Health Services as Credence Goods: A Field Experiment*

Felix Gottschalk† Wanda Mimra‡ Christian Waibel§

December 2017

Abstract

Agency problems are a defining characteristic of health care markets. We present the results from a field experiment in the market for dental care: A test patient who does not need treatment is sent to 180 dentists to receive treatment recommendations. We vary the level of information and socio-economic status of the patient and collect measures of market conditions. We observe an overtreatment recommendation rate of 28% and a striking heterogeneity in treatment recommendations. Excess capacities, measured by waiting for the next possible appointment, are associated with significantly more overtreatment. Results furthermore suggest a complex role of patient socio-economic status.

JEL classification: *C93; D82; I11; L00.*

Keywords: *Health Care; Credence Goods; Overtreatment; Market Environment*

*First and foremost we would like to thank the medical experts who collaborated with us in this project. Furthermore, we thank Loukas Balafoutas, Sheheryar Banuri, Stefan Boes, Stefano Brusoni, Winand Emons, Stefan Felder, Pierre Fleckinger, Matteo Galizzi, Hans Gersbach, Ben Greiner, Jeanne Hagenbach, Glenn Harrison, Rudolf Kerschbamer, Tobias Klein, Albert Ma, Fabrizio Mazzonna, Marc Möller, Pierre Picard, Armin Schmutzler, Matthias Sutter, Florian von Wangenheim, Daniel Wiesen, and Peter Zweifel for helpful discussions, comments and support. The authors are grateful for comments by participants at the economics seminars at Tilburg, Innbruck, Mines ParisTech, Augsburg, Paderborn, the BEH workshop (Cologne), the IHPC (London), the DGGÖ (Basel), the Swiss IO Day (Bern), and the Swiss Health Economics Workshop (Zurich). We would further like to thank Julia Burri, Kevin Grab, and Angela Peter for their excellent research assistance. Financial support by the MTEC Foundation is gratefully acknowledged.

†ETH Zurich, Zürichbergstrasse 18, 8092 Zürich, Switzerland, Email: `fgottschalk@ethz.ch`.

‡Corresponding Author. ETH Zurich, Zürichbergstrasse 18, 8092 Zürich, Switzerland, Email: `wmimra@ethz.ch`, Phone: +41 44 632 98 26.

§ETH Zurich, Zürichbergstrasse 18, 8092 Zürich, Switzerland, Email: `cwaibel@ethz.ch`.

1 Introduction

Information problems are ubiquitous in health care markets. In the patient-physician relationship, physicians often have an informational advantage vis-à-vis their patients: Whereas patients neither know exactly which disease they suffer from nor which health care service best treats their health problem, physicians are able to diagnose patients and judge which treatment is appropriate.[1] Patients thus have to rely on the physician to recommend and provide the appropriate treatment. Often, even after receiving treatment, patients cannot judge whether the treatment was appropriate or whether, for instance, a less invasive or less expensive treatment might as well have solved their health problem. Therefore, health care services are considered to be credence goods.[2]

More generally, it is difficult to assess quality objectively in health care markets. This questions whether standard market incentive systems such as competition lead to efficient outcomes. For instance, depending on financial incentives, a physician might exploit his informational advantage by overtreating. Overtreatment cannot be detected, and it is a priori not clear whether more competition in the form of a higher physician density helps to alleviate this problem. On the contrary, a higher physician density, if it leads to lower demand at the individual physician, may be compensated by inducing demand, for instance by overtreatment.[3] In theory, price competition could lead to equal mark-up prices across treatments such that there are neither under- nor overtreatment incentives, however equal mark-ups across *all* treatments is an appealing theoretical but not empirically implementable concept.

Even though there is some empirical evidence indicating that physicians react to financial incentives in treatment decisions, this evidence is typically indirect and based on highly aggregated data such as administrative data from hospitals.[4] This has the major drawback that the physician-patient interaction, including patient demand effects, communication and the actual asymmetry of information, cannot be controlled for. Then, although it might be concluded that physicians react to financial incentives, it cannot be concluded that the provision of health care services was actually not appropriate and/or inefficient due to supply

[1] Physicians also typically have better information on other dimensions than appropriateness of treatment, such as costs, compensation and their own skills.

[2] The categorization of goods and services into exhibiting search, experience and credence properties was proposed by Darby and Karni (1973). Goods for which the credence property is important are usually called credence goods, although the same good may fall into all three categories, depending on the dimensions considered. See Dulleck and Kerschbamer (2006) for an overview on the theoretical credence goods literature and Kerschbamer and Sutter (2017) for an overview on credence goods lab and field experiments.

[3] For an early discussion of the physician-induced demand hypothesis, see Evans (1974).

[4] See Clemens and Gottlieb (2014) for an overview.

side behavior.

In this paper, we address this fundamental problem and provide direct evidence of physicians' treatment decisions and their determinants with a focus on the role of the market environment as well as patient characteristics and information. To do so, we conducted a natural field experiment in the Swiss dental market based on individual physician visits[5]. In our study, a single test patient visits 180 randomly selected dentists for a checkup. At each dentist visit, the test patient asked for a diagnosis—based on an examination and the same x-ray photograph—, a treatment recommendation and a cost estimate. The test patient had a superficial caries lesion limited to the enamel which should not be treated with an invasive treatment—such as a filling[6]—according to the Swiss Dental Guidelines and four cooperating reference dentists. Thus, we focus our analysis with the case at hand at the credence goods problem of overtreatment which wastes resources and may spur adverse health effects in the long run.[7]

To analyze how market conditions affect overtreatment, we collected data from several sources for the following three measures: First, the waiting time for the next possible appointment at a dentist. This measure nicely captures short-term demand relative to available capacity at an individual dentist and indicates whether the dentist has free capacity available. Second, we have a measure of the competitive behavior of an individual physician, the physician's price level[8]. Third, we add a standard measure of competition used in the literature on physician-induced demand, physician density. The analysis is complemented by data on dentist and practice characteristics.

Furthermore, we apply a 2x2 design for our experimental variations: First, we vary the information that the patient signals to the dentist. In particular, the patient either goes as a standard patient (who simply asks for a diagnosis) or the patient informs the dentist that he has, out of curiosity, uploaded his x-ray the day before to an internet dentist platform which provides information and diagnoses, and that he is curious about what both recommendations will be. Thus, the patient signals that he will likely receive another diagnosis from an internet

[5]See Harrison and List (2004) for a categorization of field experiments and Levitt and List (2009) for an overview on field experiments. List (2007) provides a comprehensive overview of the advantages of field over lab experiments.

[6]The dental care literature usually refers to "restorations" instead of "fillings".

[7]In our case, an invasive treatment such as a filling can lead to a higher subsequent caries risk. Another fundamental problem in credence goods markets is undertreatment, which can occur for several reasons, e.g., when liability does not hold. Although this is a serious problem, we cannot address undertreatment in a similar field experiment due to ethical considerations as it would imply denying a necessary treatment during the time of the study.

[8]Dentists in Switzerland can decide, to some extent, on their overall price level. A detailed description will be provided in Section 3.

platform.[9] For simplicity of terminology, we will refer to the patient as either the *standard* patient or the *informed* test patient and the experimental conditions[10] as ST vs. INFO. The second variation is whether the patient is perceived as a patient with a lower or a relatively higher socio-economic status (LS vs. HS). This variation is implemented by modifying the physical appearance of the patient in terms of clothing and accessories.

Our study has two main contributions: First, we investigate physicians' provision of health care services on the level of individual patient-physician interactions. The design allows us to observe for each physician whether she/he provides the appropriate treatment recommendation or an overtreatment recommendation instead of observing only *aggregate* provision rates. Thus, we can provide direct evidence of overtreatment and thereby have a clean and simple measure of physician quality. Our micro approach allows us to not only observe the overtreatment behavior but also to control for the covariates on the individual level. With the experimental variation, we furthermore provide results on the role of patient SES and signalled information on overtreatment. Second, we provide results highlighting the role of the market environment and physicians' characteristics on the level of overtreatment. In particular, with waiting time for the next possible appointment—capturing demand relative to capacity—we propose a simple measure for the relevant market condition affecting short-term treatment decisions. Understanding both which are the relevant measures and how market conditions affect physicians' treatment decision under asymmetric information is an important prerequisite for market design and regulation in these markets and thus far empirical evidence is scarce.

Our central result is an overtreatment recommendation rate of 28% (50/180). Conditional on an overtreatment recommendation, mean overtreatment costs taken from the collected cost estimates amount to CHF 535 (about $550), the median being lower at CHF 444 (about $455). Regarding the treatment—the test patient has a superficial interproximal caries lesion that should not be treated by an invasive treatment such as a filling—, the suggested number of fillings at a dentist ranges from 1 to 6. Furthermore, we observe that 13 different teeth to be treated with a filling appear across all cost estimates.[11] Thus, besides our finding of a considerable overtreatment recommendation rate, we also observe a striking heterogeneity in the treatment recommendations.

[9]The patient clearly communicates that he has not yet received information from the platform. We chose this design in order not to anchor the dentist's treatment recommendation on an already received diagnosis, but to signal that the patient is likely to receive another diagnosis.

[10]Note that we refer to experimental *conditions* instead of *treatments* to distinguish between dentists' treatments and experimental conditions.

[11]This constitues a lower bound, as in several cost estimates it is not indicated which particular tooth–each tooth has a number—will be treated.

Our results on market conditions are fully consistent with a model in which unused capacity is a key driver of short-term treatment decisions. We find that a lower waiting time for the next possible appointment is associated with a significantly higher level of overtreatment recommendations. Every additional day of waiting time reduces the likelihood of receiving an overtreatment recommendation by more than a percentage point. This result suggests that physicians with free capacity are more likely to provide treatments that are not necessary. We furthermore find that the practice price level is not significant. This might seem surprising at first, however it is fully in line with a capacity-based argument: As long as the profit margin is positive and there is free capacity—which holds for the treatment case—, the treatment is performed, independent of the size of the margin. We also find that physician density does not have a significant influence on overtreatment recommendations, even if we do not control for the waiting time. This result is important in light of a vast body of empirical work that approaches the physician-induced demand hypothesis by relating treatment volumes per capita to physician density based on aggregate data.[12] The logic of the density argument is based on the notion that, in a given location with a given demand, a higher density leads to less 'true' demand and unused capacity at physicians, who compensate by inducing demand. In this study, we provide a more direct measure of demand relative to capacity, and highlight why physician density is not a good proxy for it and therefore of limited use for the causal analysis of physician-induced demand.

With regard to dentist and practice characteristics, we observe an interesting result for price transparency: There are significantly less overtreatment recommendations at physician visits for which the price level was clearly displayed in the practice. Displaying the price level in the practice is required by regulation, however 60% of practices in our sample failed to do so[13]. One interpretation of this result is that there might be different types of physicians: Those who abide by regulation and (treatment) guidelines, and those who are less prone to do so.

Regarding our experimental treatment variations, we observe the counterintuitive result of significantly less overtreatment recommendations when the test patient is a high- rather than a low socio-economic status (SES) patient, given that the patient is a standard patient. When the patient is informed, the differences diminish. These results suggest a complex role of patients' SES as well as interactions with signalled information. With respect to

[12] See, for instance, the survey and discussion in Léonard et al. (2009).

[13] At a first glance, the rate of physicians not displaying the price level in their practice may seem high but is in line with other studies. For instance the Swiss magazine K-Tipp reported in 2004 that 27 out of 35 sampled dentists in large Swiss cities did not display the price level (K-Tipp No. 13, 8/2004). The consumer organization FRC sampled 170 dentists in the French speaking part of Switzerland in 2005 and found an adherence rate of 59% (reported in Schweizer Monatsschrift für Zahnmedizin Vol 115, 10/2005).

the information variation itself, we do not find a significant change in the overtreatment recommendation rate. We observe a reduction in the overtreatment rate for a patient with lower SES when going from standard to more information, however, this difference is not significant. It is generally assumed that information provision and diagnostics from the internet increase patient information and quality of care. Our results show the limits of this argument when the case at hand has credence goods characteristics and is complex, as is the case for most health care services.

The advantages of a field experiment in health care to study the impact of patient and market characteristics on overtreatment are manifold: First, employing a test patient ensures that there is no variation of patient characteristics besides the two condition. Different patient characteristics in real patients lead to different treatment behavior and might hence confound identification (see e. g. Adler and Rehkopf, 2008). Second, a field experiment allows us to objectively know the patient's illness and the appropriate treatment. Hence, we can unambigously identify whether an individual physician recommends an overtreatment or an appropriate treatment and thus derive a simple measure of physician quality. A third methodological advantage is that the Hawthorne and experimenter demand effects Zizzo (2010) do not play a role, because physicians do not know that their decisions are recorded. Naturally, there are—mostly methodologically inherent—shortcomings of our field experiment. We discuss these in detail in light of our results in the discussion section 5.3.

The remainder of the paper is as follows: In the next section, we discuss the related literature. Section 3 provides a description of the market for dental care in Switzerland. In section 4, we present the field study and describe our data. Results are presented and discussed in section 5. Section 6 concludes.

2 Related literature

There are several strands of literature relating to our field experiment: we first present a brief overview of the empirical literature on physician-induced demand (PID) and physician behavior as well as recent audit studies in health care. We then turn to field experiments and empirical results on other credence goods markets.

Physician-induced demand and financial incentives

Early empirical studies find that an increase in the number of surgeons is associated with an increase in the number of surgeries (Fuchs, 1978; Cromwell and Mitchell, 1986). Grytten et al. (1990) show a similar effect for dentists. Grytten and Sørensen (2001), however, find that a higher physician density is not correlated with more treatments, regardless of the remuneration method. In a meta-study, Léonard et al. (2009) provide an overview and comparison of a vast literature on PID in which many studies approach the topic by analyzing the correlation between physician density and a measure of health care utilization such as annual number of procedures per general practitioner. The idea behind is that a higher density implies lower demand per physician, which is compensated by physician-induced demand. This literature generally finds a significant association between physician density and health care consumption. In our study, we provide evidence that physician density may not be a good proxy to identify (short-term) market incentives and use an alternative measure to capture short-term demand relative to capacity: the waiting time for the next possible appointment.

Gruber and Owings (1996) report that the decline in fertility rates in the US in the 1970s was partly compensated by a substitution from normal childbirth to the more profitable cesarean delivery. Using data from the Japanese prescription drug market, Iizuka (2007) finds that the mark-up on drugs influences prescription decisions, yet physicians are also found to be willing to forgo profits in order to reduce costs for patients. Clemens and Gottlieb (2014) analyze area specific price shocks following a Medicare consolidation reform and find that areas with higher payment shocks experience significant increases in health care supply. The empirical literature thus indicates that physicians react to financial incentives in their medical decisions. However, the importance and extent of physician-induced demand as well as its determinants are still an open question.

Audit studies in health care

A small number of recent audit studies uses data from direct observation of a physician-patient interaction as in our study. In Currie et al. (2011) and Currie et al. (2014), the authors sent students, trained as test patients, with identical verbally communicated flu-like complaints to physicians in Chinese hospitals. The institution setting is such that physicians prescribing medication receive kickbacks on medication bought at the hospital pharmacy. The authors analyze whether patients signaling that they are informed about inappropriate antibiotics use are prescribed less antibiotics than other patients. Currie et al. (2011, 2014) find that the signal reduces the probability of receiving an antibiotic prescription by 25 percentage points, from 64% to 39%, and that the signal also reduces drug expenditures. Lu (2014) also investigates physicians' prescribing behavior in Chinese hospitals. The variations are whether patients are insured and whether they indicate to purchase the prescribed drug at the hospital. Lu (2014) finds that physicians write prescriptions that are significantly more costly for insured than uninsured patients, but only if physicians receive kickbacks.

In a large audit study in India, Das et al. (2016) compare physician effort and treatment between private and public health care providers. Test patients are sent to physicians and communicate symptoms of one out of three predefined diseases. Das et al. (2016) find that physicians' diagnosis and treatment quality do not vary between public and private providers although private providers lacked medical qualifications. Private providers balance their worse qualification by a significantly higher effort level.[14] The level of unnecessary treatments is high under both public and private health care provision. In fact, 70% of the providers provide an unnecessary treatment. Das et al. (2016) also compare the behavior of physicians with both public and private practices and find that all quality metrics are higher in their private clinics.

In contrast to the above studies, in our design the diagnosis is not based on communicated symptoms (of different patients), but on examination of the patient and an x-ray. Our case is thus more strongly patient-specific reflecting true credence goods characteristics. This difference is important, because patient-independent information such as inappropriate antibiotics use for a flu cannot easily improve outcomes in our case, and credence goods markets in general. With overtreatment recommendations, we also provide a simple measure of physician quality. A further novelty is the analysis of the role of market conditions and the qualification of the use of physician density as an indicator of competition intensity in the context of physician-induced demand.

Credence goods field experiments

There are four field experiments in a credence goods markets framing that relate to our paper:[15] Schneider (2012); Balafoutas et al. (2013, 2017); and Kerschbamer et al. (2016).

[14]In a preceding study, Das and Hammer (2007) investigate differences in quality of health care provision in India using vignettes as well as accompanying physicians during one day in their practice. They find as well that physicians in public hospitals exert significantly less effort than physicians in private hospitals however physicians in public hospitals exert more effort than physicians in small public clinics. In this study, a physician's effort and his compentence are positively correlated: the more competent a physician is, the more effort he exerts.

[15]Due to the challenges of designing and performing field experiments in credence goods markets, the number of laboratory studies has grown in the past years. Dulleck et al. (2011)'s seminal paper investigates the impact of market institutions on expert behavior under endogeneous prices. Mimra et al. (2016a) shows that price competition may inhibit quality competition and thus lead to more inefficient market outcomes than regulated prices. Mimra et al. (2016b) show that second opinions may be an effective instrument to reduce the level of overtreatment. In contrast, Huck et al. (2016) find that insurance increases the level of overtreatment.

Schneider (2012) finds overtreatment in an undercover experiment in the market for car repairs. Balafoutas et al. (2013) and Balafoutas et al. (2017) perform field experiments in the Greek market for taxi rides. Their results in the market for taxi rides contrasts ours from health care: While they find that customers with less information are overtreated (taken on detours) more often than well informed customers, they find no significant differences in overtreatment across customers' income levels. Balafoutas et al. (2017) report evidence for second-degree moral hazard: Customers who indicate that their expenses are reimbursed by their employer are overtreated more often than those customers that do not. By varying insurance coverage, Kerschbamer et al. (2016) confirm the importance of second-degree moral hazard in the market for computer repairs. We add to this literature on credence goods with a field study in the health care market, the economically most significant credence goods market.[16]

3 The market for dental care in the canton of Zurich

The canton of Zurich is the most populous among the 26 cantons of the Swiss Confederation. In December 2015, the canton had 1.46 million inhabitants (Switzerland: 8.31 million). The two largest cities in the canton are Zurich and Winterthur with 396'000 and 108'000 inhabitants, respectively. In 2014, there were 4,217 registered dentists (57 per 100'000 population) in Switzerland among which 823 (51 per 100'000 population) were working in the canton of Zurich (BfS). Updating the population of dentists leads to (see Appendix A.2 for the construction of the dentist population) 865 dentists in the canton of Zurich among which 402 (46%) were located in the city of Zurich, 70 (8.1%) in the city of Winterthur and 393 (45.4%) in the other municipalities of the canton.

The market for dental care in the canton of Zuerich is characterized by a dual system of providers: dentists and dental hygiene practices. Whereas dentists provide dental care in the classic sense, dental hygiene practices focus on preventive measures. As a part of these, dental hygiene practices may perform x-rays, for instance to check for interproximal caries. It is not uncommon that patients bring recent x-rays taken at other providers to dentists for diagnosis, which we will use in our experimental design. Dentists are mostly self-employed and are thus residual claimants of their provided services.

The Swiss Dental Association publishes binding treatment guidelines for common cases (Schweizerische Zahnärzte-Gesellschaft, 2015a). Adult patients are essentially not insured

[16]An interesting difference compared to other credence goods markets is that social preferences such as altruism and ethical norms are presumably playing a more important role in expert behavior.

for dental care, the association estimates that 85-90 % of the expenses are paid out-of-pocket by the patients.[17] Only if patients suffer from an accident or a severe underlying disease, health insurance covers dental expenses (Schweizerische Zahnärzte-Gesellschaft, 2015b).

The prices for dental care are regulated according to the Swiss Dental Tariff (Schweizerische Zahnärzte-Gesellschaft, 2017) in the following way. The total price is a combination of two components: the number of points multiplied by the point value (PV).[18] The number of points is regulated for each treatment, i.e. it is specified in the Swiss Dental Tariff how many points can be attached to a particular treatment. To be more precise, there is a small interval of points that can be assigned to each treatment. The upper bound of this interval is 33 to 35 percent above the lower bound. Based on the difficulty of treatment dentists may choose to apply slightly more or less points. As an example, the point range in the Swiss Dental Tariff for a standard consultation for diagnosis including anamnesis, checking for caries, inspection of the oral cavity etc. and information and discussion with the patient is 18-24, with 21 being the indicated standard rate. The point value is the multiplier of the points and is chosen individually by each practice. By law, each practice has to publish the point value that the practice applies (Staatssekretariat für Wirtschaft, 2004). The regulator sets an upper limit of CHF 5.8 for the choice of the point value. As an example, if the dentist's practice has chosen a point value of CHF 4.0, a standard diagnosis consultation of 21 points is billed with 21 $\times$ CHF 4.0 = CHF 84 (about \$86).

4 The field study

4.1 Case & test patient

A field experiment in the health care markets is delimited by several strong requirements[19] with respect to the test patient's case: Physicians have to be able to diagnose the patient's condition without side effects. Furthermore, the condition must not change during the time of the experiment to ensure that all diagnoses are made under the same prerequisites. The diagnosis itself must be based on identical information for each of the visits. Additionally, treatment guidelines have to give a clear recommendation of what is the appropriate treatment and what is considered to be overtreatment. The condition of our test patient, a minor superficial caries lesion between two teeth (interproximal), satisfies all of these criteria. One

[17] Imfeld (2008) reports that even 94% of the Swiss dental care expenditures are paid out of pocket.

[18] The German expressions used in official documents are "Taxpunkte" (tax points) and "Taxpunktwert" (tax point value). For clarity we will from now on refer to "points" and "point value", respectively.

[19] See List et al. (2011); List (2011) for requirements for field experiments.

of our reference dentists took an x-ray picture at an initial visit that displayed the superficial caries lesion. Based on the x-ray and an inspection of the patient, all four reference dentists stated independently that the superficial caries lesion had not yet progressed to the dentin and should hence not be treated according to the treatment guidelines. Instead, the patient should be told to continue to well brush his teeth and that the lesion should be checked at a visit in a year's time.[20] The case is a standard case that does not require particular diagnostic effort or skill. Our four reference dentists agreed that diagnosing the interdental caries is almost effortless and hardly allows for diagnostic error.

It has to be noted that the test patient additionally suffered from two even more minor caries lesions and one shadow underneath an existing filling[21]. All four reference dentists agreed that these even more minor caries lesions clearly do not require treatment and that the shadow is clearly recognizable as not being a caries due to its sharp borders and does not require any treatment.[22]

In the study, the test patient presented the same x-ray at each visit and reported to have been shortly before at a dental hygiene practice that performed the x-ray and gave the recommendation to visit a dentist for a checkup. Due to the duality in providers (see section 3) and the fact that patients pay expenses out-of-pocket, it is not uncommon that patients present x-rays to dentists even at a first visit. Providing the x-ray ensured that all dentists would base their diagnosis on the same information.[23] A second x-ray after the study shows that the caries lesion of our test patient did not advance during the study.[24]

Overtreatment Based on the above, we define a recommendation that includes at least one filling to be an overtreatment recommendation. We furthermore use the number of suggested fillings as well as the billed amount as indicators for the extent of overtreatment.

Not only are the requirements high for the patient's condition, several criteria also have to be met by the test patient: Besides having the "correct disease", the minor superficial interdental caries lesion, the test patient had to be available for a long enough time span and be willing to conduct the large number of 180 dentist visits. Reliability was another key characteristic that our test person had to fulfill. To find candidates that would meet

[20] The indicated check-up is a year later, showing that nothing needs to be done during a considerable timeframe.

[21] The shadow does not require treatment as it corresponds to a bonding system without radiopacity.

[22] In terms of design, it would certainly have been preferable to have a test patient without these minor additional imperfections. However, from discussions with one of the reference dentists, it appears virtually impossible to find such a case.

[23] Furthermore, additional x-rays, given their side effects, are precluded for ethical reasons.

[24] Note that this second x-ray was, of course, also medically indicated as the yearly check-up.

all the criteria, we sent out more than 6'000 emails to the University of Zurich psychology department's subject pool and placed an advertisement online at the universities' platform *Marktplatz*[25]. The text in the email and advertisement was phrased very generally: A person which might have a teeth-related health problem such as a beginning caries was sought for a research project. Forty-nine persons replied to our search. We sent out a detailed questionnaire to 44 candidates who were then interviewed on the phone. The most promising candidates were invited for an interview and a visit to one of our reference dentists. The next step was an assessment of the candidates' cognitive skills and reliability. We found a male person in his twenties that fulfilled all criteria. All four reference dentists checked the test patient and independently agreed that the test patient was suffering from a minor interproximal superficial caries lesion and that no treatment was indicated.

The recruitment process was followed by an intense training of the test patient. In cooperation with the reference dentists, we worked out a detailed script that was tailored to the case and our test patient's characteristics. Once the role scripts were trained and the outfit was defined, the test patient was sent to five different dentists as real-life training sessions. The training was completed by another visit of one of the reference dentists where the test person's appearance and script was once more evaluated.

4.2 Experimental variations

We vary two test patient characteristics between the visits: the information that the test patient signals to the dentist and whether the test patient is perceived as a patient with a lower or higher socio-economic status. *Table I* summarizes the conditions and provides the number of observations per condition in parentheses.

TABLE I
Experimental conditions and number of observations.

		Information	
		Standard	Informed
SES	Low	ST-LS (45)	INFO-LS (45)
	High	ST-HS (45)	INFO-HS (45)

To indicate a *higher socio-economic status*, the test patient wore a high quality suit and high-

[25]Marktplatz is an online trading platform provided by the University of Zurich and the ETH Zurich and can be accessed here: http://www.marktplatz.uzh.ch/ (accessed on July 11th, 2017).

end accessories such as an expensive watch, a car key and a new and expensive mobile phone. The test patient specified his occupation as a translator at a bank when asked to fill out the patient form. In the *lower socio-economic status* role, the test patient wore cheap unbranded clothes, an old backpack and had no accessories. The test patient declared to be a student of translation doing an internship.[26] Note importantly that in both socio-economic status-roles the patient signals to have studied. Hence, the level of education was kept constant across the two conditions. We took pictures of the test persons' outfits under the different roles (see *Figure A.4* in *Appendix A.1.2*) in order to ensure an identical outfit on all visits.

We call the condition *informed* when the test patient indicates to the dentist that he has uploaded his x-ray—out of curiosity—to an internet platform where dentists offer free advice.[27] The test patient further states that he had not yet received a reply to his x-ray upload in order to signal that he was expecting to get another diagnosis based on the x-ray but had yet not received one.[28] In the *standard* patient role, there was no additional text to the script such that the patient went to the visits as for standard doctor visits.

4.3 The role of market conditions

The two important economic margins that the dentists face when making treatment decisions are the profit margin on treatments and the dentist's available time, i.e. remaining free capacity. Emons (1997, 2001, 2013) has highlighted the importance of capacity and the incentives to fill up remaining capacity in credence goods markets in theoretical work. In our case, available capacity in the short to medium term is essentially set by the opening hours of the dentist's practice and staffing in the practice and can be only minorly adjusted by working longer hours. Now assuming profit maximization, if the profit margin of a treatment is positive, the treatment is performed as long as there is free capacity to do so. If there are several treatments possible with different margins, then capacity is filled with treatments with highest profit margins first, but as long as free capacity remains also treatments with low margins are performed. For our case, marginal costs of fillings and some flexibility in billing ensure that the profit margin is positive for the range of possible practice price levels.

From the above, the crucial variable is the remaining free capacity in the short term. With this reasoning, whether overtreatment recommendations are given should only depend on

[26]See *Figure A.4* in *Appendix A.1.2* for a photograph of the test patient's outfits and accessory.

[27]An example for such a platform is `www.zahnforum.org` (accessed on July 13th, 2017). If dentists asked the patient on which platform he uploaded his x-ray to, the test patient referred to this platform.

[28]Note that dentists consulted via internet only compete with practices in terms of diagnosis but not treatment. The platform is intended as a platform for information and diagnosis, but not to channel patients to dentists for treatment.

whether there is free capacity or not, but *not* on the practice price level, given that profit margins can be ensured to be positive. Thus, we also should not find a significant effect of the practice price level on overtreatment unless alternative profit opportunities vary systematically with the practice price level.

In the field experiment, we measure short-term free capacity by the waiting time to the next possible appointment. The waiting time directly captures whether, in the coming weeks, demand at a dentist is such that the dentist is already fully booked or still has free capacity available. A standard measure of competition intensity in health care is physician density. There is a considerable body of empirical work on physician-induced demand in which the incentives to induce demand are proxied by physician density. The logic of the density argument is based on the notion that, in a given location with a given demand, a higher density leads to less 'true' demand at physicians, who compensate by inducing demand. This requires several underlying assumptions to be satisfied: First, the assumption that physicians have no or little reason to induce demand at lower physician density, for instance because they operate close to the capacity constraints. Second, the assumption that an increase in density reduces demand at individual physicians such that there is unused capacity in the short- to medium term, for instance if patients are equally likely to choose each physician and then there is unused capacity at each physician which leads to demand inducement.

Our first observation is that this logic can be better checked with a measure of short-term unused capacity that is less noisy than physician density, which we propose with waiting time for next possible appointment. A second observation is that physician density cannot proxy short-term unused capacity if unused capacity is highly dispersed across physicians at all levels of physician density; then there is no strong positive association between physician density and the share of physicians with unused capacity.[29] Our field experiment allows us to identify whether physician density is a good proxy for short-term demand relative to capacity, using our direct measure waiting time for the next possible appointment.

We will thus look at the following three measures for the impact of market conditions: Waiting time, the practice price level and dentist density. Based on the above observations on the role of capacity in short-term decision-making, we expect waiting time to be significantly negatively associated with overtreatment recommendations, however we do not expect a significant effect of the price level. Furthermore, we only expect dentist density to have a significant effect if waiting time is not controlled for and if we simultaneously find a strong negative correlation between physician density and waiting time, such that dentist density

[29] An explanation for this might be that perceived quality of physicians is quite heterogeneous such that demand across physicians is quite heterogeneous, even at different levels of physician density.

picks up the unused capacity if it is not controlled for with a less noisy measure.

4.4 Procedure and data collection

Random draw of dentist sample

Our database listed 865 practicing dentists in the canton of Zurich. We randomly drew 180 dentists from this dentist population.[30] Each of these 180 dentists were visited by our test patient. All visits were conducted in 2016. *Figure I* illustrates the location of the visits. Among the visits, 78 dentists (43.33%) were located in the city of Zurich, 15 (8.33%) in the city of Winterthur and 87 (48.33%) in other municipalities. These shares reflects the population of dentists well among which 402 (46%) is located in the city of Zurich, 70 (8.1%) in the city of Winterthur and 393 (45.4%) in other municipalities of the canton.

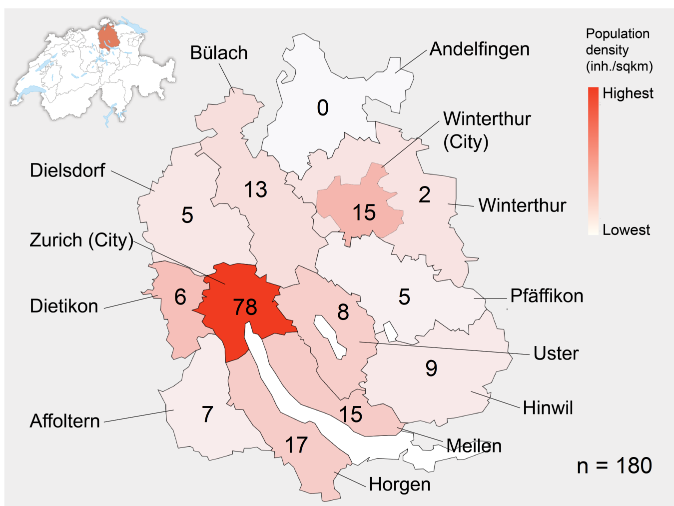


FIGURE I
Map of the canton of Zurich, Switzerland. Observations per district and population densities.

Table II presents means and standard deviations of the covariates for each of our four experimental conditions. The covariates will be presented in detail in the next section. The ANOVA analysis reveals that the randomization worked well as there are no significant differences between any pairs of means of the same covariate (*Table II*). At a first glance, the

[30] When drawing two dentists from the same practice, we randomly replaced one of the dentists by a randomly drawn dentist from the population. This procedure led to a slight oversampling of dentists in single practices.

differences in waiting time and price level displayed seems to be large. Pairwise tests, however, confirm that these differences are not significant (see *Appendix A.4* in *Tables A.2* and *A.3*).

TABLE II
Balance of covariates.

Variable	Treatment				Total
	ST-LS	INFO-LS	ST-HS	INFO-HS	
Waiting time for appointment (days)	8.756 (9.815)	8.267 (8.874)	10.58 (9.555)	7.511 (7.102)	8.778 (8.897)
Dentist density	0.561 (0.565)	0.554 (0.582)	0.719 (0.626)	0.751 (0.526)	0.646 (0.578)
Practice price level (PV)	3.840 (0.353)	3.859 (0.294)	3.922 (0.288)	3.880 (0.282)	3.875 (0.305)
Price level (PV) displayed	0.422 (0.499)	0.467 (0.505)	0.333 (0.477)	0.356 (0.484)	0.394 (0.490)
Informative webpage	0.578 (0.499)	0.733 (0.447)	0.622 (0.490)	0.711 (0.458)	0.661 (0.475)
Swiss licence age (years)	19.22 (10.90)	19.47 (10.27)	19.98 (9.635)	18.96 (10.73)	19.41 (10.31)
Practice owner	0.822 (0.387)	0.800 (0.405)	0.867 (0.344)	0.867 (0.344)	0.839 (0.369)
Median income in area (cont.)	53.09 (6.354)	52.40 (6.677)	50.32 (7.672)	51.75 (6.904)	51.89 (6.937)

Note: the table reports means and standard deviations (in brackets) for covariates over conditions.

Dentist visits

The 180 visits were conducted as follows: The test patient called the randomly selected dental practices in a randomly determined order. At each call, the test patient asked for the earliest available check-up appointment.[31] After arranging the appointment, the test person

[31] The test patient was trained to ask this question in a way that did not signal that seeing a dentist was urgent. In most cases the test patient did not choose the earliest offered date for the actual appointment.

visited the respective dentist. Visits were conducted based on the script. The script indicates that the test patient provides the dentist with the digital x-ray and tells the dentist that he has recently been at a practice for dental hygiene where the x-ray picture had been taken.[32] The patient furthermore says that the dental hygiene assistant recommended seeing a dentist which is why he was here for a check-up. If the dentist proposes a treatment, the test patient is instructed to ask for a cost estimate. After each visit, the test patient completed a detailed protocol about the visit in order to document the communication with the dentist as well as a set of dentists' and practices' characteristics.

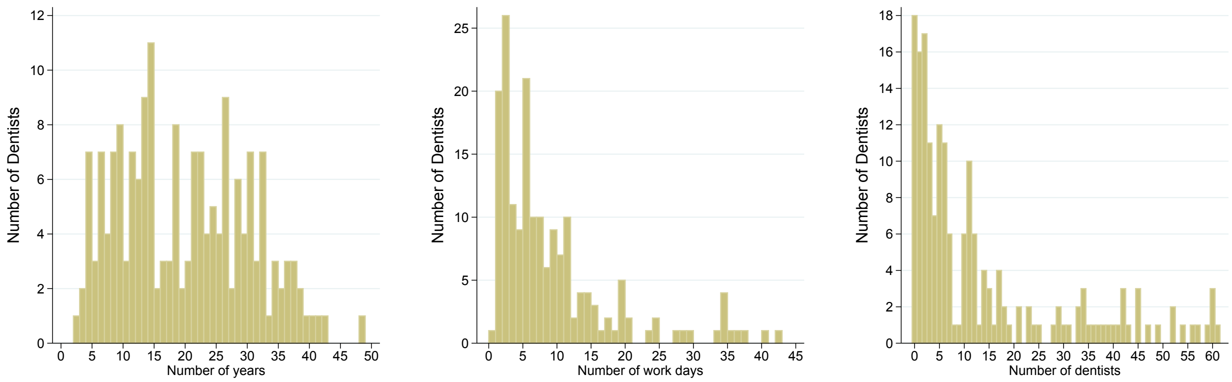
Further data sources and description of variables of interest

Our experimental data is complemented by a unique dataset that combines information about the market, practices and dentists. Table III provides an overview on the variables' descriptions and measurements. For the *dentist* data we reverted to the Swiss Medical Register (MedReg, Bundesamt für Gesundheit (2015)) and updated the register for recent changes. The Swiss Medical Register provides information on dentists' gender, nationality, education, licensing, and specialization. We complemented this registry data by information on the practices. We collected the practice information from the dentists' web-pages and during the visits. Practice characteristics include the practice age as well as information about the practice owner.

Market conditions The waiting time for the next available appointment was collected at the initial call made by the test patient. It is adjusted for weekends and closed practices due to vacations. The practice price level was taken from the diagnosis bills that the test patient received. For dentist density, we take the number of other dentists' practices within a 500 meters distance and adjust it by the number of inhabitants and workers to account for different population densities.

Practice and dentist characteristics & other variables We complement our data with the following variables for practice and dentist characteristics: First, as a measure of transparency at the level of the practice, we use whether a not the point value (price level) is displayed in the practice. Transparency about the point value is required by regulation, yet the majority of practices do not display the point value in their practice. Additionally, we include a binary variable measuring whether a dentist practice has an informative web

[32]For the time of the study, the test patient received a dental hygiene treatment in regular intervals to support the story. The date indicated on digital x-ray was regularly updated.



(a) Distribution of dentists' license (approbation) age.

(b) Distribution of the waiting time for the next possible appointment

(c) Distribution of number of dentists in 500m perimeter around practice.

FIGURE II
Distribution of key variables.

page. A web page is thereby defined as informative if it provides information on the services offered in the practice or the dentist's biography. Another important variable is since how many years the dentist had been licensed to practice in Switzerland. Besides reflecting the age of the dentist, the license age also gives an indication whether dentists still repay loans for initial investments when setting up practice, or whether these are already paid off. We also collect the data on whether or not a dentist owns the practice, making the dentist the residual claimant. As an additional variable, income within the vicinity of a practice is included in the analysis, as a higher local income may imply different practice and treatment recommendation styles.

4.5 Descriptives

We first describe the data on dentists before turning to the visit and market characteristics. The visited dentists were females in 32.2% (58/180) and males in 67.8% (122/180) of the cases. Dentists had been possessing their approbation for on average 21.7 years (sd: 9.28), with a minimum of 3 and a maximum of 49 years. The average license age was 19.41 years (sd: 10.31) (see *Figure IIa*). In total, 47.8% (86/180) of the dentist were working in a single practice and 52.2% (94/180) dentists were working in a group practice with at least one other dentist. The dentists visited obtained their diploma in Switzerland in 73.9% (133/180), in Germany in 15.6% (28/180) and in other countries in 10.6% (19/180) of the cases.

Our test patient visited the dentists on average 7.59 work days (sd: 9.71) after the date of the first possible appointment that was offered. When entering the practice, the test patient

TABLE III
Description of variables on dentists, visits, and the market.

Variable	Description
Market condition measures	
Waiting time for appointment (*short term demand*)	Number of work days between phone call and earliest offered appointment (excluding dentist vacations). Source: phone call protocols.
Dentist density (*long term competition*)	Dentist density per 1'000 workforce-adjusted inhabitants. Formula: $A/(\frac{B+C}{2}) \times 1000$, where A: number of dentists within a 500m circle around the practice; B: inhabitants in the municipality/city district (Zurich and Winterthur); C: Full-time equivalent work-force in the municipality (rural areas) or city district (Zurich and Winterthur). Source: own research.
Practice price level (PV)	Value in CHF that a practice attaches to every point it charges. The number of points that can be attached to a service is defined in the Swiss Dental Tariff. The practice has to use the same PV (point value) for all patients across all services with some exceptions for state welfare recipients. Source: diagnosis bills.
Dentist & practice variables	
Price level (PV) displayed	Indicator whether the visited practice complies with the obligation to display its price level (PV) in the practice well visible. Source: test patient protocols.
Swiss licence age	Number of years the dentist has been possessing his/her licence for working in Switzerland (in 2016). Source: MedReg
Informative webpage	Indicator variable that the webpage of the practice contains information either on the biography of the dentists working in the practice or on the offered spectrum of services. Source: own research.
Practice owner	Indicator whether the treating dentist is the owner or one of the owners of the practice. Sources: own research of webpages and phone books, telephone calls with practice assistants, MedReg.
Other variables	
Median income in area	Median income per year (in k CHF, 2013) in the municipality (and on district level in the city of Zurich) of the practice. Source: Statistical Office of the Canton of Zurich.

TABLE IV
Descriptive statistics of variables on dentist, visit, and market characteristics.

Variable	Min	Max	Mean	SD
Market Condition Measures				
Waiting time for appointment	0	43	8.78	8.90
Dentist density	0	2.33	0.65	0.58
Practice price level (PV)	2.80	4.85	3.88	0.30
Dentist & Practice Variables				
Price level (PV) displayed	0	1	0.39	0.49
Informative webpage	0	1	0.66	0.47
CH-Licence age	2	49	19.41	10.31
Practice owner	0	1	0.84	0.37
Other Variables				
Median income in area	35.24	69.90	51.89	6.94

Note: values rounded to two digits.

waited on average 6 minutes and 33 seconds (sd: 7.36, min: 0, max: 35) until he was asked to follow to the examination room. On average, the test patient spent 19 minutes and 40 seconds in the examination room (sd: 7.57, min: 5, max: 50).

Regarding the market variables, we find that the waiting time for the next appointment ranges from zero to 43 work days. On average, the next possible appointment was offered between eight and nine work days from the initial phone call (mean: 8.78 work days, sd: 8.90). The distribution of the waiting time is illustrated in *Figure IIb*.

The number of competitors within a radius of 500m ranged from 0 to 61 (see *Figure IIc*). Adjusting the number of competitors for the number of inhabitants and workers in the municipality, we find our long-term competition measure physician density to range from 0 to 2.33 with an average of 0.65 (sd: 0.58). The price level of practices ranged from factor 2.8 to 4.85 (average: 3.88, sd: 0.49). Only 39% of the practices displayed the price level in the practice. *Table IV* summarizes the descriptives.

5 Results

5.1 Treatment recommendations

5.1.1 Main result

Our test patient received an overtreatment recommendation in more than every fourth visit. More precisely, dentists suggested at least one filling in 27.78% (50/180) of all visits.

Conditional on an overtreatment recommendation, mean overtreatment costs taken from the collected cost estimates amount to CHF 535 (about $550), the median being lower at CHF 444 (about $455). Regarding the treatment, the suggested number of fillings per dentist ranges from 1 to 6. An illustration is provided in *Figure III*. Furthermore, we observe across all cost estimates that 13 different teeth are to be treated with a filling. Thus, besides our finding of a considerable overtreatment recommendation rate, we also observe a striking heterogeneity in the treatment recommendations.

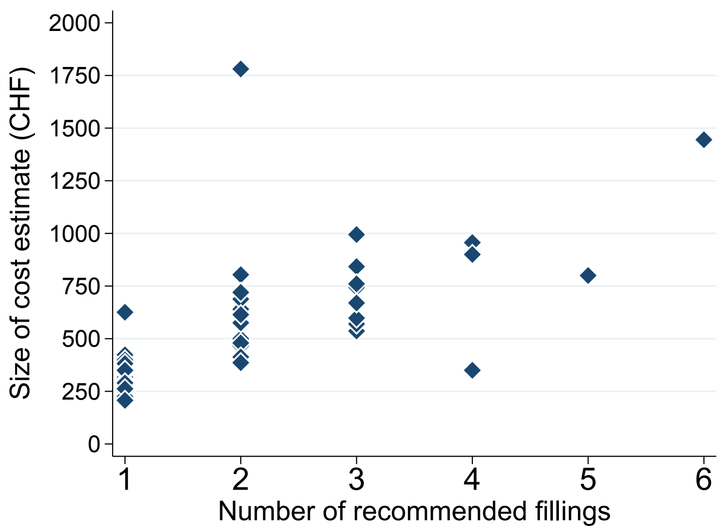


FIGURE III
Cost estimates per number of fillings for overtreatment recommendations.

We perform a parametric analysis on our binary overtreatment variable using a random effects probit regression model. The explanatory variables of interest are the two binary factor dummies, $informed$ and $high\_ses$, the interaction effect[33] between the conditions

[33]Note that the interpretation of regression coefficients of interaction terms in non-linear regressions might be misleading (see Ai and Norton (2003)). We therefore also display the results from a OLS regression in model (5) in *Table V*. Both regressions show consistent results.

$informed_i \times high\_ses$, and the different measures for competition $short\_term\_demand$, $medium\_term\_demand$,
$long\_term\_competition$, $price\_level$ and $price\_level\_displayed$. X_i reflects a vector of practice and dentist covariates. This vector includes the age of the dentist's license to practice in Switzerland and an indicator for whether the dentist is the practice owner. Further, we include an interaction term between our measure for mid-term demand, informativeness of the webpage, and the Swiss licence age. Last, we control for the median income in the area of the dentists practice. Hence, our specification is as follows:

$$
\begin{aligned}
overtreatment_i = \\
& \beta_0 + \beta_1 informed_i + \beta_2 high\_ses_i + \beta_3 (informed_i \times high\_ses_i) \\
& + \beta_4 short\_term\_demand_i + \beta_5 medium\_term\_demand_i \\
& + \beta_6 long\_term\_competition_i + \\
& + \beta_7 price\_level + \beta_8 price\_level\_displayed \\
& + \beta_8 X_i + \beta_9 median\_income\_in\_area + e_i
\end{aligned}
$$

Table V provides the results. To deepen the analysis and account for the extent of overtreatment, we run the same model with two more dependent variables, the number of recommended fillings and the cost estimate. *Figure IV* shows the distributions for both number of recommended fillings and the cost estimate. The number of recommended fillings represents count data with values between zero and six. The cost estimate size displays a typical pattern for health care costs. While there are more than 70% zero observations, the smallest cost estimate conditional on overtreatment is 56 points. The distribution further displays a long right-tail with only two observations larger than 300 points with values of 390.5 points and 419 points, respectively. We ran a negative binomial regression for the number of recommended fillings. The choice of the negative binomial model over the poisson model is akin to overdispersion of the data due to the large fraction of zeros. This choice is supported by a likelihood-ratio test. For the cost estimate size we present a GLM estimation with gamma distribution and log-link. The choice of the gamma distribution is indicated by a Modified-Park Test. *Table VI* shows the results. Model (M3) from *Table V* is displayed for comparison in column (1).[34] In the following, we start with the analysis of the experimental variations. We will present and discuss our results without considering diagnostic errors.

[34] We also considered a hurdle model with model (M3) from *Table V* in the first part and estimations of the amount of overtreatment, conditional on receiving an overtreatment recommendation, in the second part. With only 50 observation for the second part of the model, however, the poisson regression did not have the power to identify robust effects on the amount of overtreatment.

The discussion of our results in light of potential diagnostic error is provided separately in subsection 5.3.

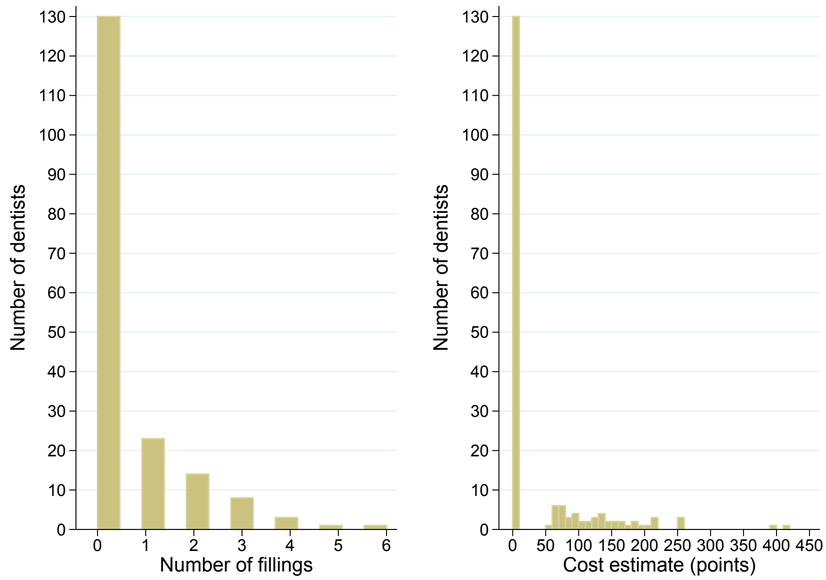


FIGURE IV
Distribution of overtreatment extent.
Number of recommended fillings (left) and cost estimate size (bin-size 10) (right).

5.1.2 The role of market conditions and practice/dentist characteristics

Besides the striking main result of an overtreatment rate of 28%, one of our key findings pertains to the role of short-term available capacity. Model (2) in *Table V* shows that a shorter waiting time for an appointment is associated with a higher likelihood to be overtreated. An additional waiting day reduces the probability to be overtreated by, on average, one percentage point. This result is already apparent in the descriptive statistics. We observe that overtreating dentists have an average waiting time for next possible appointment of 6.16 days whereas non-overtreating dentists have a waiting time of 9.78 days (Mann-Whitney U test, two-tailed (MWU): $p = 0.0008$) (see *Figure V*).

This result is in line with physicians filling up capacities by overtreating patients, which has been previously pointed out in theoretical work by Emons (1997, 2013). The low utilization of dentists' services leads to a lower profit in the short run and thus increases the incentives to provide more services than necessary, which is consistent with profit maximization and the target-income hypothesis.[35] Our result is also consistent with early empirical findings

[35] For a discussion of the target income hypothesis in the context of physician behavior, see, e.g., McGuire

TABLE V
Regressions on a binary measure of overtreatment.

	(M1)	(C1)	(C2)	(C3)	(M2)	(M3)	(M4)	(M5)
	Probit							OLS
Dependent Variable	Overtreatment (binary)							
Experimental conditions								
Information	-0.102	-0.105	-0.102	-0.101	-0.103	-0.112	-0.120*	-0.144
	(0.214)	(0.196)	(0.213)	(0.218)	(0.203)	(0.134)	(0.099)	(0.112)
High SES	-0.174**	-0.163**	-0.171**	-0.166**	-0.159**	-0.173**	-0.177**	-0.178*
	(0.026)	(0.038)	(0.029)	(0.035)	(0.045)	(0.018)	(0.012)	(0.051)
Informed x High SES	0.181	0.162	0.182	0.176	0.156	0.143	0.128	0.138
	(0.225)	(0.267)	(0.224)	(0.238)	(0.284)	(0.297)	(0.343)	(0.279)
Market & competition								
Waiting Time for Appointment (days)		-0.010**			-0.010**	-0.010**	-0.009**	-0.007**
		(0.012)			(0.014)	(0.014)	(0.026)	(0.044)
Dentist Density			-0.015		0.018	0.030	0.019	0.009
			(0.793)		(0.763)	(0.590)	(0.732)	(0.873)
Practice price level (PV)				-0.093	-0.082	-0.026	0.008	-0.008
				(0.386)	(0.454)	(0.808)	(0.940)	(0.942)
Practice & dentist variables								
Price level (PV) displayed						-0.170***	-0.154***	-0.168**
						(0.001)	(0.004)	(0.013)
Swiss Licence Age (years)						-0.003	-0.006	-0.006
						(0.530)	(0.277)	(0.258)
Informative Webpage						0.281*	0.236	0.282*
						(0.095)	(0.161)	(0.077)
Inform. Webpage x Swiss Licence Age						-0.013**	-0.011*	-0.011*
						(0.039)	(0.087)	(0.087)
Practice Owner						0.164*	0.156*	0.185*
						(0.085)	(0.093)	(0.055)
Other variables								
Median Income in Area (cont.)						0.007	0.005	0.005
						(0.135)	(0.237)	(0.256)
Overdiagnosis (conservative definition)							0.210**	0.219**
							(0.036)	(0.018)
Constant								0.161
								(0.728)
Pseudo R^2	0.017	0.048	0.017	0.021	0.050	0.160	0.184	
R^2								0.199
N	180	180	180	180	180	180	180	180

Probit regressions display average marginal effects.
p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE VI
Regression results on the likelihood (models 1 and 2) and amount (models 3 and 4) of overtreatment.

	(1)	(2)	(3)	(4)
	Probit		NBREG	GLM
Dependent variable	Overtreatment (binary)		Number of fillings	Cost estimate (points)
Experimental conditions				
Information	-0.112	-0.120*	-0.182	-15.141
	(0.134)	(0.099)	(0.314)	(0.285)
High SES	-0.173**	-0.177**	-0.444***	-39.463**
	(0.018)	(0.012)	(0.003)	(0.011)
Informed x High SES	0.143	0.128	0.221	22.545
	(0.297)	(0.343)	(0.578)	(0.259)
Market & Competition				
Waiting Time for Appointment (days)	-0.010**	-0.009**	-0.033***	-2.136***
	(0.014)	(0.026)	(0.006)	(0.001)
Dentist Density	0.030	0.019	0.158	4.122
	(0.590)	(0.732)	(0.344)	(0.655)
Practice price level (PV)	-0.026	0.008	-0.232	-2.881
	(0.808)	(0.940)	(0.335)	(0.830)
Practice & Dentist Variables				
Price level (PV) displayed	-0.170***	-0.154***	-0.301***	-26.788**
	(0.001)	(0.004)	(0.010)	(0.014)
Swiss Licence Age (years)	-0.003	-0.006	-0.006	0.723
	(0.530)	(0.277)	(0.632)	(0.362)
Informative Webpage	0.281*	0.236	0.759	75.610***
	(0.095)	(0.161)	(0.223)	(0.000)
Inform. Webpage x Swiss Licence Age	-0.013**	-0.011*	-0.035**	-3.494***
	(0.039)	(0.087)	(0.043)	(0.002)
Practice Owner	0.164*	0.156*	0.558*	47.413***
	(0.085)	(0.093)	(0.056)	(0.006)
Other Variables				
Median Income in Area (cont.)	0.007	0.005	0.026**	1.980**
	(0.135)	(0.237)	(0.028)	(0.037)
Overdiagnosis (conservative definition)		0.210**		
		(0.036)		
Pseudo R^2	0.160	0.184	0.082	
N	180	180	180	180

(1)-(2): Probit models displaying average marginal effects.
(3): Negative binomial regression displaying average marginal effects.
(4): GLM regression with Gamma-distribution and log-link displaying marginal effects at the mean.
(1)-(4): p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

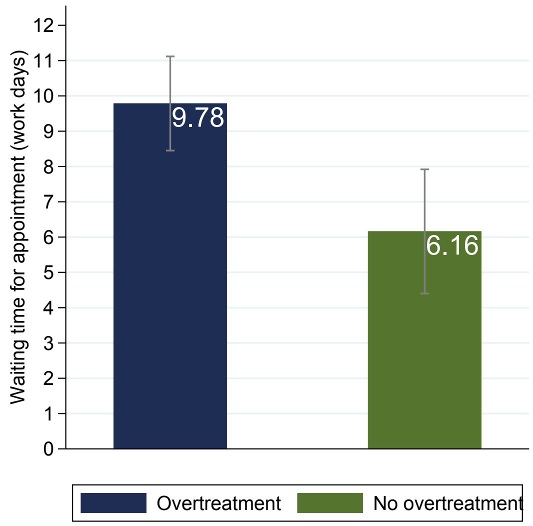


FIGURE V
Waiting time for the next possible appointment (days) by overtreatment with 90% confidence intervals.

by Marty (1998). Marty analysis data from an insurance company on physicians' individual treatment decisions and defines upward deviations from the average turnover per patient as overtreatment. He shows that physicians' idle capacities are in fact positively correlated with above average turnovers.

A possible concern with respect to our analysis may be simultaneous causality. If overtreatment was observed by patients they would possibly switch to a different dentist leading to a lower waiting time at overtreating dentists. However, patients cannot observe whether or not they received the appropriate treatment due to the credence good characteristic. Hence, the measured impact of waiting time on overtreatment does not suffer from reverse causality. With respect to the reason behind short waiting times, it might however be that these dentists are perceived by patients to provide low quality service and therefore only face low demand. Unfortunately, we cannot control separately for perceived quality. Although online rating websites start to be in place for dentists, we do not have enough data to construct a meaningful measure of perceived quality.[36] However, this does not pose a problem for our result for which we only require waiting time to be a valid proxy for short-term demand relative to capacity, where unused capacity might be the result of different factors such as low perceived quality or demand shocks.

(2000).

[36]We collected ratings data from four rating webpages. Even on Google, the webpage with the largest number of rated dentists from our sample, 61% have no rating at all, and 90% have less than three ratings.

We do not find a significant effect of our long-term competition measure, dentist density, on overtreatment. This result holds even if we do not control for the other market measures such as waiting time.[37] This result is important in light of the vast literature on physician-induced demand that approaches the topic by analyzing the correlation between physician density and a measure of health care utilization such as annual number of procedures per general practitioner. As discussed in *Section 4*, the logic of the density argument is based on the notion that, in a given location with a given demand, a higher density leads to less 'true' demand at physicians, who compensate by inducing demand. Our study allows us to shed light on the question whether physician density is a good proxy for short-term demand relative to capacity, using our more direct measure of waiting time for the next possible appointment. The result that dentist density is not significant even when not controlling for waiting time is a first indication that dentist density is not a good measure of (short-term) demand relative to capacity. If dentist density did pick up the relevant demand/capacity variable, we should observe a strong negative correlation between density and waiting time and significance of density if we do not control for waiting time, which is not the case. Indeed, looking more closely into our data, we observe that waiting time is highly dispersed across dentists at all levels of dentist density in such a way that there is no negative association between dentist density and waiting time at individual dentists. *Figure VI* displays the waiting time by dentist density for the city of Zurich, for both the city center and non-city center. The scatterplot looks similar for the whole sample (see *Figure A.3* in Appendix A.1.1)

In line with our hypothesis that a positive profit margin and free capacity but not the size of the margin are relevant—as long as alternative treatment and profit options do not vary systematically with the practive price level—, we do not find evidence that the price level itself affects overtreatment in our study. Thus, taken together, our results can be well explained based on the importance of short-term capacity considerations for treatment decisions.

Interestingly, we find that the likelihood to receive an overtreatment recommendation decreases by approximately 16 percentage points (see model (3) in *Table V*) at a dentists who displays the price level in the practice as required by regulation. A possible interpretation for the first result on price level transparency and overtreatment is that there might be different dentist types in terms of abiding to regulation (display price level in practice) and treatment guidelines (overtreatment).

[37] The result is robust to different definitions of the variable such as using an non-population adjusted measure.

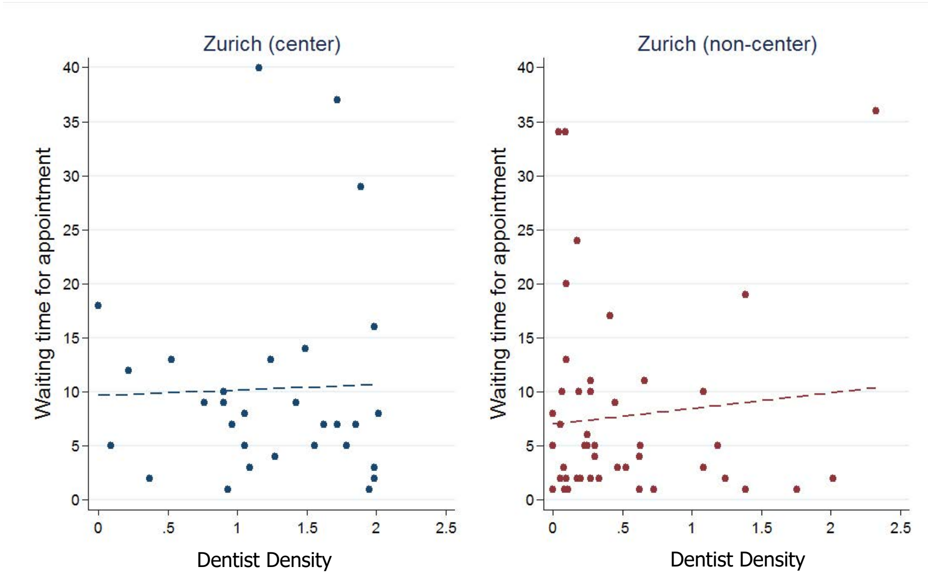


FIGURE VI
Dentist density and waiting time for the city of Zurich

We furthermore observe that dentists with an informative web page are more likely to recommend overtreatment than dentists without an informative webpage. The descriptive statistics reveal that the overtreament rate is 31.10% for dentists with an informative web page and drops by almost ten percentage points to 21.31% for dentists without an informative webpage. The effect is not significant, however (Fisher's exact test: $p = 0.218$), analogues to not being significant in regression model (M2) in *table V*, when we do not control for practice and dentist variables as well as area income. The interaction term between informative web page and the license age is however negative in (M3) and significant. This result is in line with the argument that an informative web page is a demand indicator for dentists primarily at the beginning of their career when building up the patient base. Among the dentists with a licence age below or at the median licence age of 18 years who have an informative webpage, the overtreatment rate is 43.28% (29 of 67 cases). The rate drops to 15.38% (8 of 52 cases) when the dentists are above the median licence age (Fisher's exact test: $p = 0.001$).

When treated by a practice owner, the probability to be overtreated is increased by more than 16 percentage points. Although only weakly significant, this effect is intuitive. Practice owner are residual claimants by nature while non-practice owners are often employed on a fixed income basis[38].

[38] This information was provided to us by our reference dentists.

5.1.3 Effects of socio-economic status and information

Table VII displays the overtreatment rate for each of the four experimental conditions. The probit regressions show that the likelihood to receive an overtreatment recommendation is significantly lower for a patient with a high- than a low SES (see *Table V*) in the standard condition. Being a standard patient with a high SES reduces the likelihood of receiving an overtreatment recommendation by about 17 percentage points compared to a patient with a low SES. In contrast to the standard condition, differences in SES do not translate into different overtreatment recommendation levels when the patient is informed. Both, under low and high SES, the level of overtreatment recommendations amounts to 26.67%.

TABLE VII
Overtreatment recommendations per conditions. Number of observations in parentheses.

		Information		Average
		Standard	Informed	
SES	Low	37.78% (17/45)	26.67% (12/45)	32.22% (29/90)
	High	20.00% (9/45)	26.67% (12/45)	23.33% (21/90)
Average		28.88% (26/90)	26.67% (24/90)	27.78% (50/180)

This result of lower overtreatment with a higher SES as a standard patient is surprising at first thought. In particular, a higher SES—implemented by a more expensive physical attire indicating a higher income while keeping the education level constant—might be interpreted as implying a lower price sensitivity and higher acceptance rate of costly treatment, which increases overtreatment incentives. Thus, one might have expected that the higher SES works in a similar way as the effective consumer price reduction of having (health) insurance. Kerschbamer et al. (2016) for instance show in a field experiment in the market for computer repair that customers' insurance coverage increases the repair price significantly.[39] Our result suggests that in the patient-physician interaction, another mechanism is at work for the SES variation. We will discuss the potential explanations in turn.

Our finding may be explained by the similarity of patients' and dentists' SES. Van Ryn and

[39]Huck et al. (2016) study the impact of insurance on experts' provision behavior in a lab experiment. The authors find that insurance increases the level of overtreatment and the customers' participation rate at the same time, indicating moral hazard of both experts and patients.

Burke (2000), e. g., find that patients from the lower socio-economic class are perceived more negatively by physicians than patients from the middle and high economic class. In a field experiment on taxi driver behavior, Balafoutas et al. (2013) for instance find that taxi customers with a lower perceived income — closer to the low SES taxi driver—are overcharged less often than those with a perceived high income. Contrary to our case, however, the effect of a similar SES and the economic argument of price sensitivity in Balafoutas et al. (2013) go in the same direction.

Another aspect is that the perceived likelihood to return might differ between SES for the standard patient. The importance of reputation-building might play a role.[40] It could be argued that a physician attempts to build-up reputation by not treating a patient with a minor treatment such as a filling, and that this reputation-building concern is higher for patients with a higher SES due to higher future profits from interaction. However, it is not clear whether reputation-building is actually stronger when not providing a treatment initially than when providing a treatment. Another argument related to the likelihood to return is that overtreatment for the standard patient with a lower SES might be preventive in the sense that the dentists expect this patient to be less likely to return to any dentist for check-ups in the future.[41] However, the study design limits the applicability of this argument: By going to the dentist with his x-ray from a dental hygiene practice, the patient shows a considerable interest in his dental health.[42]

A final point relates to an interpretation of a higher SES as implying *better or more information*. In the experimental design, we took care to keep the level of education constant across the variation such that the higher SES goes through higher income but not education level. However, the patient with higher SES might still be perceived as better informed. We will discuss this further below in light of our results from the experimental information variation.

More information as implemented in our experiment does not significantly reduce overtreatment. We observe a considerable drop in the rate of overtreatment recommendations from 37.78% to 26.67% between an informed and a standard patient if the SES is low. However, this difference is not statistically significant. We neither find a significant difference between an informed and a standard patient if the SES is high nor if SES treatments are pooled.

Our results suggest that it is crucial to understand and differentiate between the different

[40] In our experimental variations, due to limitations on the number of visits for our test patient, we could not run a further variation in which the role of reputation-building in repeated physician-patient interactions is analyzed.

[41] This explanation was suggested by one of our reference dentists.

[42] Furthermore, while the patient displays a lower SES compared to the other condition, the patient is by no means a case of very low social status.

types of information that are relevant in the physician-patient interaction and in particular how these are perceived by physicians. Signalling information from the online platform–even when it is on the specific case at hand—does not appear to considerably affect the treatment recommendation of dentists. This may reflect the fact that dentists do not perceive information from the platform as significantly reducing the level of asymmetric information between them and the patient. Furthermore, it might be the case is that dentists assume that they can rationalize their treatment recommendation, independent of differing diagnostics/information from the website. It is generally assumed that information provision and diagnostics from the internet increase patient information and quality of care. Our results show the limits of this argument when the service at hand has credence goods characteristics and is complex, as is the case for most health care services. The importance of the fineprint of what constitues relevant information and limits of information for credence goods is also apparent when comparing our results to the literature: In Currie et al. (2011) and Currie et al. (2014), the authors sent students, trained as test patients, with identical verbally communicated flu-like complaints and find that patients that signal that they are informed about inappropriate antibiotic use are prescribed less antibiotics than other patients. This is a case where both diagnostics and information about correct/wrong treatment is simple, unambiguous and not patient-case specific, such that it does not exhibit essential credence goods properties. Hence, signalling the corresponding information should reduce wrong prescriptions, as observed empirically. Our results point at the difficulty of information and diagnostics via the internet or other sources to address the problem of wrong treatment–be it due to diagnostic errors or physician-induced demand—for more complex cases of health services.[43]

5.1.4 The role of medical schools

Whether training at specific academic institutions affects physician behaviour has recently been analysed by Doyle et al. (2010) and Currie et al. (2016). Doyle et al. (2010) find that physicians trained at a top-ranked insitution operate at lower costs than physicians from a lower ranked institution while they achieve similar health outcomes. Currie et al. (2016) find that cardiologists trained at top-20 medical schools use more invasive procedures and are more responsive to patient conditions. We analyse potential school-effects by testing

[43]Interesting survey evidence is provided by Domenighetti et al. (1993). They find that physicians have a much lower rate of surgeries than regular patients who are not physicians or who do not have physicians (or lawyers) in their families. This evidence is consistent with the information asymmetry hypothesis, however other explanations can also account for this observation, e.g., physicians might have a generically different demand or face different prices.

whether the country in which dentists received their diploma impacts the likelihood that dentists provide an overtreatment recommendation. We group dentists into schools with a Swiss diploma (133/180, 73.89%), a diploma from Germany or Austria (29/180, 16.11%), a diploma from Eastern-European countries (12/180, 6.67%) and diplomas from other countries (6/180, 3.33%).[44] When adding school-dummies to model (M3) in table V, we do not find significant effects of school/country on overtreatment. Hence, the educational background is not important for the results in our study when controlling for other covariates such as licence age.

5.2 Diagnoses

We observe a considerable dispersion in diagnosis fees, illustrated in *Figure VII*.[45] The lowest fee, charged twice, was CHF 0, the highest fee amounted to CHF 212.65. On average, dentists charged CHF 92.62 (sd: 31.17) for the diagnosis.[46] Even when correcting for the pricel level (PV, point value) of the practice, a considerable dispersion persists (right side of *Figure VII*). In the sample, dentists charged between zero and 56 points for the diagnosis with a mean of 23.95 points (sd: 8.02). About half of the sampled dentists charged exactly 21 points for the diagnosis, which is the specified rate for a standard consultation. The average value attached to one point is CHF 3.88 (min: 2.8, max: 4.85, sd: 0.30).

With respect to our experimental conditions, diagnosis fees and points charged per diagnosis do not differ between low and high SES patients (see *Figure VIII* (left)). However, we observe a (weakly) significant difference with respect to the change in signalled information: dentists charged weakly more points for the informed patient than for the standard patient (see *Figure VIII* (right)). The average number of points charged for the standard patient is 22.91 (sd: 6.81) and 25.00 (sd: 8.99) for the informed patient (MWU: $p = 0.094$).[47] This

[44]Data on the diploma country is available on the public administrative platform MedReg. Data about the specific school at which dentists were trained is not available on MedReg and we were only able to collect it for 115 dentists in our sample. Among these, 77 of 83 dentists trained in Switzerland were trained at the University of Zurich. The decision to group all dentists in Switzerland and Germany on country level is moreover justified by the traditionally small heterogeneity in the rankings of medical schools within these countries. For instance, the QS World Ranking 2016 of the top 50 medical schools ranks three of the four Swiss Dental schools at ranks 23, 25 and 49, respectively.

[45]Our test patient received the bills for the check-up visits on average 14.89 work days after the visit (sd: 34.93). The median is only six days, however. While 34 dentists charged directly on the spot, a procedure not uncommon for first visits, seven dentists sent their bills only after 100 days or more after the visit, the maximum being 235 days.

[46]The total of diagnosis fees paid for our study amounts to CHF 16'671.

[47]When excluding the 17 dentists who only charged 14 points or less—14 points are a natural choice because it is the number attributed to the tariff item for the diagnosis of a 'recall patient'—the difference of

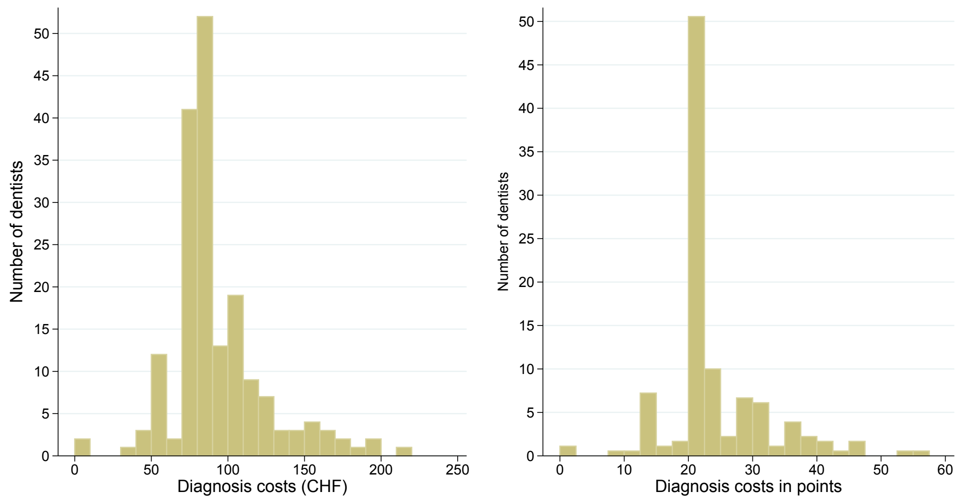


FIGURE VII
Distribution of diagnosis fees and points charged per diagnosis (n=180).

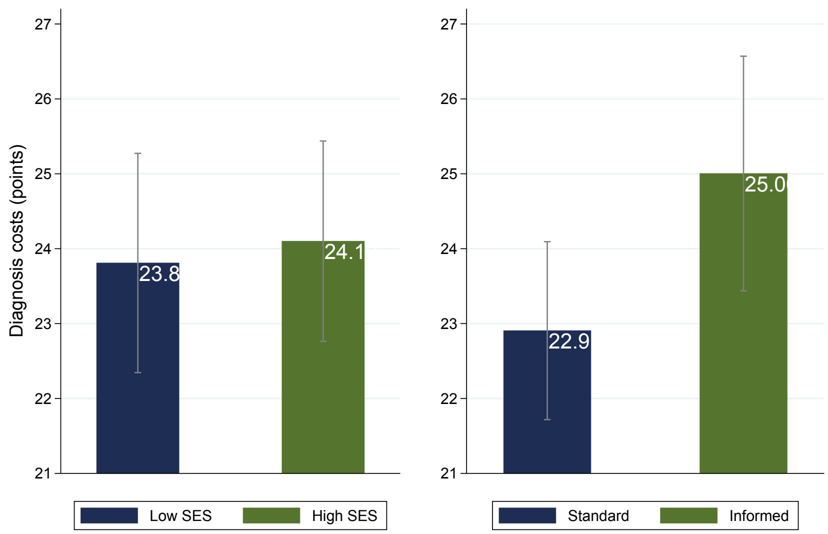


FIGURE VIII
Points charged by SES (left) and patient information (right) (n=180) with 90% confidence intervals. Note: overlapping confidence intervals are no proof that the difference between the means of two groups is not different from zero.

result holds when considering diagnosis fees without adjusting for point value. The diagnosis fee for an informed test patient is CHF 88.96 (sd: 28.08) for the standard and CHF 96.28

the means becomes more pronounced. The rationale for exclusion of these dentists is that these are likely neither interested in present nor future profits from the test patient. The average number of points charged for the informed test patient is then 23.97 (sd: 6.00) and 26.45 (sd: 8.14) for the standard patient (MWU: $p = 0.032$).

(sd: 33.74) for the informed patient (MWU: $p = 0.085$).[48]

Looking further into the diagnosis fees, we distinguish between different diagnosis items as classified by our reference dentists: consultation in points billed, further diagnosis items that are admissible and diagnosis items that constitute overdiagnosis. Overdiagnosis refers to items billed which are not needed for diagnosing the case or providing the necessary information to the patient. An example is the item 'further information about dental interventions' with 15 points. In the Swiss Dental Tariff, it is explicitly stated that this item should not be applied for information to patients about routine dental procedures, to which for instance fillings belong. We identified 27 out of 180 visits (15.00%) with overdiagnosis. *Figure IX* shows the comparison by diagnosis items, in absolute and relative terms, between the standard and the information conditions.

Figure IX shows that the main difference between standard and informed in points charged stems from a difference in overdiagnosis. This reflects the comparison of the number of visits with overdiagnosis across the information conditions: 9 out of 90 in standard compared to 18 out of 90 in informed (Fisher's exact test: $p = 0.094$). Most overdiagnosis items relate to more time spent at the visit. Looking into the time spent in the treatment room, we indeed observe that it is significantly longer for visits with overdiagnosis (25.04 minutes) than without overdiagnosis (18.60 minutes).[49] Thus, dentists appear to be spending (unnecessarily) more time, which is then billed. Interestingly, when running regressions on overdiagnosis value (CHF), we again find that the coefficient for short-term demand is significant and has the same sign as for overtreatment.[50]

Model (4) in *Table V* shows that overdiagnosis is also associated with a higher likelihood to be overtreated. Clearly, endogeneity is a problem at hand, since providing a treatment recommendation for a filling might induce more time spent with the patient that is then billed via additional diagnosis items. Indeed, we find a significantly higher time spent in the treatment room for overtreatment than for no overtreatment recommendations (22.12 minutes vs. 18.73 minutes; MWU: p=0.0242).

[48] The cumulative distribution functions of diagnosis fee and points charged are shown in *Figure A.2* in *Appendix A.1.1* separately for informed and standard patient.

[49] This provides the basis for the classification as overdiagnosis and not as overcharging at diagnosis.

[50] Our data does not provide sufficient power to identify a significant effect of short-term demand on the likelihood of receiving an overdiagnosis.

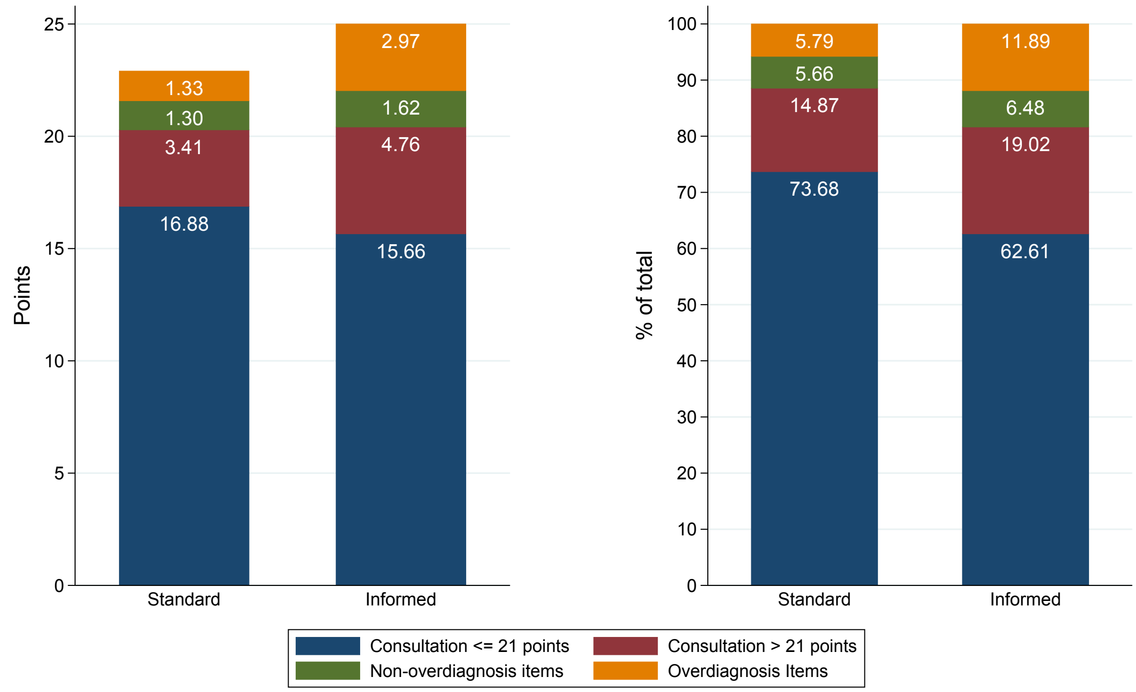


FIGURE IX
Composition of points charged absolute (left) and relative (right) scales.

5.3 Discussion

Robustness In further specifications of our regression analysis, we controlled for additional dentist, visit and market characteristics such as questions the dentists asked during the visit, whether the practice was a single practice or shared by more than one dentist, gender and membership in Swiss dental association. However, we do not find a significant impact of these possibly explanatory variables. Our results are robust against changes in the specification and against different regression models, such as the logit, negative binomial or zero-inflated model. We also investigate whether our test person's behavior changed over time as the test patient became more experienced. However, seasonal dummies do not show to have an impact on the regression results. In the study design we made use of several precautionary measures to prevent that the medical condition of our test patient changed over time. First, all diagnoses have been based on the same x-ray which our test patient brought to every visit. Second, the oral condition of our test patient was confirmed after 30, 60, 120 and 180 visits by our reference dentists and the test patient was undergoing dental hygiene regularly.

Diagnostic errors In the design, we took care to select a case for which the scope of diagnostic error should be minimal. Yet, we cannot fully exclude diagnostic errors. Regarding the information variation, diagnostic errors could explain that we do not find a significant difference between standard and informed. If dentists are not aware that their treatment recommendation is an overtreatment recommendation, an information signal from the patient does not change recommendations. Note, however, that we should not observe treatment differences between the lower and higher SES patient under standard information if diagnostic errors were the only driver of our results. Second, diagnostic errors could play a role in our result on waiting time. If there were a high correlation between dentists who make diagnostic errors and recommend overtreatment and dentists that are perceived as of low quality by the patients and the latter are less frequented by patients, then this would explain why we observe more overtreatment with a shorter waiting time. However, it is unclear that this correlation is high, as patients do not observe objective quality and might base their estimate of perceived quality on other dimensions such as friendliness of the doctor and staff.[51] Taken together, we cannot exclude that our results stem, at least in part, from diagnostic errors. However, this does not inhibit the main results: We observe a high rate of overtreatment and overtreatment is associated with a low short-term demand at the dentist.

Limitations Ideally, we would be able to control for dentist fixed effects, which is not possible with our study design. A field experiment involving real patients with sufficient power to do so is however extremely difficult to implement and we consider our experiment with 180 visits of a single patient to identify overtreatment a good start for the direct observation of physician behavior. Furthermore, regarding our experimental variations, we presented several explanations for our result on the impact of SES on overtreatment and cannot exclude alternative explanations. It is apparent that SES might operate through many channels. Our goal was to identify whether we observe an effect on the level of individual patient-physician interactions, and our results show a complex role of SES interacting with the level of information of a patient. A design that implements experimental variations to differentiate between the competing explanations of the role of SES is beyond the scope of the current paper and an interesting avenue for future research.

Last, we are able to observe the appropriateness of the treatment recommendation but cannot control for patients' overall perceived quality of a dentist. As overtreatment is not observable by patients, we are confident that our findings on waiting time are not driven by

[51] Fornara et al. (2006), e.g., show that hospital users' perceived quality of care improves when the humanization degree of the hospital environment increases. Arneill and Devlin (2002) conducted a study where they showed participants slides of doctors' waiting rooms and then asked what quality of care participants expected. Arneill and Devlin (2002) find that the perceived quality of care would be significantly higher for waiting rooms that are nicely furnished, light, contain artwork and are warm versus waiting rooms that are dark, have outdated furnishings and are cold in appearance.

reverse causality. However, we cannot exclude that general perceived quality drives short-term demand at a dentist and via this channel the likelihood to overtreat. We try to address this shortcoming by using feedback and ratings on online platforms as a proxy for perceived quality, however there are not sufficiently many patient ratings to do so in a meaningful way.

Overtreatment results and the scope for second opinions Combining our results on overtreatment and diagnosis costs, we observe that searching for a second opinion from another dentist might be worthwhile. Assume that a patient's prior for needing a treatment such as a filling is ρ. Denote by P_H the price that the patient has to pay when being treated and by P_L when not being treated. Now if physicians give a treatment recommendation with probability x to a patient who does not need treatment, and the patient's costs for searching is k, then, under risk neutrality, searching for a second opinion is worthwhile for the patient if

$$P_H > k + \frac{\rho}{\rho + (1-\rho)x} P_H + \frac{(1-\rho)x}{\rho + (1-\rho)x}(xP_H + (1-x)P_L).$$

If the customer does not search for a second opinion, she pays the price P_H for sure. If she searches for a second opinion, she incurs search costs of k and again has to pay the high price P_H if she indeed needs treatment (which happens with probability $\frac{\rho}{\rho+(1-\rho)x}$) or receives an overtreatment recommendation again (which happens with probability $\frac{(1-\rho)x^2}{\rho+(1-\rho)x}$). She only pays the lower price P_L if she does not need treatment and does not receive an overtreatment recommendation on her second opinion visit (which happens with probability $\frac{(1-\rho)x(1-x)}{\rho+(1-\rho)x}$). Using average overtreatment costs of CHF 535 for P_H, 0 for P_L, the overtreatment rate of 28% for x and average diagnosis costs of CHF 93 for k (abstracting thus for a start from other search/opportunity costs), searching for a second opinion in our case is worthwhile if $\rho < 0.47$. Assuming additional search/opportunity costs of CHF 50 on top of the additional diagnosis costs for a second opinion, this reduces ρ to 0.32.

Thus, our case illustrates that as long as both the likelihood of needing a treatment and opportunity costs are not very high, second opinions can be sensible. In many health care markets, health insurers are actually increasingly incentivizing second opinions.[52] Our results show that for cases for which both the likelihood of needing a treatment and opportunity costs are not very high, this might reduce overall costs. Furthermore, incentivizing second opinions might lead to a reduction in overtreatment rates and thus have an additional benefit.[53]

[52]In Switzerland, e.g., some insurers grant a discount of up to 15% if insurees search for a second opinion before undergoing surgeries such as artificial hip or knee joints or planned caesareans.

[53]In a neutrally framed credence goods laboratory experiment, Mimra et al. (2016b) show that the introduction of second opinions significantly reduces overtreatment rates.

6 Conclusion

We present the results from a field experiment in the market for dental care in Switzerland, for which a single test patient who did not need a treatment undertook 180 dentist visits. We provide direct evidence that overtreatment is an important phenomenon: The test patient received an overtreatment recommendation on more than every fourth visit. We also find a striking heterogeneity in treatment recommendations, with overtreatment recommendations ranging from one to six fillings at individual dentists and at least 13 different teeths recommended to be treated across the study, leading to treatment costs of on average CHF 535 (about $550) conditional on overtreatment. Using a comprehensive set of measures for market conditions, we also find that a shorter waiting time for the next possible appointment is associated with a significantly higher likelihood of receiving an overtreatment recommendation. Furthermore, in contrast to a large body of literature on physician-induced demand that relates physician density to mostly aggregate measures of health care consumption, we do not find a significant impact of dentist density on overtreatment recommendations. Our results indicate that physician density is not a good proxy for demand and treatment incentives at individual physicians; we provide a more direct measure of short-term demand relative to capacity with waiting time for the next possible appointment. In the experimental variation, we observe significantly less overtreatment recommendations for a patient with a higher socioeconomic status than a patient with a lower socioeconomic status under standard information. This difference disappears in the experimental condition in which the patient gives a signal of additional information from an online platform about the appropriate treatment to the dentist. These results highlight a complex role of the socioeconomic status for individual treatment decisions as well as interactions between the socioeconomic status and signalling information that requires further research. In particular, it is important to better understand the scope and limits of signalling information in credence goods markets in general and health care markets in particular.

References

Adler, N. E., Rehkopf, D. H., 2008. U.s. disparities in health: descriptions, causes, and mechanisms. Annu Rev Public Health 29, 235–252.

Ai, C., Norton, E. C., 2003. Interaction terms in logit and probit models. Economics Letters 80 (1), 123–129.

Arneill, A. B., Devlin, A. S., Dec. 2002. Perceived quality of care: The influence of the waiting room environment. Journal of Environmental Psychology 22 (4), 345–360.

Balafoutas, L., Beck, A., Kerschbamer, R., Sutter, M., 2013. What drives taxi drivers? a field experiment on fraud in a market for credence goods. The Review of Economic Studies 80 (3), 876–891.

Balafoutas, L., Kerschbamer, R., Sutter, M., 2017. Second-degree moral hazard in a real-world credence goods market. The Economic Journal 127 (599), 1–18.

Bundesamt für Gesundheit, 2015. Medizinalberufregister.

Clemens, J., Gottlieb, J. D., 2014. Do physicians' financial incentives affect medical treatment and patient health? American Economic Review 104 (4), 1320–1349.

Cromwell, J., Mitchell, J. B., 1986. Physician-induced demand for surgery. Journal of Health Economics 5 (4), 293–313.

Currie, J., Lin, W., Meng, J., 2014. Addressing antibiotic abuse in china: An experimental audit study. Journal of Development Economics 110, 39–51.

Currie, J., Lin, W., Zhang, W., 2011. Patient knowledge and antibiotic abuse: Evidence from an audit study in china. Journal of Health Economics 30 (5), 933–949.

Currie, J., MacLeod, W. B., Van Parys, J., 2016. Provider practice style and patient health outcomes: the case of heart attacks. Journal of health economics 47, 64–80.

Darby, M. R., Karni, E., 1973. Free competition and the optimal amount of fraud. Journal of Law and Economics 16 (1), 67–88.

Das, J., Hammer, J., May 2007. Money for nothing: The dire straits of medical practice in delhi, india. Journal of Development Economics 83 (1), 1–36.

Das, J., Holla, A., Mohpal, A., Muralidharan, K., December 2016. Quality and accountability in health care delivery: Audit-study evidence from primary care in india. American Economic Review 106 (12), 3765–99.

Domenighetti, G., Casabianca, A., Gutzwiller, F., Martinoli, S., 1993. Revisiting the most informed consumer of surgical services. the physician-patient. International Journal of Technology Assessment in Health Care 9, 505–513.

Doyle, J. J., Ewer, S. M., Wagner, T. H., 2010. Returns to physician human capital: Evidence from patients randomized to physician teams. Journal of health economics 29 (6), 866–882.

Dulleck, U., Kerschbamer, R., 2006. On doctor, mechanics, and computer specialists: The economics of credence goods. Journal of Economic Literature 44 (1), 5–42.

Dulleck, U., Kerschbamer, R., Sutter, M., 2011. The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. American Economic Review 101 (2), 526–555.

Emons, W., 1997. Credence goods and fraudulent experts. RAND Journal of Economics 28 (1), 107–119.

Emons, W., 2001. Credence goods monopolists. International Journal of Industrial Organization 19 (3-4), 375–389.

Emons, W., 2013. Incentive-compatible reimbursement schemes for physicians. Journal of Institutional and Theoretical Economics JITE 169 (4), 605–620.

Evans, R., 1974. Supplier-induced demand: Some empirical evidence and implications in the economics of health and medical care. London: Macmillian.

Fornara, F., Bonaiuto, M., Bonnes, M., Dec. 2006. Perceived hospital environment quality indicators: A study of orthopaedic units. Journal of Environmental Psychology 26 (4), 321–334.

Fuchs, V. R., 1978. The supply of surgeons and the demand for operations. Journal of Human Resources 13, 35–56.

Gruber, J., Owings, M., 1996. Physician financial incentives and cesarean section delivery. RAND Journal of Economics 27 (1), 99–123.

Grytten, J., Holst, D., Laake, P., 1990. Supplier inducement: its effect on dental services in norway. Journal of Health Economics 9 (4), 483–491.

Grytten, J., Sørensen, R., 2001. Type of contract and supplier-induced demand for primary physicians in norway. Journal of Health Economics 20 (3), 379–393.

Harrison, G. W., List, J. A., 2004. Field experiments. Journal of Economic Literature 42 (4), 1009–1055.

Huck, S., Lünser, G., Spitzer, F., Tyran, J.-R., 2016. Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment. Journal of Economic Behavior & Organization 131, 78–105.

Iizuka, T., 2007. Experts' agency problems: Evidence from the prescription drug market in japan. RAND Journal of Economics 38 (3), 844–862.

Imfeld, T., 2008. Zahnmedizin schweiz - weiterhin im dienst der oralen gesundheit? Schweizer Monatsschrift für Zahnmedizin 118 (10), 1016–1021.

Kerschbamer, R., Neururer, D., Sutter, M., 2016. Insurance coverage of customers induces dishonesty of sellers in markets for credence goods. Proceedings of the National Academy of Sciences 113 (27), 7454–7458.

Kerschbamer, R., Sutter, M., 2017. The economics of credence goods–a survey of recent lab and field experiments. CESifo Economic Studies 63 (1), 1–23.

Levitt, S. D., List, J. A., 2009. Field experiments in economics: The past, the present, and the future. European Economic Review 53 (1), 1–18.

List, J. A., 2007. Field experiments: a bridge between lab and naturally occurring data. The BE Journal of Economic Analysis & Policy 5 (2).

List, J. A., 2011. Why economists should conduct field experiments and 14 tips for pulling one off. Journal of Economic Perspectives 25 (3), 3–16.

List, J. A., Sadoff, S., Wagner, M., 2011. So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. Experimental Economics 14 (4), 439–457.

Léonard, C., Stordeur, S., Roberfroid, D., 2009. Association between physician density and health care consumption: A systematic review of the evidence. Health Policy 91 (2), 121–134.

Lu, F., 2014. Insurance coverage and agency problems in doctor prescriptions: Evidence from a field experiment in china. Journal of Development Economics 106, 156–167.

Marty, F. E., 1998. Capacity as a Determinant of the Supply for Physicians' Services. Universität Bern Abteilung für Angewandte Mikroökonomie.

McGuire, T. G., 2000. Physician agency. Vol. 1, Part A. Elsevier, book section 9, pp. 461–536.

Mimra, W., Rasch, A., Waibel, C., 2016a. Price competition and reputation in credence goods markets: Experimental evidence. Games and Economic Behavior 100, 337–352.

Mimra, W., Rasch, A., Waibel, C., 2016b. Second opinions in markets for expert services: Experimental evidence. Journal of Economic Behavior and Organization 131, Part B, 106–125.

Schneider, H. S., 2012. Agency problems and reputation in expert services: Evidence from auto repair. The Journal of Industrial Economics 60 (3), 406–433.

Schweizerische Zahnärzte-Gesellschaft, 2015a. Qualitätsleitlinien. Swiss Dental Journal.

Schweizerische Zahnärzte-Gesellschaft, 2015b. Zahnmedizin Schweiz: Berufe, Aus- und Weiterbildung, Standespolitik.
URL https://www.sso.ch/fileadmin/upload_sso/1_SSO/8_Berufsbilder/SSO_Zahnmed_dt_GzA.pdf

Schweizerische Zahnärzte-Gesellschaft, 2017. Zahnarzttarif.
URL https://www.sso.ch/fileadmin/upload_sso/3_Patienten/6_Recht_Tarif/Tarif_kurz_d.pdf

Staatssekretariat für Wirtschaft, 2004. Verordnung vom 11. Dezember 1978 über die Bekanntgabe von Preisen (PBV): Preisbekanntgabe für zahnärztliche Dienstleistungen.

Van Ryn, M., Burke, J., 2000. The effect of patient race and socio-economic status on physicians' perceptions of patients. Social science & medicine 50 (6), 813–828.

Zizzo, D. J., 2010. Experimenter demand effects in economic experiments. Experimental Economics 13 (1), 75–98.

A Appendix

A.1 Additional figures and photos

A.1.1 Figures

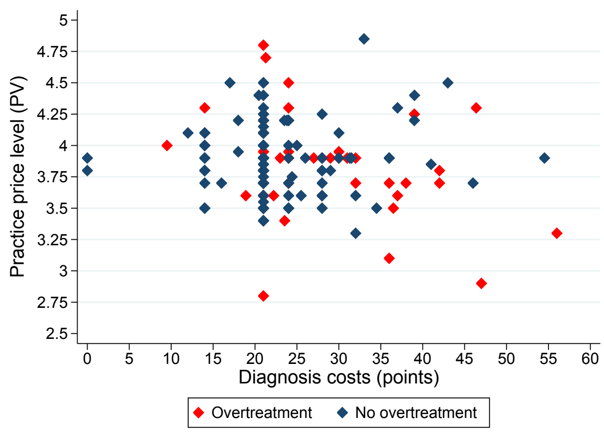


FIGURE A.1
Distribution of point values, diagnosis costs and overtreatment.

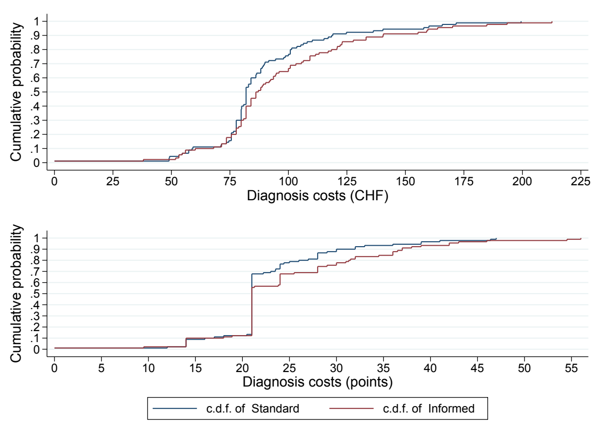


FIGURE A.2
Cumulative distribution of diagnosis fees and points charged by information status (n=180).

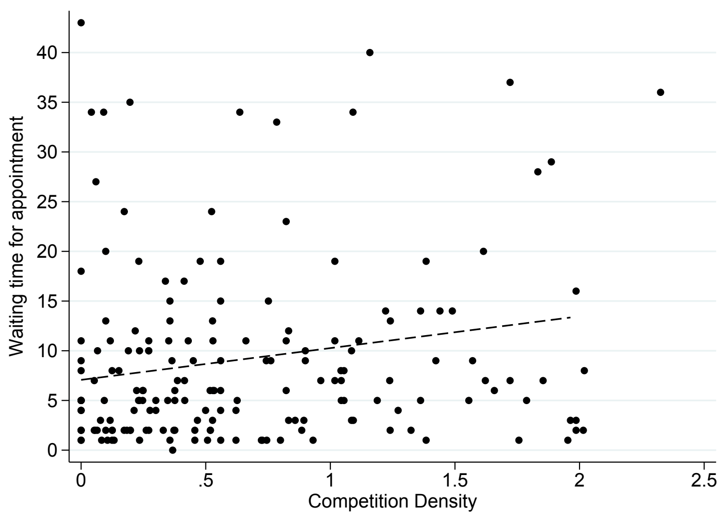


FIGURE A.3
Dentist density and waiting time in the full sample (n=180).

A.1.2 Experimental conditions

FIGURE A.4
Appearance of the test patient and accessory in the SES variation.

FIGURE A.5
Sreenshots of the mentioned web forum in the *informed* condition.

A.2 Population of dentists

The population list of dentists used in this field experiment comprised 865 entries. The list was based on the publicly available MedReg register issued by the Swiss Federal Office of Public Health. MedReg comprises all dentists in Switzerland with a valid working licence. The MedReg contains information on name, status and age of the working permission and approbation, the permission to sell pharmaceutical products[54], and some more characteristics. The register had 1'151 entries for the canton of Zurich as of October 2015. We deleted dentists with specializations such as child care, orthodontia and oral surgery as well as double entries. Moreover, we used information provided on the webpages of practices and in the yellow pages to update the database. Dentists who did not practice or who had retired were deleted and some, mostly young, dentists were added.

A.3 Detailed information on the recruitment and training process of the test patient

A.3.1 Recruitment

We searched for potential test patients through two channels. First, we advertised on the online-platform Marktplatz, run by the University of Zurich and the ETH Zurich. Second, we sent about 6'000 emails using the subject pool of the Department of Psychology at the University of Zurich. Our test patient was eventually recruited via the Marktplatz platform.[55] After telephone interviews with interested candidates, the most promising candidates were invited to visit one of our expert dentists together with one of the authors of this paper in order to check whether the candidate was suited for the study or not. The recruitment process continued with an assessment of the candidates' cognitive skills and reliability. Finally, we recruited a male person in his mid-twenties for the study in late 2015.

A.3.2 Training

After we had recruited our test patient, a detailed visiting script was developed for the test patient. The script was developed under consideration of the patient's real characteristics and histories in order to make the implementation of the roles as easy as possible. After the script had been developed, the test patient was trained accordingly. The dress for both SES

[54] In Switzerland, physicians can obtain the permission to sell drugs themselves.

[55] Marktplatz is a trading platform provided by the University of Zürich and the Swiss Federal Institute for Technology and can be reached at: http://www.marktplatz.uzh.ch/

roles was protocolled on a photograph to guarantee that they remained identical throughout the experiment. During the experiment, the test patient undertook weekly visits to our office to arrange dentist appointments and hand over the visit protocols to us. These visits were used to keep a check on the test patient's dresses. The dress for the high SES role mimics a banker's outfit and has been combined from a sales person in a classy department store in Zurich. At the time of the training sessions we also sent the test patient to five test visits. We did not use these visits for the statistical analysis in this study, but used them to make the test patient familiar with his roles and to test and improve the script. Incoming bills from all visits proof that all visits in the experiment and the medical check-up did indeed take place.

A.3.3 Effects over time

In order to eliminate time effects over the one-year period in which the visits were conducted, we implemented several procedures and checks. Each diagnosis was based on the same x-ray and the same oral condition of the test patient. Furthermore the test patient undertook test visits before the first documented visit in order to become familiar with the script and the procedures. Additionally, the test patient was provided with photographs and check lists that he was instructed to use before each visit. The test patient visited our offices on a weekly basis, such that we would have noticed changes in his appearance. Affirmatively, our results do not show any time effects. As one illustration, table A.1 shows that the inclusion of four dummies indicating the succession of visits, has no effects on the results.

A.4 Covariates and model fit

A.4.1 Balance of covariates

TABLE A.1

Regressions including dummies for 45 subsequent visits (reference category: visits 1-45).

	(M1)	(S1)	(M3)	(S3)
	Probit			
Dependent Variable	Overtreatment (binary)			
Treatments				
Information	-0.102	-0.108	-0.112	-0.120
	(0.214)	(0.190)	(0.134)	(0.113)
High SES	-0.174**	-0.176**	-0.173**	-0.173**
	(0.026)	(0.023)	(0.018)	(0.018)
Informed x High SES	0.181	0.181	0.143	0.144
	(0.225)	(0.233)	(0.297)	(0.310)
Demand & competition				
Waiting Time for Appointment (days)			-0.010**	-0.010**
			(0.014)	(0.013)
Competition Density			0.030	0.028
			(0.590)	(0.619)
Practice price level (PV)			-0.026	-0.032
			(0.808)	(0.768)
Practice & dentist variables				
Price level (PV) displayed			-0.170***	-0.161***
			(0.001)	(0.003)
Swiss Licence Age (years)			-0.003	-0.004
			(0.530)	(0.493)
Informative Webpage			0.281*	0.274
			(0.095)	(0.109)
Inform. Webpage x Swiss Licence Age			-0.013**	-0.013**
			(0.039)	(0.046)
Practice Owner			0.164*	0.159*
			(0.085)	(0.099)
Other variables				
Median Income in Area (cont.)			0.007	0.006
			(0.135)	(0.160)
Visit succession				
Visits 46-90		-0.009		0.016
		(0.925)		(0.862)
Visits 91-135		-0.068		-0.054
		(0.423)		(0.515)
Visits 135-180		-0.044		0.002
		(0.610)		(0.979)
Pseudo R^2	0.017	0.020	0.160	0.163
N	180	180	180	180

Probit regressions, displaying average marginal effects

p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A.2

p-values of pairwise comparisons of means.

	Price level displayed		
Treatment	ST-LS	INFO-LS	ST-HS
INFO-LS	0.416		
ST-HS	0.515	0.282	
INFO-HS	0.666	0.392	1.000

Test: Fisher's exact (2-sided)

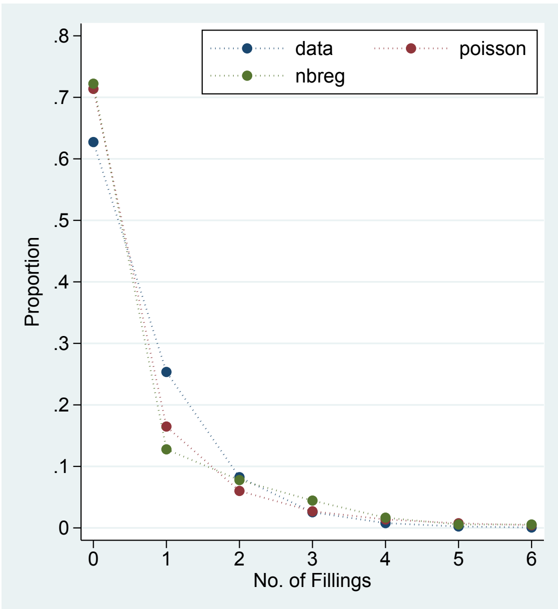
TABLE A.3

p-values of pairwise comparisons of means.

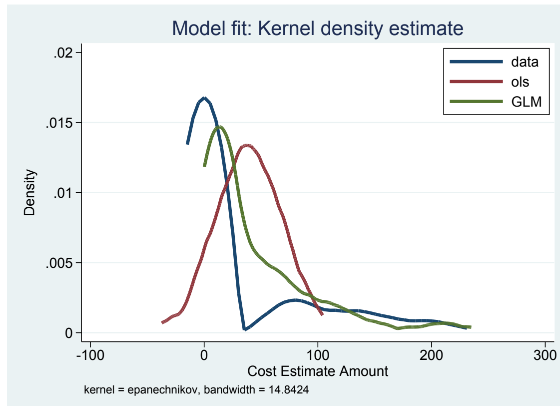
	Waiting time for appointment		
Treatment	ST-LS	INFO-LS	ST-HS
INFO-LS	0.987		
ST-HS	0.183	0.181	
INFO-HS	0.984	0.935	0.145

Test: Mann-Whitney (2-sided)

A.4.2 Model fit



(a) Comparison of the poisson and negative binomial model predicting the number of recommended fillings.



(b) Comparison of the model fit of OLS and GLM estimations predicting the cost estimate amount.

FIGURE A.6
Model fit.