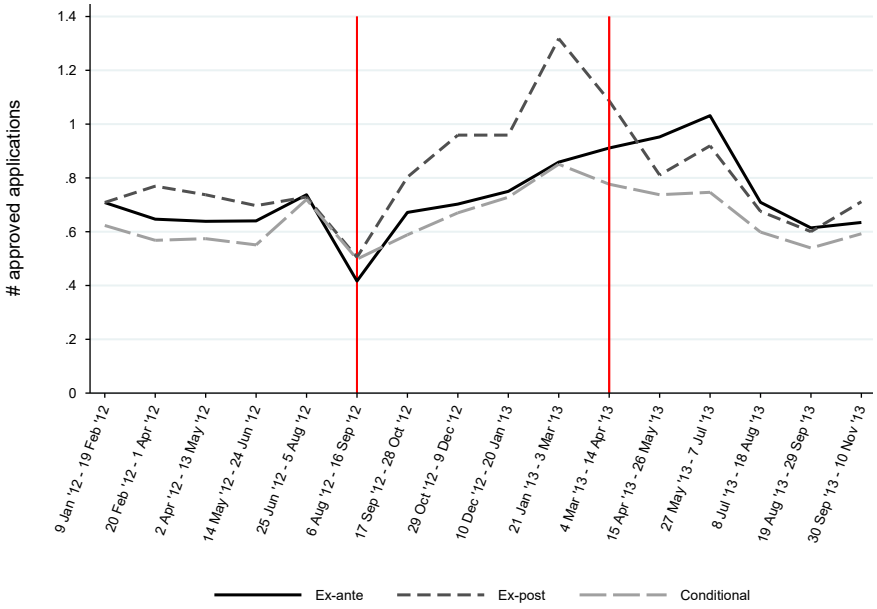
specific effects η_c, which we, therefore, model as random effects.[25] The treatment variables $T_{c,t}$ equal one if care provider c belongs to a treatment group in period t.[26] The parameters of interest are the δ's, which describe the causal effects of the audit regime. Even though ex-ante auditing was applied before the experiment, we take the ex-post audit regime as the reference group in all estimations. The ex-post audit regime is the least strict regime and our hypotheses formulated in subsection 3.3 are mainly in comparison to this regime. We report effects both in levels and in fractions. For the latter, we normalize the number of applications of a care provider in each period by dividing by the pre-treatment provider-specific average number of applications. Because of this normalization we eliminate provider specific effects.

The baseline estimation results in Table 5 show a substantial, but insignificant, decline of 0.626 applications per week when providers are exposed to the ex-ante audit regime compared to the ex-post audit regime. The decline in the conditional audit regime is 0.954 applications per week and is significant. This corresponds to a 20% decline of the number of applications in the conditional audit regime compared

[25]Fixed effects estimation gives almost the exact same estimates and standard errors. This is also the case when using OLS with cluster-robust standard errors. These estimates can be found in the online appendix.

[26]For the estimation we use data from both the pre-treatment and the experiment period. Only using the experiment period only modestly changes the estimates, but slightly increases the standard errors.

Figure 4: Weekly average number of approved re-assessment applications in the three treatment groups.



Note: The experiment period is between both vertical lines. Before and after the experiment the ex-ante audit regime was applied.

to the ex-post audit regime. Both effects confirm hypotheses 1 and 4 from our theoretical model. The larger effect of the conditional audit regime can be explained by larger opportunity costs of invalid applications in the conditional regime than in the ex-ante audit regime, due to the presence of the threat of switching from ex-post to ex-ante auditing.[27] The difference between the effects of the ex-ante and the conditional audit regime is not significant, but we reject the null hypothesis of no effect of any of the treatments as indicated by the F-test p-values in the table.

Note that 10% and 20% of the pre-experiment mean in the ex-post audit regime are smaller than the effects in levels. This suggests that the effects of the audit regime changes are different for small and large care providers. This may be caused by differences in the speed with which feedback on audits reaches the persons filing the applications within the organization, or it may relate to differences in the resources and time available for accurately filing applications. To study heterogeneous effects we interact the time trend and the treatment indicators in equation (1) with dummy variables for the size of a provider.[28,29] The estimation results in Table 6 show that

[27] In appendix A.2 we discuss the switches of providers between ex-ante and ex-post auditing in the conditional audit regime. We find that sorting of providers, based on pre-experiment approval rates, can explain the switches that we observe.

[28] The model specification equals $Y_{c,t} = \alpha + \gamma_{d_c,t} + \delta_{\text{ex-ante},d_c} T_{c,t}^{\text{ex-ante}} + \delta_{\text{conditional},d_c} T_{c,t}^{\text{conditional}} + \eta_c + \varepsilon_{c,t}$, with d_c indicating care provider type.

[29] In addition, heterogeneity in effects for providers differing in terms of the share of re-assessment

Table 5: Baseline estimates for the number of applications (per week) at the provider level (January 1, 2012 - April 7, 2013).

	# applications		**substitution**	
	level	*normalized*	*all types, excl. re-assessments*	*emergency*
	(1)	(2)	(3)	(4)
ex-ante audit regime	−0.626	−0.104	0.418	0.165
	(0.411)	(0.079)	(0.532)	(0.155)
conditional audit regime	−0.954**	−0.203***	0.763	0.485*
	(0.398)	(0.078)	(0.965)	(0.250)
mean ex-post audit regime	3.557	1.000	16.973	2.701
(before experiment)	(0.362)	(0.000)	(2.768)	(0.572)
care provider random effects	yes	no	yes	yes
p-value joint significance	0.032	0.030	0.635	0.147
p-value ex-ante = conditional	0.125	0.121	0.706	0.171
observations	9286	9286	9286	9286
# care providers	299	299	299	299

Note: Pre-experiment mean and standard error of the outcome in the ex-post audit regime are reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. The p-value for the F-test is for joint significance of the coefficient on the ex-ante audit regime and the conditional audit regime.

large providers are most responsive to the audit regime. Large care providers file 37.9% and 36.8% fewer applications when being exposed to respectively the ex-ante audit regime and the conditional regime compared to the ex-post audit regime. Small providers respond less in general. They respond most strongly to the conditional audit regime. We do not find any evidence that the medium-size care providers adapt their behavior in response to the audit regime change.

The decline in the number of re-assessment applications in the ex-ante and conditional audit regimes compared to the ex-post audit regime may be explained by substitution towards other application types. The content of standard applications and facility-based notifications is substantially different from the typical re-assessment application. Regular applications are subject to more stringent screening than re-

applications in total applications is studied. We find some evidence that providers having a small share of re-assessment applications, respond more strongly to the ex-ante and conditional audit regimes. Results are available on request.

Table 6: Heterogeneous effect estimates on normalized number of applications.

Heterogeneity in provider size		**Heterogeneity in initial approval rates**	
	(1)		(2)
ex-ante × small	−0.151	ex-ante × low	−0.025
	(0.168)		(0.091)
ex-ante × medium	0.047	ex-ante × intermediate	−0.298**
	(0.092)		(0.149)
ex-ante × large	−0.377**	ex-ante × high	−0.127
	(0.157)		(0.244)
conditional × small	−0.280*	conditional × low	−0.087
	(0.152)		(0.095)
conditional × medium	0.001	conditional × intermediate	−0.419***
	(0.104)		(0.159)
conditional × large	−0.366**	conditional × high	−0.341*
	(0.150)		(0.202)
p-value joint significance, small	0.162	p-value joint significance, low	0.612
p-value joint significance, medium	0.843	p-value joint significance, intermediate	0.027
p-value joint significance, large	0.041	p-value joint significance, high	0.178
p-value ex-ante = conditional, small	0.309	p-value ex-ante = conditional, low	0.437
p-value ex-ante = conditional, medium	0.644	p-value ex-ante = conditional, intermediate	0.381
p-value ex-ante = conditional, large	0.904	p-value ex-ante = conditional, high	0.254
controls	no	controls	no
observations	9286	observations	9286
# providers	299	# providers	299

Note: 34% (21%) of participating providers are categorized as filing few (many) re-assessment applications in the pre-experiment period, the remaining 45% of providers file an intermediate number of applications; 59% (22%) of participating providers are categorized as having low (intermediate) initial approval rates. The models include dummies for small and large providers (medium-sized providers are the reference) or dummies for low and high initial approval rates (intermediate initial approval rates is the reference). Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Reported p-values are for F-tests testing the null hypothesis of no effect of any regime for a particular type of provider.

assessment applications, but emergency care applications are audited less strictly than re-assessment applications. Hence, if care providers substitute, they most likely will substitute to emergency care applications.

To test for the presence of substitution effects, we estimate the baseline model in equation (1) having as outcomes the number of emergency care applications and the total number of applications other than re-assessment applications.[30] Estimation results are reported in columns (3) and (4) of Table 5. The estimated substitution effects are quite substantial, but hardly significant. Only for providers in the conditional audit regime we find a marginally significant substitution effect: the decline in the number of re-assessment applications is accompanied by an increase in the number of emergency care applications.

Some re-assessment applications may be easier to substitute with emergency care applications than others, which could cause compositional changes of the re-assessment applications. But also the audit regime itself may affect the composition of re-assessment applications. However, we find no compositional changes induced by the audit regimes, as shown in Table A.6 in the appendix. In particular, there is no shift in patient characteristics (age and gender) and types of re-assessment applications.[31]

5.2.1 Short-run, long-run and post-treatment effects

Figure 2 illustrated that the weekly number of applications was stable in the pre-experiment period, but increased in the fourth quarter of 2012. From the beginning of 2013, the levels remained relatively constant again. Furthermore, the figure shows a divergence of the average number of applications between the treatment groups during the experiment, and a subsequent convergence afterwards. Therefore, we extend our model to estimate separately short-run effects (fourth quarter 2012),

[30] In most cases, when an emergency care application is filed for a particular patient, two entries are recorded in the data: one for immediate care provision and another for care provision after two weeks. In all empirical analyses, we ignore this second application.

[31] We estimated the panel data model in equation (2) on re-assessment application-level data.

$$A_{i,c,t} = \alpha + \gamma_t + \zeta_{\text{ex-ante}} T_{i,c,t}^{\text{ex-ante}} + \zeta_{\text{conditional}} T_{i,c,t}^{\text{conditional}} + \eta_c + \upsilon_{i,c,t} \tag{2}$$

Here $A_{i,c,t}$ is an indicator for whether or not application i filed by provider c at time t has a particular characteristic (e.g. application for an old patient). We again include provider random effects, a general time trend and treatment indicators for the ex-ante and conditional audit regimes. The effects of interest, ζ, measure the change in the probability that the application has the characteristic A.

long-run effects (first quarter 2013) and post-experiment effects,

$$
\begin{aligned}
Y_{c,t} = \alpha + \gamma_t &+ \delta_{\text{ex-ante,short}} T_{c,t}^{\text{ex-ante,2012Q4}} + \delta_{\text{conditional,short}} T_{c,t}^{\text{conditional,2012Q4}} \\
&+ \delta_{\text{ex-ante,long}} T_{c,t}^{\text{ex-ante,2013Q1}} + \delta_{\text{conditional,long}} T_{c,t}^{\text{conditional,2013Q1}} \\
&+ \zeta_{\text{ex-ante}} P_{c,t}^{\text{ex-ante}} + \zeta_{\text{conditional}} P_{c,t}^{\text{conditional}} + \eta_c + \varepsilon_{c,t}
\end{aligned} \tag{3}
$$

We distinguish between a treatment indicator for the fourth quarter of 2012 and for the first quarter of 2013. Furthermore, P is an indicator for the post-experiment period. The parameters ζ describe possible treatment effects after the experiment ended and all care providers were again exposed to the ex-ante audit regime.

The estimation results in Table 7 show that the treatment effects are considerably more pronounced in the long-run than just after the start of the experiment. It may have taken some time for care providers to become fully aware of the implications of the audit regime change, to search for suitable alternatives for filing applications and to adapt their behavior accordingly. This is especially true for the conditional audit regime, which includes specific performance incentives. After the experiment, there is no longer a significant difference in the number of applications filed by providers in the ex-ante and conditional regimes relative to providers in the ex-post regime. Care providers thus reversed their application behavior after the experiment, indicating that the temporary change in audit regime did not cause habit formation.

5.2.2 Hawthorne effect

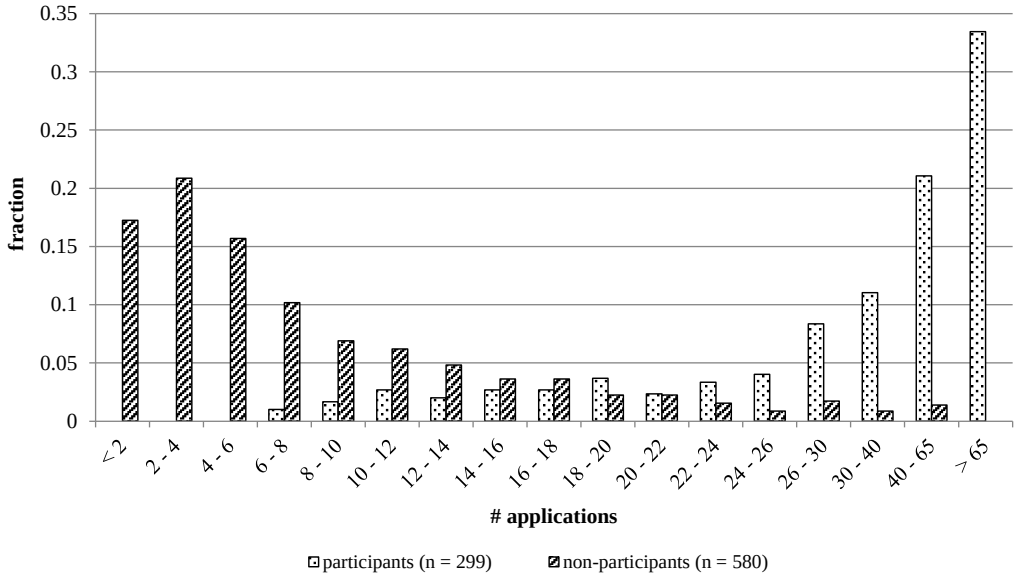
Before the start of the experiment all care providers were exposed to the ex-ante audit regime. Recall from Figure 2 that after the start of the experiment the average number of applications of care providers randomized in the ex-ante audit regime started increasing. This might be a Hawthorne effect, implying that participation in the experiment already affects outcomes. A possible explanation might be that informing providers about the experimental variation in auditing regimes for re-assessment applications draws more attention to the possibility of filing this application type, thereby influencing the number of re-assessment applications. To test for a Hawthorne effect, we compare participants in the experiment to non-participants. Recall that we selected all providers having filed at least four re-assessment applications in April 2012 to participate in the experiment. For each care provider, participating in the experiment or not, we compute the total number of re-assessment applications between February and June 2012, so the pre-experiment period. The total number of applications ranges from 6 to 642 among the participating care providers and from 1 to 64 among the non-participating care providers. Figure 5 shows the distributions of the total number of applications filed by participants and non-participants.

Table 7: Short-run, long-run and post-treatment effects (January 1, 2012 - November 24, 2013).

	# applications	
	level	*normalized*
	(1)	(2)
ex-ante regime, short-run	−0.322	−0.053
	(0.290)	(0.064)
ex-ante regime, long-run	−1.096*	−0.222*
	(0.624)	(0.134)
ex-ante regime, post-experiment	−0.489	−0.037
	(0.364)	(0.112)
conditional regime, short-run	−0.496*	−0.090
	(0.279)	(0.065)
conditional regime, long-run	−1.575***	−0.354***
	(0.609)	(0.133)
conditional regime, post-experiment	−0.662	0.013
	(0.437)	(0.114)
mean ex-post audit regime (before experiment)	3.557	1.000
	(0.362)	(0.000)
p-value, short-run	0.161	0.382
p-value, long-run	0.026	0.027
p-value, post-experiment	0.283	0.916
p-value ex-ante = conditional, short-run	0.281	0.544
p-value ex-ante = conditional, long-run	0.171	0.165
p-value ex-ante = conditional, post-experiment	0.616	0.692
care provider random effects	yes	no
time fixed effects	yes	yes
care characteristics	no	no
patient characteristics	no	no
assessor indicators	no	no
observations	13,730	13,730
# care providers	299	299

Note: Pre-experiment mean and standard error of the outcome in the ex-post audit regime are reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Reported p-values are for F-tests of the null hypothesis of no effect of any of the audit regimes in the short-run, long-run and after the experiment.

Figure 5: Distribution of number of re-assessment applications for participating and non-participating care providers, February 2012 - June 2012.

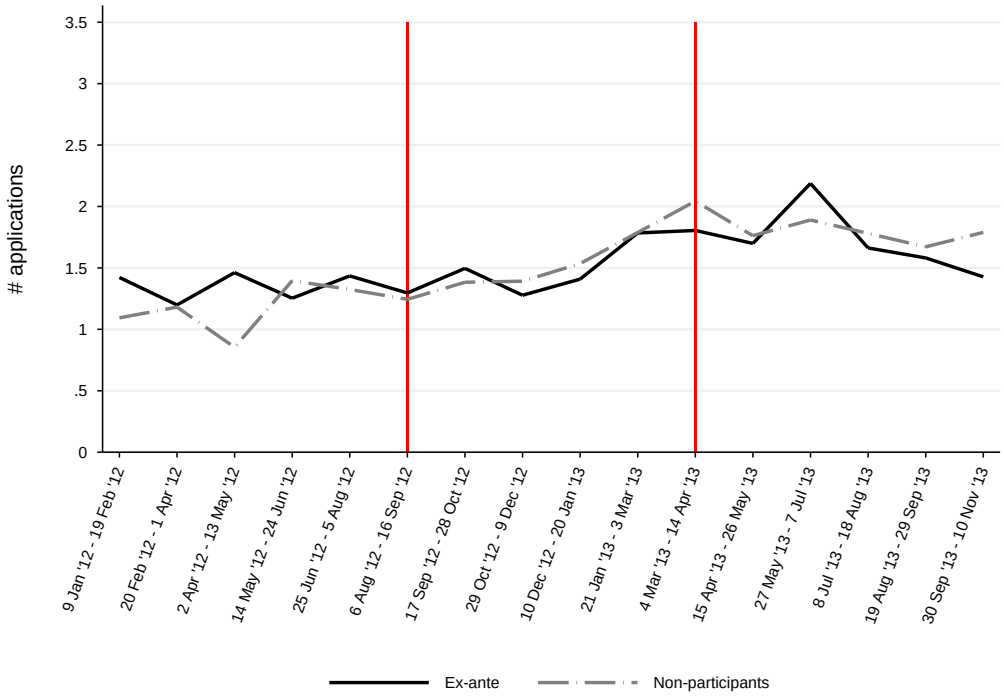


We select among both groups the care providers filing between 6 and 45 applications, so the range in which there is an overlapping support.[32] Next, we weigh the non-participating care providers, such that the weighted distribution of the total number of applications is the same as the observed distribution among the participating care providers. Thus within each cell shown in Figure 5 we give each non-participating care provider a weight based on the ratio of the number of participating and non-participating care providers in the cell. When an increasing trend in the number of applications similar to the trend for the ex-ante audited providers is found for the (weighted) group of non-participating providers, a Hawthorne effect is unlikely to explain the findings in Figure 2. Figure 6 illustrates the trends in the number of applications over time for the ex-ante regime and the non-participants. These trends are very comparable.[33] This suggests that participation in the experiment itself did not cause behavioral responses. The increase in the number of applications in the fourth quarter of 2012 thus seems a genuine time trend rather

[32]That implies that we focus on 150 of 299 participating providers and 266 of 580 non-participating providers.

[33]To test this more formally, we estimate weighted panel fixed effects models similar to our baseline specification in equation (1) for the number of applications filed by non-participants and care providers in the ex-ante audit regime. We take non-participants as the reference group and include a treatment indicator for the ex-ante audit regime. As weights we use the weights assigned to the non-participants in order to equalize the distribution of the total number of applications to that of the participating providers. The estimation results in Table A.7 in the appendix show that there is no significant difference in the (normalized) number of applications in the ex-ante audit regime compared to the non-participants.

Figure 6: Average number of re-assessment applications, providers in ex-ante regime and non-participants.



than an effect of the experiment.

5.3 Quality of applications

For the audit approval rate we estimate a linear panel data model on application-level data,

$$Q_{i,c,t} = \alpha + \gamma_t + \delta_{\text{ex-ante}} T^{\text{ex-ante}}_{i,c,t} + \delta_{\text{conditional}} T^{\text{conditional}}_{i,c,t} + \beta X_{i,c,t} + \eta_c + \upsilon_{i,c,t} \quad (4)$$

$Q_{i,c,t}$ is a dummy for approval of application i filed by provider c at time period t. Again we include provider random effects η_c and time fixed effects γ_t. Furthermore, we sequentially include various sets of application characteristics, $X_{i,c,t}$. These include indicators for the type of re-assessment, indicators for types and amounts of care, patient characteristics and indicators for the assessor performing the audit. The causal effect estimates (δ) represent percentage point changes in the approval probability.

The estimation results are presented in Table 8. The approval probability declines by about four percentage points when providers are exposed to the ex-ante audit regime instead of the ex-post audit regime. We estimate a 2.5 percentage points decrease for switching to the conditional audit regime.[34] We can reject the

[34] As for the number of applications, we also estimated short-run, long-run and post-treatment

null hypothesis that both the ex-ante audit regime and the conditional audit regime simultaneously have no effect on the approval probability. This also rules out that assessors devoted equal effort to ex-ante and ex-post auditing (see hypotheses 2 and 4 of the theoretical model). The direction of the effects is consistent with assessors devoting less effort to audits that cannot have (direct) consequences. The estimates for the treatment effects are robust against including covariates. Only including the assessor indicators in column (4) causes the effects to change slightly. This might be influenced by the fact that eight of the 20 assessors performed audits only in the pre-experiment period or only during the experiment. We discuss the role of the assessor in explaining application quality in ex-post versus ex-ante audits in more detail below.

effects on the audit approval rate. We do not find differences in short-run and long-run effects. The effects on the quality of applications are immediate and quite persistent. This may be explained by the potential role of the assessors in the effect on application quality studied in subsection 5.3.2. The effects are no longer present after the experiment, which again suggests the absence of habit formation. The estimation results are shown in Table A.10 in the appendix.

Table 8: Baseline estimates for the audit approval rate at the application level (January 1, 2012 - April 7, 2013).

	approval			
	(1)	(2)	(3)	(4)
ex-ante audit regime	-0.039^{***}	-0.040^{***}	-0.040^{***}	-0.045^{***}
	(0.013)	(0.013)	(0.013)	(0.014)
conditional audit regime	-0.027^{**}	-0.025^{*}	-0.024^{*}	-0.019
	(0.013)	(0.013)	(0.013)	(0.015)
mean ex-post audit regime	0.827	0.827	0.827	0.827
(before experiment)	(0.014)	(0.014)	(0.014)	(0.014)
p-value joint significance	0.008	0.007	0.006	0.006
p-value ex-ante = conditional	0.363	0.246	0.246	0.068
care provider random effects	yes	yes	yes	yes
time fixed effects	yes	yes	yes	yes
care characteristics	no	yes	yes	yes
patient characteristics	no	no	yes	yes
assessor indicators	no	no	no	yes
# applications	16,876	16,875	16,872	16,872
# care providers	299	299	299	299

Note: Pre-experiment mean and standard error of the outcome in the ex-post audit regime are reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. F-test p-values are reported for testing whether the coefficient on the ex-ante audit regime and the conditional audit regime are simultaneously equal to zero. The general time trend is accounted for by two week period dummies. Care characteristics include indicators for type of re-assessment, indicators for all care types (personal care, nursing care, individual assistance, group assistance, treatment and facility-based care). Furthermore, indicators for various levels of intensity of personal care, and similarly for nursing care, individual and group assistance are included. Patient characteristics include a gender dummy, a marital status dummy, and an indicator for whether an individual is older than 80 years.

Providers in the conditional audit regime who had performance levels around the target approval rate of 0.92 before the experiment may be more likely to respond to the audit regime changes by changing application quality compared to providers far below or far above the threshold. Therefore, we expect effects on the audit approval rate to be heterogeneous with initial approval rate. We interact the time trend and the treatment indicators in equation (4) with a dummy variable for the initial approval rate category.[35] The estimation results in Table 9 show that care providers

[35]We estimate the panel data model $Q_{i,c,t} = \alpha + \gamma_{d_c,t} + \delta_{\text{ex-ante},d_c} T^{\text{ex-ante}}_{i,c,t} +$

with intermediate and, to a lesser extent, high initial approval rates reduce their number of applications in response to the conditional audit regime. Furthermore, only providers with low initial approval rates have a significant reduction in audit approval rate in response to the conditional audit regime.[36]

5.3.1 Approved applications

In the ex-ante and the conditional audit regime fewer applications are filed and approval rates are lower than in the ex-post regime. This implies that there are also fewer approved applications in these regimes where auditing can have consequences. The number of approved applications is the main indicator for public funding of long-term care. Therefore, we estimate equation 1 again, but with the normalized number of approved applications as outcome measure.

Table 10 presents the effects of the audit regimes on approved applications. The table shows that both the ex-ante and the conditional audit regime have a significantly lower number of approved applications. This is consistent with hypotheses 3 and 6 of the theoretical model. Recall that in the ex-post regime also disapproved applications receive funding for care provision. Combining these effects shows that in the ex-ante and the conditional audit regime funding is given to slightly over 60% of the funded applications in the ex-post regime. Stricter auditing thus has a very substantial impact on public expenditures on long-term care for three reasons. Fewer applications are filed, more applications are disapproved and disapproved applications are not funded.

$\delta_{\text{conditional},d_c} T_{i,c,t}^{\text{conditional}} + \beta X_{i,c,t} + \eta_c + \varepsilon_{i,c,t}$.

[36] There can be heterogeneity across types of patients for whom an application is filed. For some types of patients it is easier to make an accurate assessment of care needs and as a result to file an application that will be approved in an audit. We consider heterogeneous effects by age and gender and find no heterogeneity by gender. Approval rates decrease considerably more for the elderly, both in the ex-ante and the conditional audit regime. The estimation results are available on request.

Table 9: Heterogeneous effect estimates in approval rates.

Heterogeneity in provider size		**Heterogeneity in initial approval rates**	
	(1)		(2)
ex-ante × small	−0.080**	ex-ante × low	−0.039
	(0.041)		(0.017
ex-ante × medium	−0.018	ex-ante × intermediate	−0.040
	(0.022)		(0.021
ex-ante × large	−0.042**	ex-ante × high	−0.044
	(0.019)		(0.039
conditional × small	−0.050	conditional × low	−0.030
	(0.035)		(0.018
conditional × medium	−0.008	conditional × intermediate	−0.018
	(0.023)		(0.022
conditional × large	−0.029*	conditional × high	0.018
	(0.017)		(0.033
mean ex-post, small	0.801	mean ex-post, low	0.752
	(0.034)		(0.010
mean ex-post, medium	0.851	mean ex-post, intermediate	0.892
	(0.017)		(0.005
mean ex-post, large	0.822	mean ex-post, high	0.988
	(0.014)		(0.005
p-value, small	0.127	p-value, low	0.057
p-value medium	0.674	p-value, intermediate	0.170
p-value, large	0.058	p-value, high	0.334
p-value ex-ante = conditional, small	0.374	p-value ex-ante = conditional, low	0.422
p-value ex-ante = conditional, intermediate	0.463	p-value ex-ante = conditional, medium	0.198
p-value ex-ante = conditional, large	0.187	p-value ex-ante = conditional, high	0.100
controls	yes	controls	yes
observations	16,872	observations	16,872
# providers	299	# providers	299

Note: 34% (21%) of participating providers are categorized as filing few (many) re-assessment applications in the pre-experiment period, the remaining 45% of providers file an intermediate number of applications; 59% (22%) of participating providers are categorized as having low (intermediate) initial approval rates. Patient characteristics are included as controls. The models include dummies for small and large providers (medium-sized providers are the reference) or dummies for low and high initial approval rates (intermediate initial approval rates is the reference). Pre-experiment mean (over providers) and standard error of the outcome in the ex-post audit regime by provider type are reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Reported p-values are for F-tests testing the null hypothesis of no effect of any regime for a particular type of provider.

Table 10: Baseline estimates for the normalized number of approved applications (per week) at the provider level (January 1, 2012 - April 7, 2013).

	level
	(1)
ex-ante audit regime	−0.381**
	(0.155)
conditional audit regime	−0.409**
	(0.158)
mean ex-post audit regime	0.827
(before experiment)	(0.041)
care provider random effects	yes
p-value joint significance	0.030
p-value ex-ante = conditional	0.761
observations	9222
# care providers	297

Note: Pre-experiment mean and standard error of the outcome in the ex-post audit regime are reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. The p-value for the F-test is for joint significance of the coefficient on the ex-ante audit regime and the conditional audit regime.

5.3.2 The role of the assessor

The absence of possibilities for adjustments to the type and/or amount of care to be delivered after an ex-post audit may cause that assessors value ex-post audits differently than ex-ante audits. Even though assessors were explicitly instructed not to change anything in the way audits were done, differences in attitude towards ex-ante and ex-post audits can have consequences for their strictness. Which application is in fact audited, cannot be influenced by the assessor, but the strictness of auditing can. The empirical results obtained in the previous subsections are consistent with the hypotheses based on our theoretical model that assessors devote less effort to ex-post audits. Below we compare the results of ex-ante and ex-post audits and consider the assessors.

Table 11 shows descriptives on the reasons for disapproval in ex-ante audits and ex-post audits before and during the experiment, where multiple reasons may apply to a single disapproval. Comparing ex-ante audited applications before and during the experiment, the table illustrates only modest differences. More substantial differences are observed between ex-ante and ex-post audited applications during the

Table 11: Reasons for audit disapproval before and during the experiment.

	before	**during**	
	ex-ante audited	*ex-ante audited*	*ex-post audited*
number of registered reasons	1.761	1.668	0.897
different type of care required	0.261	0.274	0.261
too much care requested	0.328	0.344	0.376
too little care requested	0.048	0.077	0.014
disorder and/or limitations not appropriate	0.592	0.544	0.179
other reasons	0.357	0.283	0.261
# audited applications	1383	788	574

Note: Fractions do not sum to one as there may be multiple reasons underlying one disapproval. Ex-ante (ex-post) audited applications may be filed either by care providers in the ex-ante (ex-post) regime or care providers in the conditional regime. Obviously, ex-post audited applications are necessarily filed during the experiment.

experiment. For ex-post audited applications the average number of registered reasons for disapproval is considerably lower. In the ex-post regime assessors have fewer incentives to extensively motivate an audit decision and recall that ex-post audit decisions were administered in another system than ex-post audits. In particular, there is a substantial reduction in *too little care requested* and *disorder or limitations not appropriate* in the ex-post audit regime.[37] Recall that ex-post audits occur both in the ex-post audit regime and in the conditional audit regime. If ex-post audits are less rigorous, this can inflate approval rates in the ex-post regime and, to a lesser extent, the conditional audit regime.

Of 16 assessors auditing re-assessment applications during the experiment, only three performed both ex-post and ex-ante audits. We exploit this to investigate whether ex-post audits are less strict than ex-ante audits. For this purpose we consider the panel data model,

$$Q_{i,a,c,t} = \alpha_c + \gamma_t + \xi_a + \theta P_{i,a,c,t} + \eta_{i,a,c,t} \tag{5}$$

where $Q_{i,a,c,t}$ is the dummy for approval in an audit of application i filed by provider c at time t and audited by assessor a. We include provider fixed effects and a general time trend. Provider fixed effects are required because in the conditional audit regime, the choice for ex-ante or ex-post audit is not random. ξ_a is an assessor

[37]These shifts can be (partly) attributed to differences in data quality of registrations of disapproval reasons for ex-ante and ex-post audits and hence do not relate to behavioral responses of care providers. Because of this we do not present more formal analyses of the effect of the audit regime on disapproval reasons.

Table 12: Effect of ex-post auditing on approval probability (September 17, 2012 - April 7, 2013).

	approval	
	(1)	(2)
ex-post audit	−0.004	0.040
	(0.025)	(0.027)
constant	0.827***	0.856***
	(0.024)	(0.062)
provider fixed effects	yes	yes
time fixed effects	yes	yes
assessor fixed effects	no	yes
# audited applications	8358	8358
# care providers	287	287
# assessors	16	16

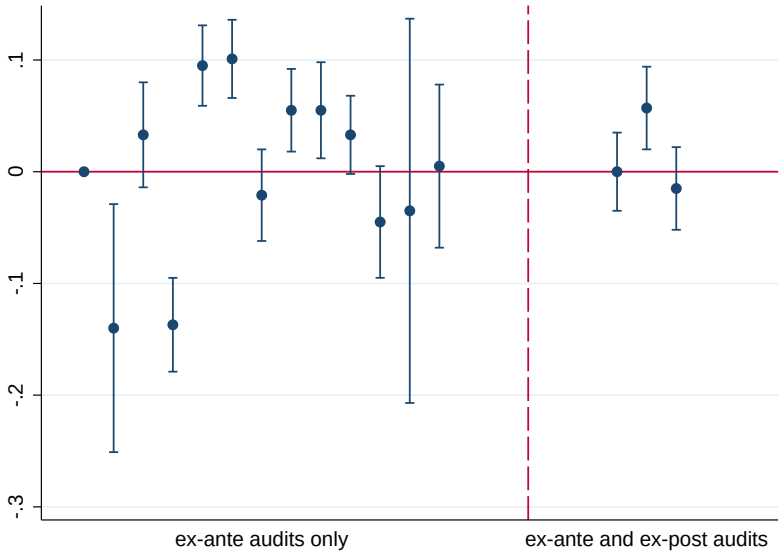
Note: Both models are estimated on audits during the experiment only. For 163 out of 8,521 audited applications (during the experiment) it is unknown which assessor performed the audit. These observations are excluded from the analyses. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

fixed effect accounting for differences in audit strictness between assessors. $P_{i,a,c,t}$ indicates whether application i filed by provider c was audited ex-post ($P_{i,a,c,t} = 1$) or ex-ante by assessor a. The effect of interest θ measures the change in the probability of approval in case of ex-post instead of ex-ante audit.

Table 12 shows the estimation results. In column (1) assessor fixed effects are excluded and estimation is done on data for all 16 assessors. This shows that the overall difference in approval probability between ex-ante and ex-post audited applications is very small. When adding the assessor fixed effects in column (2), we identify the effect on those assessors having performed both ex-ante and ex-post audits. Now, ex-post audits have about a four percentage point higher approval rate. This is consistent with our earlier finding that approval rates are highest in the ex-post audit regime. It also supports our theoretical model in which assessors devote less effort to auditing applications that do not have (direct) consequences. We should note that the estimated effect of ex-post audit is not significant.

To characterize the three assessors doing ex-post audits during the experiment,

Figure 7: Estimated assessor fixed effects.



we estimate the following model using only pre-experiment data.

$$Q_{i,a,c,t} = \alpha_c + \gamma_t + \xi_a + \eta_{i,a,c,t} \tag{6}$$

Figure 7 shows the resulting point estimates and 95%-confidence intervals for the assessor fixed effects ξ_a. The figure illustrates that, although there are differences in approval probabilities across assessors, the three assessors who performed ex-post audits, are not very different in terms of approval rates from other assessors.

Hence, we do not find strong evidence that the lower approval probability in the ex-ante and conditional audit regimes can be attributed to differences in the strictness of assessors.

6 Conclusion

Public resources for long-term care services can be managed by a gatekeeper deciding on the approval of applications for these services. However, often care is needed quickly, which limits the gatekeeper in its possibilities to audit applications. In this paper we use a large-scale field experiment to study the effectiveness of different audit policies in the Dutch market for long-term care. In this market care providers file applications on behalf of the patients. We consider three audit regimes that differ in the degree of enforcement and the presence of performance incentives: ex-ante auditing, ex-post auditing and conditional auditing. In the ex-ante audit regime care is only provided after approval of the gatekeeper, while in the ex-post audit regime care is provided immediately after the application. The gatekeeper then only

provides feedback on the quality of the application, but there are no consequences for the provision of care. In the conditional audit regime the timing of the audit depends on recent approval rates of the care provider. Those with high recent approval rates are audited ex-post, those with low recent approval rates are audited ex-ante. Thus, in this conditional audit regime enforcement (ex-ante auditing) is targeted towards poorly performing care providers. We provide an intuitive and simple theoretical model that provides testable hypotheses.

Our results show that increasing the degree of enforcement (the ex-ante compared to the ex-post audit regime) reduces the number of applications by 10%. A more substantial reduction of 20% is found for the conditional audit regime as compared to the ex-post regime. A substantial part of these declines in the number of applications can be attributed to substitution between types of applications. Furthermore, we find negative effects of both the ex-ante and the conditional audit regime on the quality of applications. We find some evidence that this can partly be explained by assessors valuing ex-post audits less than ex-ante audits. Low auditing effort of assessors when incentives are low, may also explain the finding byLindeboom et al. (2016) that varying the audit rate of ex-post audits does not affect the number of applications and the approval rate. The treatment effects on both the number of applications and the audit approval rate disappear shortly after the end of the experiment, so the temporary change in the audit regime did not cause any habit formation. More enforcement and incentives reduce the number of (approved) applications, which supports the traditional model by Allingham and Sandmo (1972) and Harrington (1988).

Long-term care expenses are rising across all developed countries. Trends towards lower fertility rates and increasing life expectancy imply that this is likely to remain one of the biggest challenges for health care systems. The main policy implication of our results is that audit rules can be an important tool in managing applications for long-term care services and targeting care to those who need it. We find that a gatekeeper is important when aiming to curb the trends towards expansion of long-term care expenditures. In particular, auditing is most efficient when there are incentives such as in the conditional audit regime (as compared to the unincentivized ex-post audit regime). The question whether the decrease in the number of funded applications improves targeting of resources to those who need the care most cannot be answered by our analysis. However, our theoretical model and the observed change in the approval rate suggest that when enforcement and incentives in auditing are less strong, invalid applications (i.e. type I classification errors) are more likely to receive funding. This conditional audit regime is also efficient in terms that it causes fewer delays in access to care services than the ex-ante regime. However, auditing efficiency might be somewhat limited by a shift towards

other application types. Nevertheless, these types are typically associated with less extensive care. The lesson for policy makers is thus that (some) care providers file applications to types of care where access is most easily obtained.

References

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3 - 4):323 – 338.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249 – 275.

Blundell, W. (2020). When threats become credible: A natural experiment of environmental enforcement from Florida. *Journal of Environmental Economics and Management*, 101:102288.

Bohnet, I., Frey, B. S., and Huck, S. (2001). More order with less law: On contract enforcement, trust, and crowding. *American Political Science Review*, 95(1):131 – 144.

Cason, T. N., Friesen, L., and Gangadharan, L. (2016). Regulatory performance of audit tournaments and compliance observability. *European Economic Review*, (85):288 – 306.

Cason, T. N. and Gangadharan, L. (2006). An experimental study of compliance and leverage in auditing and regulatory enforcement. *Economic Inquiry*, 44(2):352 – 366.

Clark, J., Friesen, L., and Muller, A. (2004). The good, the bad, and the regulator: An experimental test of two conditional audit schemes. *Economic Inquiry*, 42(1):69 – 87.

Colombo, F., Llena-Nozal, A., Mercier, J., and Tjadens, F. (2011). Help wanted? Providing and paying for long-term care. OECD Health Policy Studies, OECD Publishing.

Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611 – 1630.

Friesen, L. (2003). Targeting enforcement to improve compliance with environmental regulations. *Journal of Environmental Economics and Management*, 46(1):72 – 85.

Gilpatric, S. M., Vossler, C. A., and Liu, L. (2015). Using competition to stimulate regulatory compliance: A tournament-based dynamic targeting mechanism. *Journal of Economic Behavior and Organization*, 119:182 – 196.

Gilpatric, S. M., Vossler, C. A., and McKee, M. (2011). Regulatory enforcement with competitive endogenous audit mechanisms. *RAND Journal of Economics*, 42(2):292 – 312.

Greenberg, J. (1984). Avoiding tax avoidance: A (repeated) game-theoretic approach. *Journal of Economic Theory*, 32(1):1 – 13.

Harrington, W. (1988). Enforcement leverage when penalties are restricted. *Journal of Public Economics*, 37(1):29 – 53.

Hartman, L., Hesselius, P., and Johansson, P. (2013). Effects of eligibility screening in the sickness insurance: Evidence from a field experiment. *Labour Economics*, 20:48–56.

Landsberger, M. and Meilijson, I. (1982). Incentive generating state dependent penalty system. *Journal of Public Economics*, 19(3):333 – 352.

Lindeboom, M., Van der Klaauw, B., and Vriend, S. (2016). Audit rates and compliance: A field experiment in care provision. *Journal of Economic Behavior and Organization*, 131(B):160 – 173.

Mendoza, J. P. and Wielhouwer, J. L. (2015). Only the carrot, not the stick: Incorporating trust into the enforcement of regulation. *PLoS ONE*, 10(2):1 – 18.

Mot, E. (2010). The Dutch system of long-term care. CPB Document No. 204.

Nederlandse Zorgautoriteit (2013). Marktscan en beleidsbrief AWBZ: Weergave van de markt 2010 - 2012. Utrecht: NZa.

Thomson, S., Osborn, R., Squires, D., and Jun, M. (2013). International profiles of health care systems. The Commonwealth Fund.

Willemé, P. (2010). The long-term care system for the elderly in Belgium. ENERPI Research Report No. 70.

Appendices

A.1 Construction of the panel data set

Starting from application-level data containing the exact date of each application, a few choices are made to construct an aggregated panel data set. We want the start of the experiment at the beginning of a time period. This affects the observation period when we only want to use full time periods. Alternatively, we could weigh the numbers of applications and audits in the incomplete periods by the number of weeks that these periods cover. Comparing the results in Tables A.8 (weighted) and A.4 (exclusive) in the appendix, we find results to be robust with respect to this choice. Furthermore, the end of the experiment does not necessarily fall exactly at the beginning of a new time period when aggregating the data. Therefore, we drop the period containing the end of the experiment in the regression analysis.

In our main analysis, we aggregate the data in time periods of two weeks. Aggregating the data to weekly periods leads to many provider-period observations with zero re-assessment applications (24%), while this is limited to 12% of provider-period observations when aggregating the data for two week time periods. For the number of audits this is even more pronounced with zero audits in 56% (38%) of the provider-period observations when aggregating to weeks (two weeks). We checked the robustness of our estimated effects on the number of applications to this aggregation choice. Results can be found in Table A.4 in appendix A.3.

Related to this aggregation issue, a further robustness check is concerned with the issue of potentially serially correlated errors stressed by Bertrand et al. (2004). In our main analysis, we account for this by reporting cluster-robust standard errors. An alternative solution, mentioned by Bertrand et al. (2004), is to collapse the data into one pre-treatment and one post-treatment period and estimate a simple difference-in-differences specification on this collapsed data set. The estimation results are presented in Table A.9 in appendix A.3. Columns (1) and (2) show coefficients similar in magnitude to our baseline estimates reported in Table 5. We observe some differences in the magnitude of the estimated effects on the audit approval rate (columns (3) and (4)). Including a pre/during-experiment dummy instead of dummies for two week periods yields estimates that are comparable to our baseline results (Table 8 column (1)). However, random effects estimation on the pre/during-experiment aggregated data set yields larger coefficients and standard errors.

A second choice relates to the treatment of providers that were inactive at the beginning or at the end of the time range for which we have data available. We do not observe the exact date at which a provider became (in)active. We approximate this by comparing the provider-specific length of a period at the beginning/end

without applications to the length of intermediate periods of zero applications. If the length of a period of inactivity at the beginning (end) exceeds the length of the longest intermediate period of inactivity by more than two weeks, we define the provider as inactive at the beginning (end) of our data range. There are four ways to treat inactivity of these providers in the construction of the panel data set.

(a) Disregard inactivity by putting numbers of applications and audits to zero whenever no applications have been filed in a particular time period (balanced panel).

(b) Include providers as soon as an application is filed.

(c) Include providers from the time period $t + 1$ ($t - 1$) onwards if activity starts (ends) within time period t.

(d) Weigh the number of applications and audits by the number of weeks of activity in the first (last) time period of activity.

In our main analysis we include providers as soon as they filed an application, irrespective of whether or not this coincided with the start of a new time period (option (b)). We investigated the robustness of results to the alternative ways of dealing with inactivity. The estimation results for our baseline model are reported in Table A.5 in appendix A.3. Estimates are only modestly affected when using the balanced panel data set. Estimates obtained using procedures (c) and (d) yield coefficients similar to our baseline results reported in section 5. As a final robustness check, the table also shows the results of regressions on a data set from which providers inactive at the end of the data range have been completely removed. Note that for the effects on the audit approval rate, which are estimated on application-level data, this issue does not show up.

A.2 Dynamics in the conditional audit regime

The conditional audit regime successfully reduced the number of re-assessment applications. Furthermore, audit approval rates in the conditional regime are somewhat lower. Behavioral responses can be investigated in more detail by looking at dynamics between ex-ante and ex-post auditing of providers in the conditional audit regime. Recall that providers were audited ex-ante (ex-post) as long as their approval rate is less than (at least) 0.92. There were twelve occasions for switching the timing of audit. Descriptives on these switches are provided in Table A.1. Over time, a declining fraction of providers is subject to ex-ante auditing. This illustrates that, on average, care providers had performance levels above the 0.92-threshold. However, in each update we observe providers switching in both directions.

Table A.1: Updates of the timing of audit in the conditional audit regime.

update	**% ex-ante**	**% ex-ante → ex-post**	**% ex-post → ex-ante**
0 (17 Sep '12)	100.0	-	-
1 (15 Oct '12)	57.9	42.1	0.0
2 (5 Nov '12)	61.1	9.5	12.6
3 (19 Nov '12)	56.8	5.3	1.1
4 (3 Dec '12)	56.8	6.3	6.3
5 (17 Dec '12)	52.6	6.3	2.1
6 (31 Dec '12)	53.7	6.3	7.4
7 (14 Jan '13)	53.7	3.2	3.2
8 (28 Jan '13)	55.8	7.4	9.5
9 (11 Feb '13)	55.8	5.3	5.3
10 (25 Feb '13)	50.5	8.4	3.2
11 (11 Mar '13)	51.6	3.2	4.2
12 (25 Mar '13)	48.4	10.5	7.4

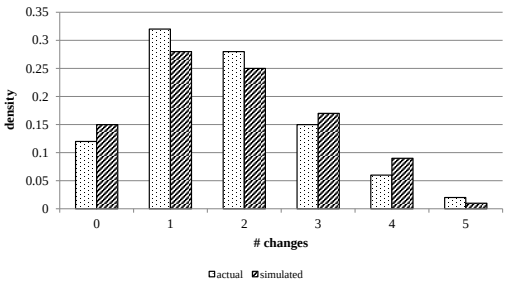
Note: In parentheses in the first column is the date at which the update became effective. Reported are percentages of providers in the conditional audit regime. Four care providers did not have any audited application during the experiment. For them the timing of audit remained at ex-ante without this being due to poor performance. We excluded these providers from the figures in this table.

The light-colored bars in Figure A.1a show the distribution of the actual number of switches for providers (over all twelve updates). Providers switched at most five times, but on average a provider has 1.8 switches during the experiment. Figure A.1b shows the distribution of the actual number of update rounds in which a care provider was subject to ex-ante auditing. The figure illustrates that only 11% of care providers performed consistently well; they moved to ex-post auditing immediately after the start of the experiment and remained being audited ex-post. On the other hand, 12% of care providers performed consistently poor: they have always been subject to ex-ante auditing.

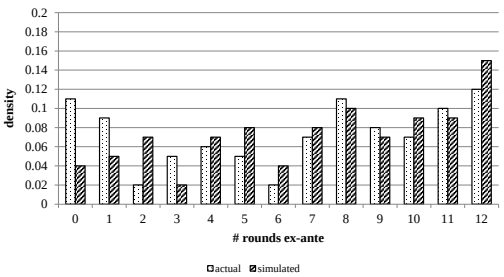
Switches between ex-ante and ex-post auditing in the conditional audit regime can be the result of sorting based on care providers' intrinsic motivation and/or ability to file applications accurately, or could reflect actual changes in performance as a result of the audit regime change. To determine whether the full variation in the timing of audit can be attributed to sorting, we consider simulated conditional updates for providers in the ex-ante audit regime. That is, we apply the rules used for updating the timing of audit in the conditional audit regime to providers in the ex-ante audit regime using application data on the experiment period. This exploits the absence of actual behavioral responses to performance incentives in the conditional regime by care providers in the ex-ante audit regime. As a result, simulated variation

Figure A.1: Actual and simulated dynamics in the timing of audit.

(a) Changes in the timing of audit



(b) Frequency of ex-ante auditing



in the timing of audit for these providers is the result of sorting only. Table A.2 shows statistics on the simulated updates for providers in the ex-ante audit regime. For each update we test for equality of the actual and simulated fraction of providers subject to ex-ante auditing and for equality of the actual and simulated fraction of providers switching in a particular direction by means of a χ^2-test and report p-values in the table. The results illustrate some differences for the first update, with fewer providers switching to ex-post auditing in the simulated updates. However, in subsequent updates we do not observe significant differences. Hence, the results suggest observed dynamics in the timing of audit in the conditional audit regime to be largely the result of sorting on pre-experiment approval rates.

In addition, Figure A.1a shows that the actual and simulated distributions of the number of switches are quite similar. This is confirmed by a formal χ^2-test for equality of the actual and simulated distributions (p-value = 0.872). Figure A.1b shows some differences in the actual and simulated distributions of the number of rounds of ex-ante auditing. The actual distribution shows a larger fraction of providers with only very few rounds (zero or one) of ex-ante auditing. Nevertheless, a formal χ^2-test cannot reject the null hypothesis of equality of the distributions (p-value = 0.553). We thus conclude that observed dynamics in the timing of audit in the conditional audit regime can largely be attributed to sorting based on unobserved intrinsic quality of providers.

Table A.2: Descriptives on simulated updates of the timing of audit in the ex-ante audit regime.

update	**% ex-ante**		**% ex-ante → ex-post**		**% ex-post → ex-ante**	
		p-value		*p-value*		*p-value*
0 (17 Sep '12)	100.0		-		-	
1 (15 Oct '12)	70.5	[0.069]	29.5	[0.069]	0.0	
2 (5 Nov '12)	65.3	[0.547]	12.6	[0.488]	7.4	[0.227]
3 (19 Nov '12)	65.3	[0.234]	7.4	[0.551]	7.4	[0.030]
4 (3 Dec '12)	67.4	[0.135]	6.3	[1.000]	8.4	[0.579]
5 (17 Dec '12)	66.3	[0.055]	4.2	[0.516]	3.2	[0.650]
6 (31 Dec '12)	64.2	[0.140]	5.3	[0.756]	3.2	[0.194]
7 (14 Jan '13)	56.8	[0.662]	11.6	[0.026]	4.2	[0.700]
8 (28 Jan '13)	54.7	[0.884]	9.5	[0.601]	7.4	[0.601]
9 (11 Feb '13)	52.6	[0.662]	9.5	[0.267]	7.4	[0.551]
10 (25 Feb '13)	51.6	[0.885]	7.4	[0.788]	6.3	[0.306]
11 (11 Mar '13)	54.7	[0.663]	2.1	[0.650]	5.3	[0.733]
12 (25 Mar '13)	51.6	[0.663]	8.4	[0.620]	5.3	[0.551]

Note: Reported are percentages of providers in the ex-ante audit regime. Six care providers did not have any audited application during the experiment. We excluded these providers from the figures in this table. In square brackets are χ^2-test p-values for equality of actual and simulated statistics.

A.3 Additional tables

Table A.3: Descriptive statistics on outcomes during the experiment (September 17, 2012 - April 7, 2013).

	ex-ante	**ex-post**	**conditional**
re-assessments			
applications (per week)	4.13	4.88	3.70
	(0.46)	(0.61)	(0.47)
audits (per week)	0.96	1.23	0.87
	(0.10)	(0.15)	(0.11)
approval rate	0.82	0.87	0.81
	(0.01)	(0.01)	(0.02)
other applications (per week)			
standard applications	6.62	7.22	8.75
	(1.48)	(1.64)	(1.87)
facility-based notifications	1.74	2.16	1.73
	(0.33)	(0.41)	(0.35)
emergency care applications	4.11	4.54	5.86
	(0.76)	(0.97)	(1.33)
regular applications	3.93	4.11	3.80
	(0.37)	(0.48)	(0.45)
# care providers	96	99	96

Note: Standard error of means in parentheses. Reported p-values are for Kruskal-Wallis rank tests for equality of populations.

Table A.4: Robustness check: level of time aggregation.

	# applications				
	(1)	(2)	(3)	(4)	(5)
ex-ante audit regime	−0.626	−0.652	−0.614	−0.534	−0.505
	(0.411)	(0.419)	(0.413)	(0.379)	(0.390)
conditional audit regime	−0.954**	−0.963**	−0.938**	−0.829**	−0.766**
	(0.398)	(0.404)	(0.399)	(0.366)	(0.376)
mean ex-post audit regime (before experiment)	3.557	3.551	3.557	3.554	3.552
	(0.362)	(0.362)	(0.362)	(0.362)	(0.362)
	# applications (normalized)				
	(1)	(2)	(3)	(4)	(5)
ex-ante audit regime	−0.104	−0.105	−0.103	−0.089	−0.080
	(0.079)	(0.080)	(0.079)	(0.070)	(0.074)
conditional audit regime	−0.203***	−0.200**	−0.198**	−0.173**	−0.162**
	(0.078)	(0.079)	(0.078)	(0.072)	(0.074)
mean ex-post audit regime (before experiment)	1.000	1.000	1.000	1.000	1.000
period length	2 weeks	1 week	4 weeks	8 weeks	12 weeks
observations	9286	19127	4650	2048	1466
# care providers	299	299	299	299	299
# time periods	32	66	16	7	5

Note: Column (1) repeats the baseline results. Pre-experiment mean of the outcome variable in the ex-post audit regime is reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table A.5: Robustness check: inactive care providers.

	# applications				
	baseline	*balanced*	*excluded*	*weighted*	*removed*
	(b)	(a)	(c)	(d)	
ex-ante audit regime	−0.626	−0.708*	−0.623	−0.622	−0.639
	(0.411)	(0.417)	(0.412)	(0.411)	(0.419)
conditional audit regime	−0.954**	−0.915**	−0.949**	−0.957**	−0.982**
	(0.398)	(0.403)	(0.398)	(0.398)	(0.403)
mean ex-post audit regime (before experiment)	3.552	3.557	3.559	3.558	3.514
	(0.362)	(0.362)	(0.362)	(0.362)	(0.360)
	# applications (normalized)				
	baseline	*balanced*	*excluded*	*weighted*	*removed*
	(b)	(a)	(c)	(d)	
ex-ante audit regime	−0.104	−0.171**	−0.102	−0.103	−0.102
	(0.079)	(0.085)	(0.079)	(0.079)	(0.080)
conditional audit regime	−0.203***	−0.211**	−0.196**	−0.208***	−0.204**
	(0.078)	(0.082)	(0.078)	(0.078)	(0.079)
mean ex-post audit regime (before experiment)	1.000	1.000	1.000	1.000	1.000
observations	9286	9568	9273	9286	8952
# care providers	299	299	299	299	281

Note: Column (1) repeats the baseline results, where a care provider is included in the data as soon as an application is filed. In column (2) we disregard inactivity and set the number of applications to zero when no application has been filed in a particular period, resulting in a balanced panel. Column (3) additionally, compared to the baseline, excludes periods in which care providers start (stop) to be active. In column (4) the number of applications in the first (last) period of activity is weighted by the number of weeks in which the provider was active. For details on these alternatives, see appendix A.1. Pre-experiment mean of the outcome variable in the ex-post audit regime is reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table A.6: Composition effects in the number of re-assessment applications at the application level.

	patient types		**re-assessment types**		
	women	*older than 80*	*type 1*	*type 2*	*type 3*
ex-ante audit regime	−0.006	0.001	−0.014	0.007	0.006
	(0.010)	(0.009)	(0.013)	(0.015)	(0.009)
conditional audit regime	−0.009	0.006	0.007	0.008	−0.014
	(0.009)	(0.007)	(0.012)	(0.016)	(0.011)
intercept	0.585***	0.281***	0.709***	0.118***	0.124***
	(0.018)	(0.020)	(0.020)	(0.011)	(0.013)
# applications	72,446	72,446	72,446	72,446	72,446
# care providers	299	299	299	299	299

Note: Type 1 re-assessment applications concern changes of home-based care, type 2 applications concern prolongation of facility-based care without changes, and type 3 re-assessments concern changes of facility-based care assessments. Each model includes care provider random effects. The general time trend is accounted for by including dummy variables for each period of two weeks. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table A.7: Estimates of a Hawthorne effect for the number of applications.

	# applications	
	level	*normalized*
	(1)	(2)
ex-ante audit regime	0.119	−0.014
	(0.135)	(0.099)
mean (weighted) non-participants, before	1.293	1.000
observations	9545	9545
# care providers	314	314

Note: The results reported are obtained from weighted regression, using weights assigned to non-participants to equalize their distribution of applications filed to those for the participants in the experiment. Time and provider fixed effects are included. Pre-experiment mean of the outcome for the non-participants is reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table A.8: Robustness check: including incomplete periods.

	# applications			
	(1)	(2)	(3)	(4)
ex-ante audit regime	−0.682	−0.721	−0.704	−0.813
	(0.429)	(0.449)	(0.450)	(0.502)
conditional audit regime	−0.976**	−0.981**	−1.000**	−0.996**
	(0.413)	(0.431)	(0.433)	(0.479)
mean ex-post audit regime (before experiment)	3.551	3.531	3.548	3.489
	(0.362)	(0.363)	(0.363)	(0.370)
period length	2 weeks	4 weeks	8 weeks	12 weeks
observations	9854	5218	2616	2034
# care providers	299	299	299	299
# time periods	34	18	9	7
	# applications (normalized)			
	(1)	(2)	(3)	(4)
ex-ante audit regime	−0.107	−0.107	−0.114	−0.131
	(0.083)	(0.089)	(0.089)	(0.116)
conditional audit regime	−0.196**	−0.184**	−0.204**	−0.192*
	(0.081)	(0.086)	(0.085)	(0.110)
mean ex-post audit regime (before experiment)	1.000	1.000	1.000	1.000
period length	2 weeks	4 weeks	8 weeks	12 weeks
observations	9854	5218	2616	2034
# care providers	299	299	299	299
# time periods	34	18	9	7

Note: Pre-experiment mean of the outcome variable in the ex-post audit regime is reported as a reference. Provider random effects and time fixed effects are included. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table A.9: Robustness check: accounting for serial correlation.

	# applications		**approval**	
	level	*normalized*	*application-level*	*aggregated*
	(1)	(2)	(3)	(4)
ex-ante audit regime	−0.656	−0.114	−0.042***	−0.051***
	(0.461)	(0.084)	(0.013)	(0.019)
conditional audit regime	−0.913**	−0.171**	−0.028**	−0.056**
	(0.443)	(0.082)	(0.013)	(0.024)
mean ex-post audit regime (before experiment)	3.545	1.000	0.829	0.829
	(0.362)	(0.000)	(0.014)	(0.014)
observations	590	590	16,876	586
# care providers	299	299	299	299

Note: The models include provider random effects, a treatment period indicator and interactions between treatment group and the treatment period indicator. As in the main analyses, the model in column (2) does not include provider random effects. No additional controls are included. Pre-experiment mean of the outcome variable in the ex-post audit regime is reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%.

Table A.10: Short-run, long-run and post-treatment effects on the audit approval probability (January 1, 2012 - November 24, 2013).

	approval
ex-ante regime, short-run	-0.048^{***}
	(0.017)
ex-ante regime, long-run	-0.034
	(0.016)
ex-ante regime, post-experiment	0.005
	(0.015)
conditional regime, short-run	-0.024
	(0.021)
conditional regime, long-run	-0.013
	(0.018)
conditional regime, post-experiment	0.017
	(0.014)
mean ex-post audit regime (before experiment)	0.827
	(0.014)
p-value, short-run	0.022
p-value, long-run	0.102
p-value, post-experiment	0.489
care provider random effects	yes
time fixed effects	yes
care characteristics	yes
patient characteristics	yes
assessor indicators	no
observations	24,170
# care providers	299

Note: Pre-experiment mean and standard error of the outcome in the ex-post audit regime are reported as a reference. Cluster-robust standard errors in parentheses, clustered by care provider. * significant at 10%, ** significant at 5% and *** significant at 1%. Reported p-values are for F-tests of the null hypothesis of no effect of any of the audit regimes in the short-run, long-run and after the experiment.