DISCUSSION PAPER SERIES

No. 10572

THE EFFECT OF AUDIT REGIMES ON APPLICATIONS FOR LONG-TERM CARE

Maarten Lindeboom, Bas van der Klaauw
and Sandra Vriend

PUBLIC ECONOMICS

Centre for Economic Policy Research

ISSN 0265-8003

THE EFFECT OF AUDIT REGIMES ON APPLICATIONS FOR LONG-TERM CARE

Maarten Lindeboom, Bas van der Klaauw and Sandra Vriend

Discussion Paper No. 10572
May 2015
Submitted 21 April 2015

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **PUBLIC ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Maarten Lindeboom, Bas van der Klaauw and Sandra Vriend

THE EFFECT OF AUDIT REGIMES ON APPLICATIONS FOR LONG-TERM CARE[†]

Abstract

This paper studies the effects of various audit regimes, differing in the degree of control 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 the degree of control reduces the number of applications and that introducing performance incentives reduces this even further. However, we find evidence for substitution with other types of long-term care services, which are often less extensive. 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.

JEL Classification: C93, H51 and I18
Keywords: auditing, field experiment, incentives and long-term care

Maarten Lindeboom mlindeboom@vu.nl
VU University Amsterdam and Tinbergen Institute

Bas van der Klaauw b.vander.klaauw@vu.nl
VU University Amsterdam, Tinbergen Institute and CEPR

Sandra Vriend s.vriend@vu.nl
VU University Amsterdam and Tinbergen Institute

[†] Earlier versions of the paper circulated with the title *Control-based or trust-based auditing: A field experiment in 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 conference and seminar participants in Aarhus, Alicante, Amsterdam, Barcelona, Braga, Ljubljana and Rotterdam for useful comments.

1 Introduction

In the provision of long-term care a trade-off has to be made between providing services quickly when needed and efficient spending of public resources. Many countries introduced a so-called gatekeeper, which is an institution responsible for controlling access to long-term care services. The gatekeeper often seeks for effective audit policies to manage the provision of long-term care.[1] The strictness of an audit regime is determined by several instruments, such as audit frequency, degree of control, (financial) sanctions, and performance incentives. The choices of these instruments are important determinants of the speed at which an application can be approved and care can be provided.

Our study concerns the Netherlands, where long-term care involves chronic care, either institutionalized or home-based, for the elderly and the mentally or physically impaired. When a patient requires long-term care services, a care provider files an application for (public) funding to the gatekeeper. The gatekeeper can either approve the application or perform an audit, which can result in an approval, denial or adjustment of the application.[2] Care providers do not start the care provision until approval is given, causing that audits delay the moment at which the patient receives services. This is referred to as *ex-ante* auditing.

In this paper, we conduct a large-scale field experiment to compare the ex-ante auditing regime to two alternative audit regimes. The first is an *ex-post* audit regime, which allows that care provision starts immediately after filing the application. Unlike after an ex-ante audit, the requested services cannot be denied or adjusted in an ex-post audit. The degree of control of the gatekeeper is thus smaller when applying ex-post auditing compared to ex-ante auditing. Both types of auditing do not allow for (financial) sanctions in case of disapproval. Prior to our experiment ex-post auditing was already used for cheaper and less extensive types of long-term care (e.g., Lindeboom et al., 2014). The reason for this field experiment is that the gatekeeper believes that many care providers are intrinsically motivated to file

[1] Long-term care systems differ considerably across countries, not only in terms of the types of care covered, but also in terms of the systems implemented to manage expenditures. In particular, differences exist in terms of whether access decisions are made at the local (municipal or regional) or at the national level (Colombo et al., 2011). Secondly, in many countries, care providers make the assessment of needs, without an independent institution checking care needs prior to provision of services, whereas in other countries a separate institution makes need assessments. For instance, in Belgium, care providers make the assessment and this is combined with frequent ex-post random evaluations of the assessments (Willemé, 2010). 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] 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 have been requested by the care provider.

applications correctly and in accordance with the actual care needs of patients. In their opinion ex-post auditing can be equally effective as ex-ante auditing, while ensuring immediate access to care services.

In addition to the ex-ante and the ex-post audit regime, we introduce a *conditional* audit regime. In this regime, the timing of audits is adjusted regularly during the experiment based on recent approval rates. Care providers that perform well in terms 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 control, but control is targeted towards poorly performing providers. This regime yields fewer delays in care provision than the ex-ante audit regime, but introduces incentives for care providers to comply. Care providers are randomized in one of the three audit regimes. We study the effects of the audit regime on application behavior of care providers, in particular the quantity and quality of applications.

The theoretical literature on auditing started 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. In contrast to the traditional model, behavioral models focus on the role of intrinsic motivation and trust. 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. Mendoza and Wielhouwer (2015) show in a game-theoretical model that trust-based regulation is feasible if agents care sufficiently about the future and if there are some costs of control to the agent. Falk and Kosfeld (2006) find in a laboratory experiment that most agents reduce performance when the principal chooses to control. This suggests that control crowds out intrinsic motivation of agents, labeled the hidden costs of control. With only weak incentives available to the principal, trusting agents provides more favorable outcomes (Falk and Kosfeld, 2006). Although the studies above concern individuals, similar studies have been done for (non-profit) organizations. Bengtsson and Engström (2014) conduct a field experiment where they switch from a fully trust-based regime to one with increased monitoring for some non-profit organizations, but they do not find evidence for a crowding out effect.

Our study also relates to the literature on external interventions such as sanctions or rewards, or in our case a delay in provision, possibly crowding out intrinsic motivation. In multiple settings a crowding-out effect due to the presence of external (financial) interventions has been found. For instance, Frey and Oberholzer-Gee (1997) find that the willingness of individuals to agree with unwanted projects in their neighborhood decreases when compensation is offered. Gneezy and Rustichini (2000b) find that, even though higher monetary rewards increase performance, the

introduction of financial incentives in itself may have detrimental effects on performance in IQ-tests and on doing volunteer work. Leuven et al. (2010) find a similar result when studying financial rewards for academic achievement. Finally, Gneezy and Rustichini (2000a) conclude that the introduction of a fine increases the number of parents picking up their children late from day care.[3]

These branches of the literature help us in defining hypotheses on the effects of the audit regime of care provider application behavior. Following the lines of thought of the traditional model, care providers should file fewer (invalid) applications of higher quality under the ex-ante and conditional audit regimes as compared to the ex-post audit regime. On the contrary, in our setting, intrinsic motivation can be interpreted as the willingness of providers to file applications according to the rules of the gatekeeper. This would lead us to hypothesize compliance to be highest in the ex-post audit regime, translating into fewer (invalid) applications and higher approval rates under the ex-post audit regime as compared to the ex-ante and conditional audit regimes.

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. To a large extent, this seems to be the result of substitution to other types of long-term care applications. Furthermore, we find an (insignificant) 10% 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. We provide some evidence that part of this effect can be attributed to behavior of assessors who screen ex-post audits less accurately.

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.

[3]Reviews of this literature are provided by, for example, Fehr and Gächter (2000), Frey and Jegen (2001) and Gneezy et al. (2011). Various mechanisms are suggested for this effect (listed by, e.g., Falk and Kosfeld, 2006): agents may feel insulted (e.g., Gneezy and Rustichini, 2000b), the usage of incentives may provide new information on the importance or costs of the activity (e.g., Gneezy and Rustichini, 2000a; Bénabou and Tirole, 2003), or incentives may be in contradiction with social norms of fairness, trust or cooperation (e.g., Fehr and List, 2004; Sliwka, 2007).

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).[4] This covers chronic care, either institutionalized 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 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, s/he should contact his/her preferred care provider. The care provider files a request to the gatekeeper who then makes an assessment of the needs, stating amongst others the type and amount of care services the patient is eligible for.[5] When the application is approved by the gatekeeper, financial resources are made available so that the provider can start delivering the care to the patient. Patients are allowed to switch providers during care provision, which gives care providers an incentive to provide good-quality care.[6] In 2012, there were around 1,600 providers of which almost 1,400 provided extramural and 800 provided intramural care. Intramural care providers are required to be non-profit institutions, while extramural care providers can also be for-profit organizations. Care providers differ considerably in size, some are large-scale nursing homes, whereas other providers consist of a very small number of care givers who visit patients at their homes.

A provider can choose between various types of applications, depending on the long-term care history of the patient and the care needs. We focus on one type of application, the so-called re-assessment applications.[7] Re-assessment applications are used to prolong the provision of existing care services, possibly adjusted in, for instance, intensity.[8] It is not possible to use a re-assessment application in case of

[4]The description of the institutional details in this section is similar to Lindeboom et al. (2014) and draws on Mot (2010), Nederlandse Zorgautoriteit (2013) and information from the Center for Care Assessment (CIZ).

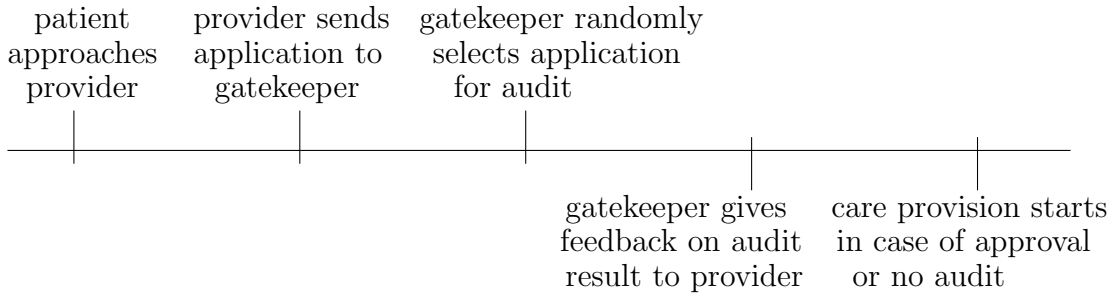
[5]A patient can file a request directly to the gatekeeper, but around 85% of the applications are filed by care providers.

[6]In the period from July 2010 to July 2011, about 2.5% of those receiving extramural care through a re-assessment application switched provider, while this was 8% for intramural care (Nederlandse Zorgautoriteit, 2013).

[7]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.

[8]Types of care services are: personal care (e.g., help with showering), nursing (e.g., wound

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



a first or a complex application. In 2012, around 70,000 re-assessment applications were filed. Re-assessment applications usually result in care provision for a relatively long period of up to 15 years.[9] Hence, the associated expenses are substantial and often much larger than for other types of applications.

A random sample of around 25% of the re-assessment applications are 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. An audit may result in an approval or disapproval. A disapproval can either be an adjustment or denial of the application, meaning that requested and adjudged care services can differ. Several reasons for disapproval are registered. 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 on the audit from the assessor, explaining what went wrong and how the quality of future applications can be improved.

When a re-assessment application is selected for audit, the financial resources for providing care are not yet made available. This only happens after the result of the audit is known. Therefore, the care provider has to wait with care provision until approval from the gatekeeper. The audit result should be made available within two weeks (on average, it is known within a week).

Besides the re-assessment applications that are the main focus of our field experiment, providers can choice 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

care), assistance (help in organizing practical matters in daily life), treatment (e.g., rehabilitation) and stay in an institution (e.g., nursing home).

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

that providers can file are '*emergency care applications*', '*standard applications*', '*intramural 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. Intramural notifications concern intramural care for patients older than 80 years. 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.

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 adjust or deny applications. In particular, no (financial) sanctions can be applied in case of disapproval. Ex-post auditing is thus less strict than ex-ante auditing. 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.[10] 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.[11] These updates were communicated

[10]The threshold audit approval rate of 0.92 is determined by the target approval rate that the gatekeeper has in mind.

[11]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,

to the care providers supported by a list of results of the audits. All care providers in all treatment groups received the same detailed feedback about the outcomes of their own audits.

In each of the treatment groups, the probability that an application is selected for audit is held constant. Assessors were instructed to audit along the same procedures 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.[12] Participation was compulsory. Care providers were informed by letter and e-mail about the set-up of the experiment.[13] Finally, a brief notice of the experiment was posted in an online information bulletin of the gatekeeper in October 2012.

Table 1 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.

The gatekeeper assigned three assessors to perform the ex-post audits. These assessors also perform ex-ante audits, but they could distinguish whether the audit is ex-ante or ex-post. Table 2 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 results are not likely to be driven by differences between assessors. However, approval rates of ex-post audits are substantially higher then of ex-ante audits. We return to the role of the assessors in more detail in subsection 5.3.1.

and the second adjustment three weeks after the first update.

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

[13]An indication for 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 1: 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	

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

3.3 Hypotheses

A patient who contacts a long-term care provider usually dislikes waiting until the provision of care starts. Furthermore, since long-term care is publicly funded, both the patient and the care provider benefit from increasing the amount of care. However, to obtain funding for providing long-term care, providers should devote costly effort to file an application. The gatekeeper randomly selects a number of applications for audit and may disapprove the application if the requested care does not match the patient's needs or if information in the application is incomplete. Requesting too much care or not devoting enough effort to the application increases the probability of a disapproval during an audit.

According to the traditional model of Allingham and Sandmo (1972), care providers should weigh the benefits of requesting funding for more care and devoting more effort to filing the application against the expected costs of noncompliance given the degree of control and the level of punishment. This traditional model implies that sufficient control and extensive punishment by the gatekeeper can induce all care providers to comply. This means that care providers file complete applications, do

Table 2: 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	-

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).

not request too much care and do not make invalid applications. In our setting, the gatekeeper can only choose between ex-ante and ex-post auditing. Since ex-post auditing does not provide any (financial) incentive for compliance, a prediction of the traditional model is that under ex-ante (and also conditional) auditing care providers should file fewer applications and have a higher approval rate.

This traditional model does not yield a clear prediction about how the conditional audit regime relates to ex-ante auditing. On the one hand, in the conditional audit regime care providers may alter periods of ex-ante auditing with periods of ex-post auditing. While being exposed to ex-post auditing the care providers may have a lower degree of compliance since the audits do not have immediate consequences. On the other hand, the result of an audit in the conditional audit regime has consequences for how future audits will be performed. So it depends on the discount rate of the care providers, the strictness of audits and the costs of effort when filing an application, how behavior of care providers changes when being exposed to the conditional audit regime.

In contrast to this traditional model, behavioral theories focus on the role of intrinsic motivation. In our setting intrinsic motivation can be interpreted as the willingness of care providers to file applications according to the rules of the gatekeeper. It is often argued that external interventions, like control and punishment, can crowd out intrinsic motivation (see e.g., Gneezy et al. (2011) and other references in the introduction). Increasing the degree of control as is done in ex-ante auditing compared to ex-post auditing may reduce effort put in filing applications and can reduce compliance. This may also be the case for the conditional audit regime. If intrinsic motivation is the driving factor, the prediction is that compli-

ance is highest in the ex-post audit regime. And if disapprovals are associated with invalid applications, then the number of applications is lowest in the ex-post audit regime.

Trust is often mentioned as a pathway through which external interventions crowd out intrinsic motivation. Mendoza and Wielhouwer (2015) mention two requirements for feasible trust-based auditing, i.e. non-decreasing compliance despite less control. They argue that trust-based auditing can be feasible when there are certain costs of control to agents and when agents sufficiently value the future. In our setting, control-based auditing is associated with a delay in care provision. Since care providers are competing for patients, they care about quick provision of services to patients. This delay implies a cost to the care provider. Regarding the second requirement, care providers typically have repeated interactions with the gatekeeper. Building a trust relation can, therefore, be valuable for providers. Hence, trust-based auditing can be feasible in our setting so that detrimental effects on compliance need not be expected.

Care providers are likely to be heterogeneous, some providers may behave according to the traditional model (referred to as traditional care providers in the remainder), while others may have a stronger intrinsic motivation. In case of such a mix of providers, the conditional audit regime would probably yield the largest compliance. Care providers that are intrinsically motivated may interpret this as a regime where trust can be built when applications are approved and dissolves after a disapproval. Once they become exposed to ex-post auditing, they may feel trusted and maintain a high level of compliance. For the traditional care providers this audit regime imposes the incentives to perform well, because there are no direct punishments for noncompliance and they can provide care quicker when their performance remains sufficiently high to be exposed to ex-post auditing. The crucial feature of the conditional audit regime is the performance threshold for switching between ex-ante and ex-post auditing. If this threshold is set too high, it is too costly for the traditional care providers to comply and the intrinsically motivated care providers may get stuck in ex-ante auditing.

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 5.2.2 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.[14]

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 for this audit decision. Data on applications provided by the gatekeeper are complemented with linked micro-data from Statistics Netherlands, providing information on basic patient characteristics including age, gender and marital status.[15]

Descriptive statistics on application and patient characteristics for whom re-assessment applications are filed, are provided in Table 3. 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 extramural care. A single application may be concerned with multiple types of care. The most frequent type of care is personal care, followed by assistance and institutionalized 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 percent 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, by treatment group are provided in Table A.3 in the appendix.

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. The resulting panel data set is unbalanced, because providers may become inactive due to closing down activities or merging with another provider.[16] For ease of interpretation we normalize the number of applications to weekly averages for each care provider.

[14]Only in subsection 5.2.3 we also consider applications of care providers not participating in the experiment.

[15]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.

[16]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.

Table 3: Application characteristics and patient characteristics (January 1, 2012 - April 7, 2013).

	ex-ante	**ex-post**	**conditional**
Types of re-assessments			
change in extramural care	69%	73%	75%
prolong (unchanged) intramural care	16%	13%	11%
change in intramural 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%
institutionalized care	31%	27%	25%
Patient characteristics			
female	57%	58%	59%
age (in years)	62.26	64.30	66.16
older than 80	31%	33%	36%
married	22%	23%	24%
# applications	23,550	26,859	22,037

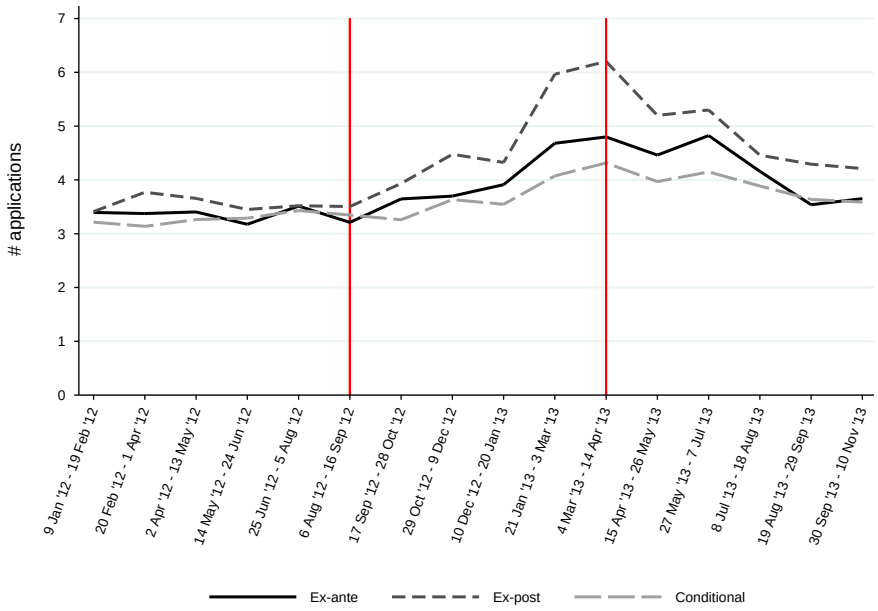
5 Results

Our key interests are the quantity and quality of applications, the latter being proxied by the audit approval rate. 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 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 group and smallest in the conditional audit group. After the experiment ended, the number of applications in the treatment groups converged again.

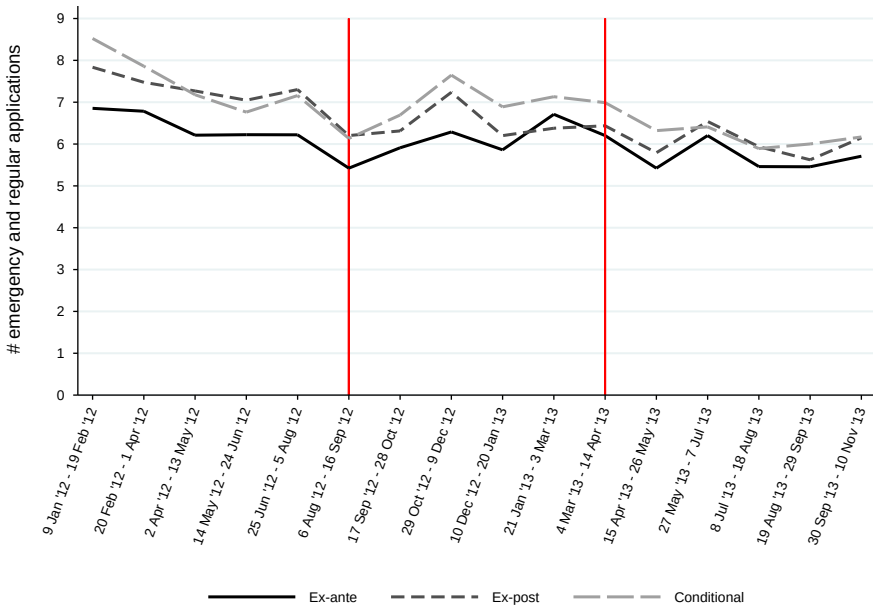
Figure 2: Weekly average number of re-assessment applications in the three treatment groups.



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

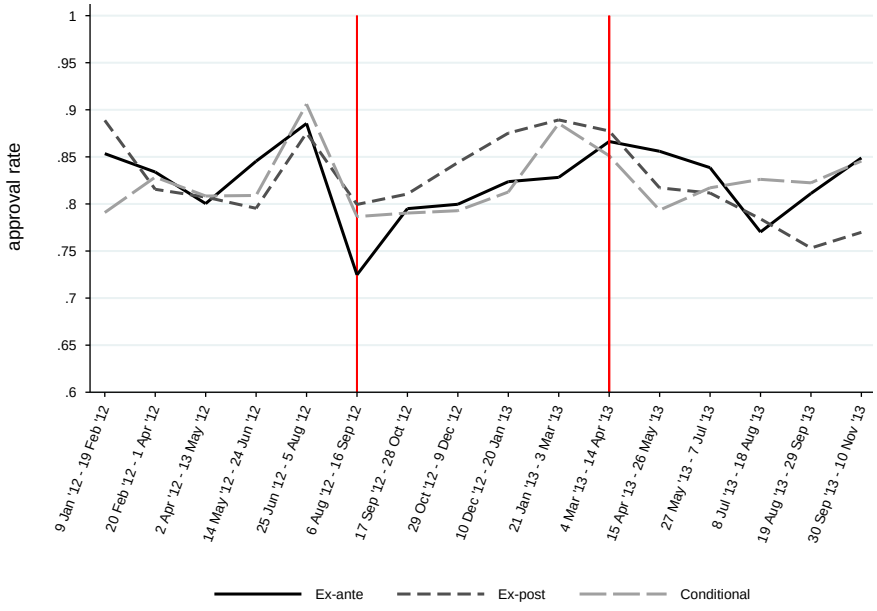
The change in the number of re-assessment applications might be explained by substitution with other application types. If substitution is present, it is most likely towards emergency care applications and regular applications. Other types of applications are very different in content from re-assessment applications. Figure 3 illustrates the trend in these two application types. The figure shows a decreasing pattern in the number of regular and emergency care applications before the experiment. We observe a change during the experiment, where the number of regular and emergency care applications slightly increases for providers in the ex-ante and conditional audit regimes, and remains relatively constant in the ex-post audit regime. This suggests that there is indeed substitution.

Figure 3: Weekly average number of emergency care and regular 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.

Figure 4: 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.

Our final outcome of interest is the audit approval rate. Figure 4 shows approval rates over time by treatment group. Despite the increase in the number of applica-

tions in the ex-post audit group, there are only modest differences across treatment groups with the approval rate being slightly higher in the ex-post audit group.

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,

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

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.[17] Care provider fixed effects (α_c) are included to control for (time-invariant) characteristics of care providers. The treatment variables $T_{c,t}$ equal one if care provider c belongs to a treatment group in period t. 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 in subsection 3.3 are in comparison to this regime. The ex-ante regime increases control for all care providers irrespective of performance. The conditional audit regime increases control only for poorly performing care providers, but introduces an incentive to perform well. 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. This normalization eliminates provider fixed effects.

The baseline estimation results in Table 4 show a substantial, but insignificant, decline of 0.624 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 0.924 applications per week and significant. When focussing on the fractional change, we see a 10% decline in the ex-ante audit regime and a 20% decline in the conditional audit regime compared to the ex-post audit regime, but only the latter is significant. The larger effect of the conditional audit regime can be explained by larger opportunity costs of false 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.[18] An alternative explanation, as discussed in the hypotheses

[17] 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.

[18] In section A.2 in the appendix we discuss the switches of providers between ex-ante and ex-

Table 4: Baseline estimates for the number of applications (per week) at the provider level (January 1, 2012 - April 7, 2013).

	# applications	
	level	*normalized*
	(1)	(2)
ex-ante audit regime	−0.624	−0.104
	(0.407)	(0.079)
conditional audit regime	−0.949***	−0.203***
	(0.394)	(0.078)
mean ex-post audit regime (before experiment)	3.557	1.000
	(0.362)	(0.000)
care provider fixed effects	yes	no
observations	9286	9286
# care providers	299	299

Notes: 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%.

in subsection 3.3, may be the presence of a mix of traditional care providers and intrinsically motivated care providers, where both respond to the conditional audit regime but only the traditional care providers respond to the ex-ante audit regime.

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.[19] 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 a dummy variable for the size of a provider.[20] The estimation results in Table A.6 in the appendix show that large providers are most responsive to the audit regime.

post auditing in the conditional audit regime. There we find that sorting of providers, based on pre-experiment approval rates, can explain the switches that we observe.

[19]We also considered heterogeneity in effects for providers differing in terms of the share of re-assessment applications in total applications. We found some evidence for providers having a small share of re-assessment applications responding more strongly to the ex-ante and the conditional audit regimes. Results are available on request.

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

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.

5.2.1 Substitution effects

We find that the number of re-assessment applications in the ex-ante and conditional audit regimes declines compared to the ex-post audit regime. An explanation for this can be substitution towards other application types. The content of standard applications and intramural notifications is substantially different from the typical re-assessment application, so if substitution is present, it is most likely towards emergency care applications and regular applications. Emergency care applications are audited less strict than re-assessment applications, while regular applications have a stricter audit regime.

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, the number of regular applications and the total number of applications other than re-assessment applications.[21] Estimation results are reported in Table 5. We find positive effects on both the number of regular applications as well as the number of emergency care applications. Most of the effects are quite large but insignificant. The substitution effects are largest for the conditional audit regime, where we also found the largest decline in the number of re-assessment applications. From column (1) we observe that these substitution effects, especially for the conditional regime, have about the same magnitude as the negative effect on re-assessment applications reported in column (1) of Table 4, i.e., there seems to be almost one-to-one substitution although again the effects are insignificant. This suggests that if auditing becomes more costly to the provider, providers choose the type of application more carefully. Stricter auditing hence seems to be less effective when only partially applied. The results are in agreement with the presence of traditional care providers searching for opportunities to minimize the effort of making applications.

Some re-assessment applications may be easier to substitute with emergency care and regular applications, which can cause compositional changes of the re-assessment applications. But also the audit regime itself may affect the composition

[21] 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. We consider this as a single application.

Table 5: Substitution effects at the provider level.

	all types	**regular**	**emergency**
ex-ante audit regime	0.382	0.304	0.126
	(0.531)	(0.214)	(0.150)
conditional audit regime	0.740	0.301	0.463*
	(0.960)	(0.241)	(0.247)
mean ex-post audit regime (before experiment)	16.973	4.485	2.701
	(2.768)	(0.579)	(0.572)
observations	9286	9286	9286
# care providers	299	299	299

Notes: All effects are in levels. Each regression includes provider and time fixed effects. Pre-experiment mean and standard error of the outcome in the ex-post 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%.

of re-assessment applications. However, we find no compositional changes induced by the audit regimes, as shown in Table A.7 in the appendix. In particular, there is no shift in patient characteristics (age and gender) and types of re-assessment applications.[22]

5.2.2 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 showed 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),

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

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

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 fixed 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,

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

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 θ 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 6 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.

Table 6: Short-run, long-run and post-treatment effects (January 1, 2012 - November 24, 2013).

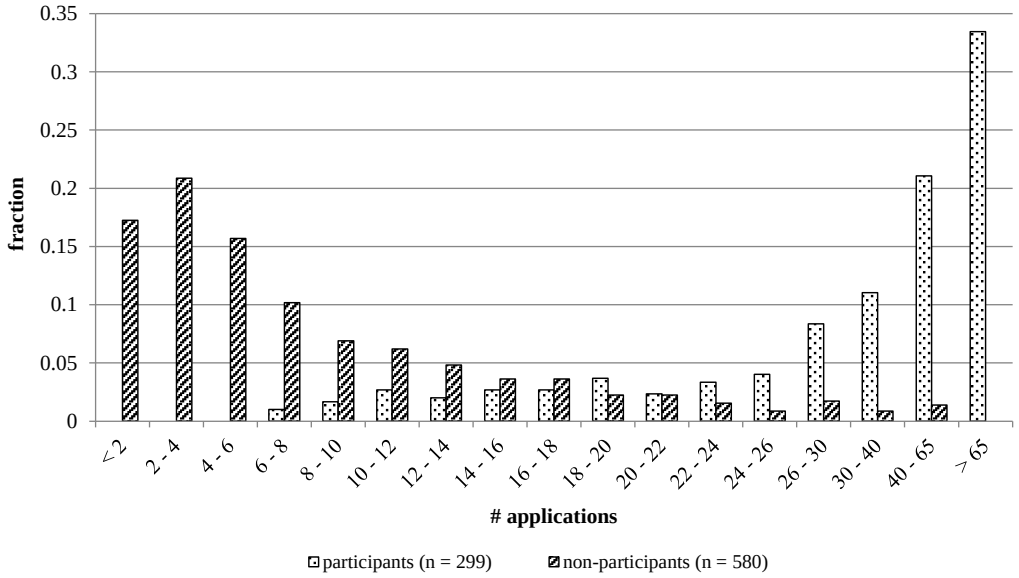
	# applications	
	level	*normalized*
	(1)	(2)
ex-ante regime, short-run	−0.319	−0.053
	(0.287)	(0.064)
ex-ante regime, long-run	−1.095*	−0.222*
	(0.619)	(0.134)
ex-ante regime, post-experiment	−0.487	−0.037
	(0.361)	(0.112)
conditional regime, short-run	−0.491*	−0.090
	(0.277)	(0.065)
conditional regime, long-run	−1.571***	−0.354***
	(0.605)	(0.133)
conditional regime, post-experiment	−0.658	0.013
	(0.436)	(0.114)
mean ex-post audit regime (before experiment)	3.557	1.000
	(0.362)	(0.000)
p-value, short-run	0.165	0.382
p-value, long-run	0.026	0.027
p-value, post-experiment	0.285	0.916
care provider fixed effects	yes	yes
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

Notes: 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.

5.2.3 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

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

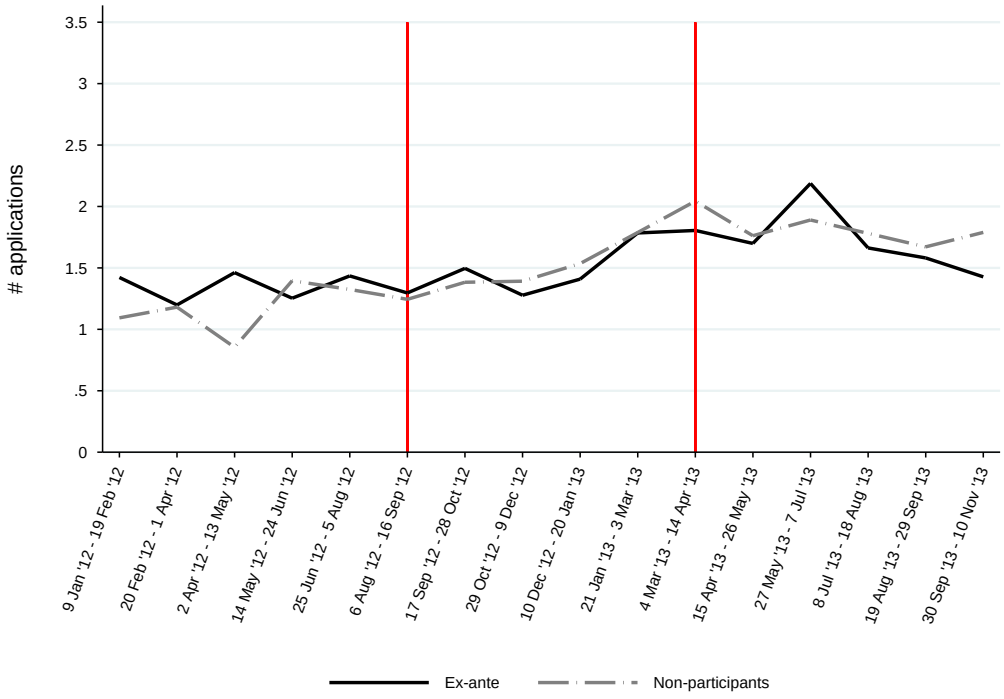


started increasing. This might be a Hawthorne effect, implying that participation in the experiment already affects outcomes. A possible explanation might be that informing providers about the experimental variation in auditing regimes for re-assessment applications draws more attention to the possibility 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.

We select among both groups the care providers filing between 6 and 45 applications, so the range in which there is an overlapping support.[23] 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. 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

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

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



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.[24] 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 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,

$$Q_{i,c,t} = \alpha_c + \gamma_t + \delta_{\text{ex-ante}} T_{i,c,t}^{\text{ex-ante}} + \delta_{\text{conditional}} T_{i,c,t}^{\text{conditional}} + \beta X_{i,c,t} + \eta_{i,c,t} \qquad (4)$$

[24]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 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.8 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.

Table 7: 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	−0.038**	−0.039**	−0.039**	−0.047***
	(0.015)	(0.016)	(0.016)	(0.017)
conditional audit regime	−0.026*	−0.028*	−0.028*	−0.029*
	(0.015)	(0.015)	(0.015)	(0.016)
intercept	0.884***	0.956***	0.956***	0.924***
	(0.023)	(0.026)	(0.026)	(0.043)
care provider fixed 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

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

$Q_{i,c,t}$ is a dummy for approval of application i filed by provider c at time period t. Again we include provider fixed effects α_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. Again, the ex-post audit regime, which is the least strict, is the reference group.

The estimation results are presented in Table 7. 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 three percentage points decrease for switching to the conditional audit regime.[25] The estimates for

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

the treatment effects are robust against including covariates. Only including the assessor indicators in column (4) causes the effects to increase 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 below. The direction of the effects is consistent with the presence of intrinsically motivated care providers, i.e. increasing the degree of control crowds out intrinsic motivation thereby reducing compliance.

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.[26] The estimation results in Table A.6 in the appendix show that care providers 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.[27]

5.3.1 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. Table 8 shows descriptives on the reasons for disapproval in ex-ante audits and ex-post audits before and during the experiment, where multiple reasons can be given for a single

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. 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.11 in the appendix.

[26] We estimate the panel data model $Q_{i,c,t} = \alpha_c + \gamma_{d_c,t} + \delta_{\text{ex-ante},d_c} T^{\text{ex-ante}}_{i,c,t} + \delta_{\text{conditional},d_c} T^{\text{conditional}}_{i,c,t} + \beta X_{i,c,t} + \varepsilon_{i,c,t}$.

[27] There can also be heterogeneity across types of patients for whom an application is filed. For some types of patients it may be 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 considered heterogeneous effects by age and gender and found 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 8: Reasons for audit disapproval before and during the experiment.

	before	**during**	
	ex-ante audited	*ex-ante audited*	*ex-post audited*
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

Notes: 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.

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 experiment. For ex-post audited applications often only a single reason for disapproval is stated. Furthermore, we see a substantial reduction in *too little care requested* and *disorder or limitations not appropriate*. 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 the 16 assessors auditing re-assessment applications during the experiment, only three performed both ex-post and ex-ante audits. We can 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 again include provider fixed effects and a general time trend. ξ_a is an assessor 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 an ex-ante audit.

Table 9 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

Table 9: Effect of ex-post auditing on approval probability.

	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

Notes: For 163 out of 8,521 audited applications (during the experiment) it is unknown which auditor 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%.

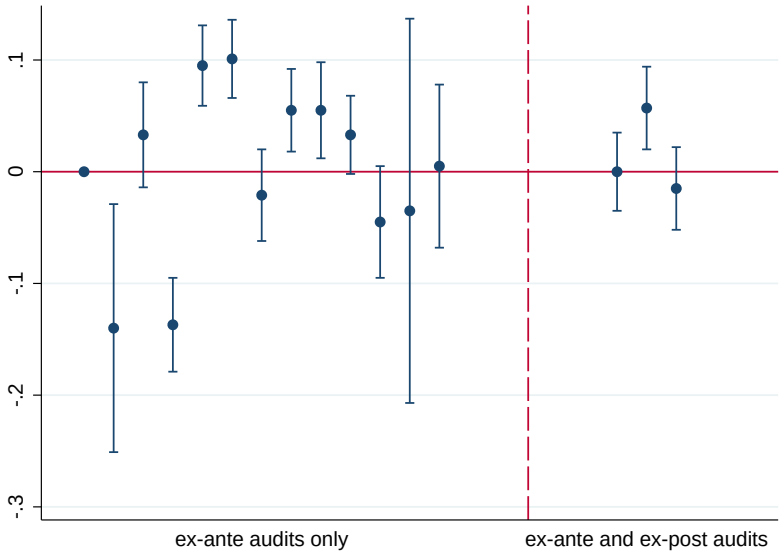
overall difference in approval probability between ex-ante and ex-post audited applications is very small. When adding the assessor fixed effects in column (2), we identify the effect on those assessors having performed both ex-ante and ex-post audits. Now, ex-post audits have about a four percentage point higher approval rate, which is consistent with our earlier finding that approval rates are highest in the ex-post audit regime. The effect on the audit approval rate may be partly attributed to a different attitude of assessors towards ex-ante and ex-post audits. Nevertheless, the estimated effect of ex-post audit is insignificant, which prevents us from drawing strong conclusions on this.

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

Figure 7: Estimated assessor fixed effects.



Hence, we do not find 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, which decides about rewarding applications for these services. However, often care is needed quickly, which limits the gatekeeper in its possibilities to audit applications. In this paper we discussed a large-scale field experiment to study the effectiveness of different audit policies in the Dutch market for long-term care. These audit policies differ in the degree of control, which is operationalized by the timing of audit and the presence of performance incentives. In particular, the timing of an audit can be ex-ante (before the start of care provision) or ex-post (after the start of care provision). With ex-ante auditing the gatekeeper makes the final decision on the care services that can be provided to the patient. The gatekeeper then has a large amount of control, but this comes at the cost of delay in provision. Ex-post auditing allows for immediate provision of care services, but comes at the risk of inefficiency in spending.

Our results show that increasing the degree of control (the ex-ante compared to the ex-post audit regime) reduces the number of applications by 10%. This effect is large in magnitude, but insignificant. A more substantial and significant reduction of 20% is found for the conditional audit regime as compared to the ex-post regime.

In this conditional audit regime control (ex-ante auditing) is targeted towards poorly performing care providers, which yields as a performance incentive that control is low (ex-post auditing) when the performance of a care provider is sufficiently good. A large part of these declines in the number of applications can be attributed to substitution between types of applications. Furthermore, we found negative effects of both the ex-ante and the conditional audit regime on the quality of applications. We found (weak) evidence that this can partly be explained by assessors valuing ex-post audits less than ex-ante audits. The estimated effects on both quantity and quality of applications disappear shortly after the end of the experiment, so the temporary change in the audit regime did not cause any habit formation.

More control and incentives reduce the number of applications, which supports the traditional model by Allingham and Sandmo (1972). But the results for the audit approval rate suggest that care providers are (often) intrinsically motivated to file applications correctly. However, differences in the attitude of assessors towards ex-ante and ex-post audits provide an alternative explanation. When only focussing on long-term care expenditures the conditional audit regime should be preferred because of the reduction in the number of applications and the shift towards other application types, which are typically associated with less extensive care. Furthermore, in the conditional audit regime the gatekeeper causes fewer delays in care provision due to auditing. Nevertheless, the choice of audit policy should ultimately also take into account targeting efficiency of funding, in particular whether the decline in the number of applications does not come at the cost of the availability of care services to patients truly in need of those services.

References

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

Bénabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70(3):489 – 520.

Bengtsson, N. and Engström, P. (2014). Replacing trust with control: A field test of motivation crowd out theory. *Economic Journal*, 124(577):833 – 858.

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.

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.

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.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159 – 181.

Fehr, E. and List, J. A. (2004). The hidden costs and returns of incentives: Trust and trustworthiness among CEOs. *Journal of the European Economic Association*, 2(5):743 – 771.

Frey, B. S. and Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, 15(5):589 – 611.

Frey, B. S. and Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review*, 87(4):746 – 755.

Gneezy, U., Meier, S., and Rey-Biel, S. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191 – 210.

Gneezy, U. and Rustichini, A. (2000a). A fine is a price. *Journal of Legal Studies*, 29(1):1 – 17.

Gneezy, U. and Rustichini, A. (2000b). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115(3):791 – 810.

Leuven, E., Oosterbeek, H., and Van der Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6):1243 – 1265.

Lindeboom, M., Van der Klaauw, B., and Vriend, S. (2014). Audit rates and compliance: A field experiment in long-term care. Tinbergen Institute Discussion Paper 2014-038/V.

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):e0117212.

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.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, 97(3):999 – 1012.

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.9 (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 analyses, 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.10 in appendix A.3. Columns (1) and (2) show coefficients similar in magnitude to our baseline estimates reported in Table 4. 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 7 column (1)). However, fixed 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 analyses 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.

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

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.

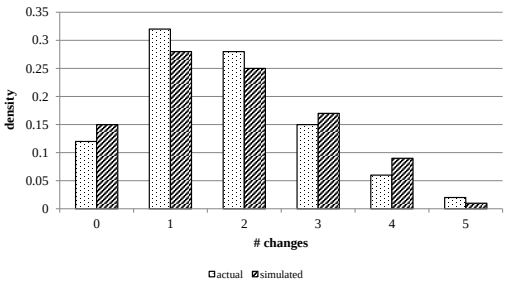
A.2 Dynamics in the conditional audit regime

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.

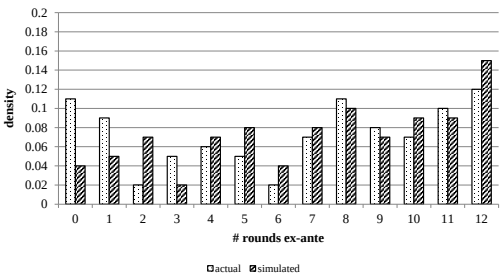
The light-colored bars in Figure A.1a show the distribution of the actual number of switches for providers (over all twelve updates). Providers switched at most five times, but on average a provider has 1.8 switches during the experiment. Figure A.1b shows the distribution of the actual number of update 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

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



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 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-

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]

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.

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.

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)
intramural 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

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	−0.624	−0.651	−0.611	−0.528	−0.496
	(0.407)	(0.416)	(0.407)	(0.369)	(0.379)
conditional audit regime	−0.949**	−0.960**	−0.931**	−0.818**	−0.751**
	(0.394)	(0.402)	(0.394)	(0.357)	(0.367)
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

Notes: 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	−0.624	−0.703*	−0.622	−0.620	−0.639
	(0.407)	(0.414)	(0.408)	(0.407)	(0.415)
conditional audit regime	−0.949**	−0.908**	−0.944**	−0.953**	−0.979**
	(0.394)	(0.400)	(0.395)	(0.394)	(0.400)
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

Notes: 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.6: Heterogeneous effect estimates.

Panel A: heterogeneity in provider size			**Panel B: heterogeneity in initial approval rates**		
	# applications (norm.)	***approval***		***# applications (norm.)***	***approval***
	(1)	(2)		(3)	(4)
ex-ante × small	−0.148	−0.071	ex-ante × low	−0.024	−0.045***
	(0.169)	(0.047)		(0.091)	(0.017)
ex-ante × medium	0.046	−0.010	ex-ante × intermediate	−0.302**	−0.047*
	(0.091)	(0.031)		(0.149)	(0.026)
ex-ante × large	−0.379**	−0.044**	ex-ante × high	−0.126	−0.076*
	(0.157)	(0.022)		(0.246)	(0.044)
conditional × small	−0.280*	−0.059	conditional × low	−0.087	−0.043**
	(0.153)	(0.044)		(0.095)	(0.020)
conditional × medium	0.002	0.000	conditional × intermediate	−0.419***	−0.022
	(0.104)	(0.033)		(0.159)	(0.026)
conditional × large	−0.368**	−0.035*	conditional × high	−0.341*	0.027
	(0.150)	(0.018)		(0.203)	(0.042)
mean ex-post, small	1.000	0.801	mean ex-post, low	1.000	0.752
	(0.000)	(0.034)		(0.000)	(0.016)
mean ex-post, medium	1.000	0.851	mean ex-post, intermediate	1.000	0.892
	(0.000)	(0.017)		(0.000)	(0.005)
mean ex-post, large	1.000	0.822	mean ex-post, high	1.000	0.988
	(0.000)	(0.014)		(0.000)	(0.005)
p-value, small	0.165	0.243	p-value, low	0.610	0.015
p-value medium	0.849	0.897	p-value, intermediate	0.027	0.203
p-value, large	0.042	0.069	p-value, high	0.178	0.056
controls	no	yes	controls	no	yes
observations	9286	16,872	observations	9286	16,872
# providers	299	299	# providers	299	299

Notes: 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. In columns (2) and (4) care and patient characteristics are included as controls. 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 A.7: 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	0.006	0.006
	(0.010)	(0.009)	(0.013)	(0.015)	(0.009)
conditional audit regime	−0.011	0.006	0.007	0.009	−0.014
	(0.009)	(0.007)	(0.011)	(0.016)	(0.011)
intercept	0.598***	0.302***	0.754***	0.102***	0.101***
	(0.015)	(0.013)	(0.011)	(0.008)	(0.010)
# applications	72,446	72,446	72,446	72,446	72,446
# care providers	299	299	299	299	299

Notes: Type 1 re-assessment applications concern changes of extramural care, type 2 applications concern prolongation of intramural care without changes, and type 3 re-assessments concern changes of intramural care assessments. Each model includes care provider fixed 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.8: 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

Notes: 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.9: Robustness check: including incomplete periods.

	# applications			
	(1)	(2)	(3)	(4)
ex-ante audit regime	−0.681	−0.720	−0.700	−0.812*
	(0.425)	(0.444)	(0.441)	(0.491)
conditional audit regime	−0.971**	−0.973**	−0.989**	−0.973**
	(0.409)	(0.426)	(0.425)	(0.469)
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	−0.106	−0.106	−0.113	−0.129
	(0.083)	(0.089)	(0.089)	(0.116)
conditional audit regime	−0.196**	−0.184**	−0.204**	−0.190*
	(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

Notes: 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: Robustness check: accounting for serial correlation.

	# applications		**approval**	
	level	*normalized*	*application-level*	*aggregated*
	(1)	(2)	(3)	(4)
ex-ante audit regime	−0.631	−0.114	−0.042***	−0.058**
	(0.424)	(0.084)	(0.015)	(0.024)
conditional audit regime	−0.868**	−0.171**	−0.027*	−0.064**
	(0.409)	(0.082)	(0.015)	(0.028)
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

Notes: The models include provider fixed 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 fixed 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.11: 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^{***}
	(0.018)
ex-ante regime, long-run	-0.028
	(0.019)
ex-ante regime, post-experiment	0.012
	(0.015)
conditional regime, short-run	-0.033
	(0.020)
conditional regime, long-run	-0.021
	(0.019)
conditional regime, post-experiment	0.005
	(0.015)
mean ex-post audit regime (before experiment)	0.827
	(0.014)
p-value, short-run	0.023
p-value, long-run	0.309
p-value, post-experiment	0.741
care provider fixed effects	yes
time fixed effects	yes
care characteristics	yes
patient characteristics	yes
assessor indicators	no
observations	24,170
# care providers	299

Notes: 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.