Implementation of altered provider incentives for a more individual-risk-based assignment of dental recall intervals
evidence from a health systems reform in Denmark

Gabel, Frank; Kalmus, Olivier; Rosing, Kasper; Trescher, Anna-Lena; Listl, Stefan

Published in:
Health Economics

DOI:
10.1002/hec.3997

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Implementation of altered provider incentives for a more individual-risk-based assignment of dental recall intervals: evidence from a health systems reform in Denmark

Frank Gabel | Olivier Kalmus | Kasper Rosing | Anna-Lena Trescher | Stefan Listl

1Department of Conservative Dentistry - Section for Translational Health Economics, Heidelberg University Hospital, Heidelberg, Germany
2Department of Dentistry - Quality and Safety of Oral Healthcare, Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands
3Department of Odontology, Section 1, Community Dentistry, University of Copenhagen, Denmark

Correspondence
Stefan Listl. Department of Dentistry - Quality and Safety of Oral Healthcare, Radboud University Medical Center, Radboud Institute for Health Sciences, Philips van Leydenlaan 25, 6525 EX, Nijmegen, The Netherlands.
Email: stefan.listl@radboudumc.nl

Funding information
European Commission’s Horizon 2020 research and innovation program, Grant/Award Number: 635183

ABSTRACT

Equipping health systems with suitable incentives for efficient resource allocation remains a major health policy challenge. This study examines the impacts of 2015 regulatory changes in Danish dental care which aimed at effectuating a transition from six-to-twelve-monthly dental recall intervals, for every patient, towards a model where patients with higher need receive dental recalls systematically more frequently than patients with lower need. Exploiting administrative data from the years 2012-2016 from the Danish National Health Insurance database containing 72,155,539 treatment claims for 3,759,721 unique patients, we estimated a series of interrupted time-series regression models with patient-level fixed-effects. In comparison to the pre-reform period, the proportion of patients with recall intervals of up to 6 months was by 1.2%-points larger post-implementation; that of patients with 6-12-monthly recalls increased by 0.7%-points; that of patients with more than 12-monthly dental recalls decreased by 1.9%-points. The composition of care shifted more substantially: the proportion of treatment sessions including preventive care increased by 31.5%-points (95%-CI: 31.4;31.6); that of sessions including scaling increased by 24.1%-points (24.0;24.2); that of sessions including diagnostics decreased by 34.5%-points (34.4;34.6). These findings suggest that dental care providers may have responded differently to regulatory changes than intended by the health policy.

KEYWORDS
Public Health, Public Health Systems Research, Dental Care, Risk-based care, Denmark

1 INTRODUCTION

Implementing health system and practice changes through incentives is complex. Health care payment reform has been described as a social and learning process for the stakeholders involved: in order to be successful, it has been suggested that payment reforms need to be based on collaboration and coordination between payers and providers and not be undermined by competitive relationships between stakeholders (Conrad et al. 2016). Embedding changes in payment...
systems, such that the incentives for the various stakeholders involved are aligned in a sensible way, is not an easy task (Chaudoir, Dugan, & Barr, 2013). In particular, given the nature of the doctor-patient relationship in terms of information asymmetry and supplier-induced demand (Arrow, 1963; McGuire, 2000), altered incentive structures might result in unforeseen changes in health care utilization and health outcomes.

More specifically, previous work by Grembowski and colleagues suggests that various factors can foster or hinder successful health system transformation (Conrad, Vaughn, Grembowski, & Marcus-Smith, 2016; Grembowski & Marcus-Smith, 2018). The described factors include social conditions such as: the engagement of committed leaders (“champions”) to encourage diverse stakeholders to participate in payment and system reforms; a common stakeholder-consented vision of future reforms; a culture of collaboration and cooperation; social capital (including norms of reciprocity, trust in others, civic participation); and collective efficacy, that is shared beliefs in a group's collective ability to produce desired results. In addition, support factors include: sufficient monetary resources for system transformation; a reliable data infrastructure with harmonized performance indicators for planning and monitoring purposes; and a legal infrastructure in support of payment reform (Conrad et al., 2016; Grembowski & Marcus-Smith, 2018).

A recent example from pay-for-performance implementation within dental care highlights substantial complexity in dental care system transition such as: the customization of adequate incentives for various types of providers, securing information systems to provide sensible performance metrics, and obtaining commitment among dental care providers (Conrad et al., 2018). The existing literature on supplier-induced demand in dentistry provides further evidence for possible complexities involved when seeking to implement changes within dental care (see e.g. Birch, 1988; Grembowski & Milgrom 1988; Chalkley & Listl, 2018; Gottschalk, Mimra, & Waibel, 2018).

Health care is an expensive undertaking, with most high-income countries spending from 10 to 14 percent of their GDP on health care and costs increasing rapidly (Dieleman et al., 2016; OECD, 2019). Dental diseases account for about 5% of health spending (Listl, Galloway, Mossey, & Marcenes, 2015). In order to allocate (oral) health care resources in a rational way, it is sensible to make careful efforts to improve the alignment of supply and need. In the context of preventing and managing chronic diseases such as dental caries, a particularly appealing approach is to introduce personalized recall intervals according to a patient's current disease level and individual risk of future disease (Bader, 2005; Darcey & Ashley, 2011). In terms of the order of magnitude of costs, dental check-ups were estimated to account for more than 10% of dental expenditures (Listl & Chalkley, 2014). The principle underpinning the risk-based recall approach is that extending recall intervals for those individuals classified as 'low risk' should not incur any undue detrimental effects on their oral health status, reduce their overall resource consumption and reduce risk of overdiagnosis and overtreatment. More frequent recalls can then be adopted for those individuals with the greatest need who are classified as 'high risk'. Improving the alignment of supply and need provides potential efficiency gains based on promoting the most beneficial type and amount of health care use given available resources.

While determining the length of dental recall intervals according to individual patients' needs could be a promising alternative to fixed recall intervals, the optimal length of the time interval between two diagnostic dental examinations has given rise to intensive debate (Beirne, Clarkson, & Worthington, 2007; Sheiham, 2000). On the one hand, it has been argued that routine dental check-ups can have a positive impact on perceived quality of life and in terms of reducing the incidence of dental diseases (Afonso-Souza et al., 2007). On the other hand, the tradition of prompting all patients to attend for regular examinations every six months (Hahn, Kraus, & Hooper-Lane, 2017) or yearly, has been challenged. It has been suggested that there is no "one-size-fits-all" approach to dental recall intervals and patients may benefit from strategies revolving around tailored prevention (Tomar, 2011). Along similar lines, it has also been hypothesized that by lengthening or individualizing dental examination recall intervals, access to dental care could be widened to a greater number of people (McGuire, 2000). A Finnish expert group previously recommended prolonging the average oral health examination interval for children and adolescents to 1.5-2.0 years, taking into account the risk of each individual; yet this was not implemented via official guidelines (Lahti, Hausen, Widstr, & Eerola, 2001).

As it stands, however, the evidence on potential implications of various recall practices - including in terms of equity of access and health care costs - remains unclear. Findings from a previous health technology assessment suggest that less frequent dental check-ups are associated with reduced treatment and little evidence of an adverse impact on dental health (Davenport et al., 2003). Beirne et al. (2007) found no high-quality RCTs supporting any kind of fixed recall interval. Previously, there has been only one RCT which investigated different recall intervals from a local field project in Norway, which suggested both significant improvements in children's dental health and efficiency gains; thereby, risk was assessed using qualitative clinical judgment (Wang & Holst, 1995). More recently, a randomized controlled trial has been set up in the UK to investigate the risk-adjusted assignment of recall intervals according to NICE...
guidelines (Clarkson et al., 2018). Similar (ongoing) research substantiates the relevance of addressing the question of whether a fixed or an individual-risk-based recall policy would be more appropriate. According to the most recent Cochrane review on recall intervals for oral health in primary care patients, the evidence to support or refute the practice of encouraging patients to attend for dental check-ups at six-monthly intervals remains unclear (Riley et al. 2013). A similar state of knowledge applies to periodontal recalls (Farooqi, Wehler, Gibson, Jurasic, & Jones, 2015).

To the best knowledge of the authors, there have not yet been larger-scale implementations of dental remuneration systems aligned with individual disease risk and the present study is the first to scrutinize the effects of introducing a risk-based remuneration system in dentistry for an entire country’s population. Using administrative data from the Danish Health Authority, consisting of about 72 million treatment items allowed the authors to model changes in utilization patterns in detail and examine ramifications related to physician agency and demand induced by suppliers of dental services. As the remuneration regulations were dependent on patients’ oral disease risk, it was hypothesized that there would be shifts in the extent and composition of treatment sessions following the reform.

2  |  METHODS

2.1  |  Institutional background

In Denmark, dental care for adults is usually provided by private dental practitioners. Dental care expenses are partly covered by self-payment and from general taxation financed payments from the National Health Insurance. All adult citizens are eligible for compensation. For persons under the age of 18, dental care is provided in public dental clinics financed by general taxation and without additional out-of-pocket expenses (Danish Health Act, 2018). According to WHO criteria, the Scandinavian countries belong to the so-called very low and low-caries prevalence countries (Petersen, 2003). The use of dental services is comparatively high in these countries, with 64% of the adult population and 77% of the population under 18 years of age visiting a dentist at least once a year (Nihtilä, 2010). Dentists in Denmark are paid using the fee-for-service payment model in which each item of treatment is paid for separately, giving an incentive for dentists to provide more treatments because payment is dependent on the quantity, rather than quality of care.

2.2  |  The 2015 reform

In 2013, The Danish Health Authority issued new guidelines for dental recall intervals. From April 1, 2015, a new collective agreement was negotiated between the Danish Regions and the Danish Dental Association, incorporating the 2013 guidelines (Regionernes Lønningsog Takstnævn, 2014). The collective agreement describes the dental services delivered in adult dental care and sets the level of remuneration paid from the Danish National Health Insurance. In this paper, this is designated as the “2015 reform”.

Since then, dentists have been required to risk-classify their patients into three distinct classes according to their current oral health status and the assessed risk of future oral disease. Healthy patients (free from active oral disease and free from risk factors for future oral disease) should be categorized as “green”, at-risk patients (active oral-disease and/or presence of risk factors for oral disease which are modifiable, for instance poor oral hygiene) should be categorized as “yellow” and high-risk patients (active oral disease and/or risk factors for oral disease, which are not modifiable, for instance chronic general disease with known influence on oral health) should be categorized as “red”. The recommended dental recall intervals vary across these risk-groups. Patients categorized as either “yellow” or “red” are advised to attend for check-ups more frequently while healthy patients are incentivized to attend for check-ups less frequently (Figure 1). Additionally, in part, the risk classification determines which treatments can be remunerated. Most notably, remunerating “Individual Preventive Treatment (IPT)” in diagnostic check-ups is now restricted to patients characterized as either yellow or red. Also, claiming remuneration for newly created codes concerning “focused examination (FE)” is only possible if patients are classified as yellow or red. This way, at-risk patients should both undergo a more thorough treatment (through IPT) and visit the dentist more frequently (through FE). The “Status Examination (SE)” is to be performed regularly (every 12-24 months) for all patients. Further details of the Danish treatment approach are shown in Figure 1 and in Table 1. From the dentists’ perspectives, the reform was anticipated as likely to result in reductions in revenues from treating patients in the low risk group (green category) but in increases
FIGURE 1  Treatment pathways according to risk grouping. Patients can switch between risk groups when their disease status changes. The dental recall intervals are now 12-24 months in all tracks with additional sessions pertaining to yellow and red risk groups [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1  Description of treatment items exclusively available to at-risk patients (IPT, FE) and all patients in regular intervals (SE) following the 2015 reform

**Individual Preventive Treatment (IPT - treatment code 2920)**
IPT can be used in the presence of active caries, gingivitis, mucositis around implants, marginal periodontitis and periimplantitis or other oral disorders that require preventive treatment.
In order to perform this service, patients need to be classified as being in either the yellow or red risk group.

**Status examination (SE - treatment code 1114/1115)**
The SE service forms the basis for the planning of necessary preventive efforts and treatment activities until the next SE or FE service can be carried out. It involves determining the interval until the next examination based on risk assessment and the individual need of a patient as well as risk categorizations into green, yellow, or red categories.

**Diagnostic Examination (DE - treatment codes 1111/1112/1113)**
A diagnostic examination is a service carried out for new patients as a basis for diagnosis and treatment planning of oral disorders. It mainly contains an in-depth assessment of oral health status and the identification of risk factors. The DE involves risk categorizations into green, yellow, or red categories.

**Focused examination (FE - treatment code 1116)**
A focused examination is a follow-up examination focused on a current disease problem. It also includes updates on treatment planning regarding this disease problem and, accordingly, future recall interval lengths.
in earnings from treating patients in higher risk groups for whom provision of preventive care (IPT) during follow-up examinations became mandatory. Yet dentists who exceeded a maximum limit of health insurance reimbursements were also subject to restitution of payments exceeding the respective threshold.

2.3 Data

The authors have obtained unique administrative data with patient-level information on the services provided by dental practitioners in Denmark over a 5-year period from 2012 to 2016. The data comprise all treatment claims achieved in the Danish National Health Insurance database between 2012 and 2016. Usable variables included the treatment performed, patient age and sex, date of treatment, municipality of both patient and dental practice as well as cost of treatment. In total, 72,155,539 claims from a total of 3,759,721 unique patients in 25,533,311 distinct treatment sessions were investigated. A single observation was formed by a claim handed into the health insurance by a dentist. The raw data presented as very homogenous and were not indicative of missing data. We performed several sense checks to exclude observations with typing errors or missing comas. A detailed description of the raw dataset and variables can be found in the Appendix (Table A2). The data used were pseudonymized in accordance with Danish jurisdiction and no ethical clearance was required for purposes of this research project.

2.4 Analytical approach

In order to analyze the impacts of the reform, we compared the utilization of dental services thought to be potentially affected by the reform. The dental services under scrutiny were divided into several treatment baskets: preventive (IPT, see Figure 1), diagnostic (FE, SE and DE see Figure 1), scaling, X-ray, periodontal and fillings. For each of these categories, a binary variable was set to 1 for a particular session (one session consists of several treatments, all performed within a day) if at least one corresponding code was remunerated in that session. Otherwise, it was set to 0.

To identify the impact of the 2015 reform, we compared the utilization of dental services before and after the introduction of the reform using an interrupted time series design (Lopez Bernal, Cummins, & Gasparrini, 2016) at session level (Ordinary Least Squares (OLS) model with binary treatment variables). To achieve a better view of the effects of implementing reforms using individual risk classes, we performed multiple interrupted time series analyses on utilization of different treatment patterns, which is a valid way of evaluating large-scale implementations in analyzing data known or thought to be affected by interventions (Bauer, Damschroder, Hagedorn, Smith, & Kilbourne, 2015).

The search for a suitable set of confounders explaining residual variance in our model was primarily subject to data availability. Our dataset is based on reimbursement claims which were submitted to a health insurance and our set of available variables to correct for (potential) confounding was limited. In addition, if patient characteristics are not fully contained in the available explanatory variables, OLS regression models may be biased. In order to deal with potential problems of this kind, we exploited the longitudinal character of our data and included patient-level fixed effects, giving us the most complete control for unobserved heterogeneity. Taken all together, aiming at a sensible identification strategy and employing backward elimination of variables, the model specification described below was devised. A more detailed description of the dataset, corresponding transformations and included/excluded variables can be found in the Appendix (Table A2).

For our regressions, we used the following equations:

\[ y_t = \beta_0 + \beta_1 \text{time}_t + \beta_2 \text{gender} + \beta_3 \text{age} + \beta_4 \text{municipality} + \beta_5 \text{reform} + \beta_6 \text{reform \times time}_t + \epsilon \quad \text{(OLS)} \]

\[ y_{it} = \alpha_i + \beta \text{reform}_{it} + \beta \text{reform \times time}_{it} + \epsilon_{it} \quad \text{Fixed Effects} \]

In this equation, \( y \) denotes the binary outcome variable at hand at time point \( t \) (within an individual patient \( i \)), i.e. whether a particular session contained a treatment code of interest or not. The variables “gender” and “municipality” constitute categorical control variables, the variable “time” (i.e. month of treatment) and “age” a control variable assumed to influence the outcome linearly and the binary variable reform is 1 if an observation pertains to the
period after the reform and 0 otherwise. The variable reform x time captures possible changes in trend after the reform was put into place.

2.5 Robustness Checks

In the main analyses, we restricted the period under investigation to one year before and after the reform, excluding yearly seasonal trends. As a robustness check, we chose to exclude the year 2015 from our analyses altogether and subsequently compared utilization in the 19 months before and after the reform year 2015, excluding anticipatory effects.

Ideally, detailed information on patients’ dental risk categories would be available. However, as such information was unavailable, we carried out auxiliary analyses using known correlates of dental disease risks as proxy variables. First, we applied socio-economic status (constructed according to information on the average income of the municipality in which the patient lives (Statistikbanken.dk, 2017); we assumed that patients usually seek dental care close to where they are living) as proxy for dental disease risk; existing evidence suggests that persons with higher socioeconomic status tend to have lower dental disease risks than persons with lower socioeconomic status (Peres et al., 2019). Second, considering that the cumulative nature of tissue destruction in caries and periodontitis has been reported to result in higher severity and extent at older ages, we used “age” as a further risk proxy (López, Smith, Göstemeyer, & Schwendicke, 2017). Third, persons who had previously received fillings or periodontal treatment were considered a “proxied high-risk group” and persons without fillings or periodontal treatment could be considered a “proxied lower risk group”.

3 RESULTS

Table 2 shows summary statistics for dependent and explanatory variables. The mean age of patients was 49.4 years. Overall, in most sessions, scaling and diagnostic codes were utilized. Fillings and periodontal treatments were performed in 32% and 22% of sessions, respectively, while the preventive code was used in about 15% of all sessions. Radiographs were taken in about every 6th session.

Table 3 shows descriptive statistics regarding changes in dental utilization and recall interval characteristics before and after the reform. The average number of dentist visits and recalls per patient was shown to have increased slightly. The average number of days between dental visits per patient reduced by about seven days and that of dental recalls reduced by about six days. The proportion of patients visiting the dentist every 6 months or more often increased by about 0.9%, whereas the proportion of patients with 6-12-monthly or more than 12-monthly dental visits decreased somewhat. In comparison to the pre-reform period, the proportion of patients with recall intervals of up to 6 months was by 1.2%-points larger post-implementation; that of patients with 6-12-monthly recalls increased by 0.7%-points; that of patients with more than 12-monthly dental recalls decreased by 1.9%-points.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Summary statistics of dependent and independent variables of our total sample (averages across treatment sessions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample size: 25,533,311 treatment claims</td>
<td>Description</td>
</tr>
<tr>
<td>age in years</td>
<td>49.4 (16.9)</td>
</tr>
<tr>
<td>sex Women/men in percent</td>
<td>53.8/46.2</td>
</tr>
<tr>
<td>total number of sessions mean number of yearly sessions</td>
<td>2.41 (0.74)</td>
</tr>
<tr>
<td>preventive (IPT) equals 1 if a preventive code was remunerated in a particular session</td>
<td>0.15</td>
</tr>
<tr>
<td>diagnostic (SE, DE, EDE, FE) equals 1 if a diagnostic code (with FE) was remunerated in a particular session</td>
<td>0.75</td>
</tr>
<tr>
<td>scaling equals 1 if scaling was remunerated in a particular session</td>
<td>0.58</td>
</tr>
<tr>
<td>fillings equals 1 if fillings were remunerated in a particular session</td>
<td>0.32</td>
</tr>
<tr>
<td>periodontal equals 1 if periodontal treatments were remunerated in a particular session</td>
<td>0.22</td>
</tr>
<tr>
<td>radiographs equals 1 if radiographs were remunerated in a particular session</td>
<td>0.15</td>
</tr>
<tr>
<td>surgical treatments equals 1 if an operation or tooth extraction was performed in a particular session</td>
<td>0.07</td>
</tr>
</tbody>
</table>

A detailed description of the presented variables can be found in the Appendix (Table A2).
Table 3 Summary statistics of the frequency of dental visits recalls before and after the reform

<table>
<thead>
<tr>
<th>sample size:</th>
<th>3,759,721 unique patients</th>
<th>before the 2015 reform</th>
<th>after the 2015 reform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of dental visits per patient (per year)</td>
<td>1.83</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>Average number of dental recalls* per patient (per year)</td>
<td>1.29</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>Average number of days between dental visits (per patient)</td>
<td>148.5</td>
<td>141.6</td>
<td></td>
</tr>
<tr>
<td>Average number of days between dental recalls* (per patient)</td>
<td>230.6</td>
<td>225.3</td>
<td></td>
</tr>
<tr>
<td>% of patients with a 6-month interval between dental visits (or more frequently)</td>
<td>20.4%</td>
<td>21.3%</td>
<td></td>
</tr>
<tr>
<td>% of patients with a 6-month interval between dental recalls* (or more frequently)</td>
<td>16.3%</td>
<td>17.5%</td>
<td></td>
</tr>
<tr>
<td>% of patients with a 6-12-month interval between dental visits</td>
<td>33.4%</td>
<td>32.6%</td>
<td></td>
</tr>
<tr>
<td>% of patients with a 6-12-month interval between dental recalls*</td>
<td>36.1%</td>
<td>36.8%</td>
<td></td>
</tr>
<tr>
<td>% of patients with a more than 12-month interval between dental visits</td>
<td>46.2%</td>
<td>46.1%</td>
<td></td>
</tr>
<tr>
<td>% of patients with a more than 12-month interval between dental recalls*</td>
<td>47.6%</td>
<td>45.7%</td>
<td></td>
</tr>
</tbody>
</table>

*dental recalls were defined as treatment sessions which only included only SE and/or DE (see Figure 1). Patients with 0 visits to the dentist were not considered as their recall intervals can’t be recovered with claims-data based observations.

Figure 2 illustrates the trajectories of treatment codes over time for the years 2012-2016. The proportion of treatment sessions including preventive items or scaling is shown to have become larger after the regulatory changes in April 2015; the proportion of diagnostic items was shown to have become lower.

Table 4 displays results from regression analyses on the utilization of certain treatment codes represented by independent, mostly binary variables (except for “total number of sessions” which is a count variable) following the 2015 reform. The second column contains point estimates from OLS regressions. A total of 5,420,552 sessions were included in this regression. The total number of sessions did only change marginally (point estimate: -0.003 [95%-Confidence_Interval: -0.005; -0.002] as compared to the pre-reform baseline). The proportion of sessions containing preventive codes and the proportion of sessions with scaling increased by 0.301 [0.300, 0.302] %-points and 0.225 [0.224; 0.226] %-points, respectively. At the same time, fewer sessions contained diagnostic codes (-0.298 [-0.299; 0.297] %-points).

The two columns on the right of Table 4 show parameter estimates from fixed-effects regressions when considering both the 12 months before and after the reform (third column) and, as a robustness check, both 2014 and 2016 excluding 2015 (fourth column). In the third column (12 months pre/after reform), the results indicate an increase in the

![Figure 2](https://example.com/figure2.png) Trajectories of the proportion of treatment claims containing codes for preventive care, diagnostic care, scaling, and filling treatment sessions taking place in the period of 2012 to 2017. The vertical gray line depicts April 1, 2015, when the reform commenced [Colour figure can be viewed at wileyonlinelibrary.com]
proportion of sessions including the preventive code by about a third (0.310 [95%-Confidence-Interval: 0.309; 0.311] %-points). The proportion of diagnostic sessions decreased by 0.345 [-0.346; -0.344] %-points and scaling experienced a sizable increase as well (0.241 [0.240; 0.242] %-points). Treatment codes apart from preventive and diagnostic codes and scaling did not exhibit large variations: fillings slightly decreased while the number of radiographs remained relatively stable. Results in the fourth column (19 months pre-/post-reform) are similar, with exceptions being different signs for the total number of sessions and periodontal treatment (relative to preventive, diagnostic and scaling codes, however, these effect sizes are much smaller). By and large, our results are robust to changes in the observation periods and the general tendency of variations in utilization remains.

When splitting the sample according to various proxies for dental disease risk, the results were mixed (Table 5). Parameter estimates differed relatively little when differentiating between persons with high vs. low income although effects for codes related to preventive and diagnostic services were stronger for the high-risk group. There was more variation between parameter estimates when differentiating between persons in young vs. old age; again, the estimated effects of regulatory changes on utilization of preventive, diagnostic and scaling items are larger for persons with higher risk (older age) than for persons with lower risk (younger age) but all have the same sign (more prevention and scaling but less diagnostics after regulatory changes). When using previous treatment experience as dental disease risk proxy, the most substantial differences in parameter estimates were found for the total number of sessions and periodontal treatment (relative to preventive, diagnostic and scaling codes, however, these effect sizes are much smaller). By and large, our results are robust to changes in the observation periods and the general tendency of variations in utilization remains.

When splitting the sample according to various proxies for dental disease risk, the results were mixed (Table 5). Parameter estimates differed relatively little when differentiating between persons with high vs. low income although effects for codes related to preventive and diagnostic services were stronger for the high-risk group. There was more variation between parameter estimates when differentiating between persons in young vs. old age; again, the estimated effects of regulatory changes on utilization of preventive, diagnostic and scaling items are larger for persons with higher risk (older age) than for persons with lower risk (younger age) but all have the same sign (more prevention and scaling but less diagnostics after regulatory changes). When using previous treatment experience as dental disease risk proxy, the most substantial differences in parameter estimates were found for the total number of sessions and periodontal treatment (after regulatory changes: fewer sessions for high risk patients; more sessions for low risk patients), scaling (after regulatory changes: nearly four-fold more sessions with scaling for low-risk than for high-risk patients), and fillings (after regulatory changes: fewer fillings for high-risk patients but more fillings for low-risk patients). By and large, the robustness checks neither provide clear evidence in support of, or against treatment patterns having become more risk-oriented in response to regulatory changes.

| TABLE 4 | OLS and Fixed Effects regression results for effects of the 2015 reform |
|-----------------|-----------------|-----------------|-----------------|
|                 | OLS             | Fixed Effects (Patient Level) | Fixed Effects (Patient Level) |
|                 |                 | 12 months pre. vs. 12 months post reform | 19 months pre. vs. 19 months post reform (excluding 2015) |
| total number of sessions | -0.003 [-0.005; -0.002] | -0.012 [-0.014; -0.010] | 0.018 [0.016; 0.020] |
| preventive (IPT) | 0.301 [0.300, 0.302] | 0.310 [0.309, 0.311] | 0.315 [0.314; 0.316] |
| diagnostic (SE, DE, EDE, FE) | -0.298 [-0.299; 0.297] | -0.345 [-0.346; -0.344] | -0.364 [-0.365; -0.363] |
| scaling | 0.225 [0.224; 0.226] | 0.241 [0.240; 0.242] | 0.282 [0.281; 0.283] |
| fillings | -0.029 [-0.030; -0.028] | -0.041 [-0.042; -0.039] | -0.041 [-0.043; -0.040] |
| periodontal treatment | -0.005 [-0.006; -0.004] | -0.032 [-0.033; -0.031] | 0.008 [0.007; 0.009] |
| radiographs | -0.008 [-0.010; -0.007] | -0.041 [-0.042; -0.040] | -0.016 [-0.017; -0.015] |
| surgical treatments | -0.003 [-0.004; -0.002] | -0.012 [-0.013; -0.011] | -0.018 [-0.020; -0.017] |
| Patient fixed effects | NO | YES | YES |
| N (observations = sessions) | 5,420,552 | 3,181,824 | 3,259,848 |

OLS (second column) and linear individual fixed effects regression (third and fourth column) using the number of sessions containing preventive, diagnostic, scaling, fillings and periodontal codes as well as codes related to surgical procedures and extractions (“surgical treatments”) resp. as independent variables. In our OLS model, we used age, sex and municipality of patients as confounders. In our Fixed Effects model, we did not use additional controls as no time-varying confounders could be identified (we tried to control for age, but this did not add explanatory power). 95% Confidence Intervals in brackets.

4 | DISCUSSION

Our findings indicate significant and quantitatively large shifts in treatment patterns in response to 2015 regulatory changes to provider incentives in Denmark. By having dentists classify patients into three distinct risk groups, these changes were intended to effect a transition from six-to-twelve-monthly dental recall intervals for every patient towards a more patient-centered model in which patients with higher need should receive dental recalls systematically more frequently than patients with lower need.

In comparison to the pre-reform period, our findings suggest that the proportion of patients with dental recalls every 6 months or more often increased by 1.2%-points, the proportion of patients with 6-12-monthly recalls increased by about 0.7%-points and the proportion of patients with more than 12-monthly dental recalls decreased by about 1.9%-
### TABLE 5  Robustness check: sub-Group Fixed Effects regression results of interrupted time-series modeling on utilization of treatment groups

<table>
<thead>
<tr>
<th></th>
<th>proxy: income (median split @ income)</th>
<th>proxy: age (median split @ age 51)</th>
<th>proxy: treatment experience (0 treatments = low risk)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 3,181,824 (observations = sessions)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total number of sessions</td>
<td>high risk [0.016, 0.020]</td>
<td>low risk [0.015, 0.021]</td>
<td>high risk [-0.021, -0.018]</td>
</tr>
<tr>
<td></td>
<td>high risk [0.019, 0.024]</td>
<td>low risk [0.019, 0.023]</td>
<td>low risk [0.019, 0.022]</td>
</tr>
<tr>
<td>Preventive</td>
<td>high risk [0.321, 0.324]</td>
<td>low risk [0.304, 0.308]</td>
<td>high risk [0.322, 0.324]</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>high risk [-0.329, -0.327]</td>
<td>low risk [-0.388, -0.384]</td>
<td>high risk [-0.379, -0.377]</td>
</tr>
<tr>
<td>Scaling</td>
<td>high risk [0.271, 0.273]</td>
<td>low risk [0.302, 0.306]</td>
<td>high risk [0.326, 0.328]</td>
</tr>
<tr>
<td>Fillings</td>
<td>high risk [-0.043, -0.039]</td>
<td>low risk [-0.045, -0.040]</td>
<td>high risk [-0.091, -0.088]</td>
</tr>
<tr>
<td>Periodontal treatment</td>
<td>high risk [-0.016, -0.013]</td>
<td>low risk [-0.019, -0.016]</td>
<td>high risk [0.125, 0.129]</td>
</tr>
<tr>
<td>Radiographs</td>
<td>high risk [0.005, 0.008]</td>
<td>low risk [0.009, 0.094]</td>
<td>high risk [-0.003, -0.001]</td>
</tr>
<tr>
<td>Surgical treatments</td>
<td>high risk [-0.06, -0.000]</td>
<td>low risk [0.021, 0.024]</td>
<td>high risk [0.017, 0.024]</td>
</tr>
</tbody>
</table>

Linear individual fixed effects regression on subgroups (exposed to different proxy variables for dental disease risk) using the number of sessions containing preventive, diagnostic, scaling, fillings and periodontal codes as well as codes related to surgical procedures and extractions ("surgical treatments") resp. as independent variables. In our model, we did not use additional controls as no time-varying confounders could be identified (we tried to control for age, but this did not add explanatory power). 95% Confidence Intervals in brackets.
points. While this distribution of recall intervals changed only to a relatively small extent, the composition of utilized care items shifted substantially. For the time period following regulatory changes, we observed substantial increases in preventive services and scaling as well as a substantial decrease in diagnostic services. Given the comparatively low dental disease burden in Denmark, these findings may appear to be somewhat against expectations.

In international comparisons, the Danish population has comparatively good oral health. According to WHO criteria, the Scandinavian countries belong to the so-called very low and low-caries prevalence countries (Petersen, 2003; Silveira Moreira, 2012), but there is still room for improvement due to disparities related to social inequalities (Rosing, 2015). However, in a recent review, the Danish Health Authority reported that following the 2015 reform, only a minority of Danish patients were classified as having a low oral disease risk (Sundhedsstyrelsen, 2016). It was remarkable that - despite good oral health - the majority of examined individuals were categorized as belonging to the at-risk groups. However, following regulatory changes, most patients (79%) were classified as either being at-risk or high-risk (Sundhedsstyrelsen, 2017). Following the reform, there were many payments for the FE-code (Table 1) pertaining to patients with active disease status only. This code also made it mandatory to perform IPT (unlike before the reform, where preventive treatment was to be performed only when found to be necessary by the dentist), contributing to the steep rise in the usage of this code. Also, the number of “scalings” increased in a similar manner, while the use of diagnostic codes decreased - as most patients were assigned to the yellow and red tracks, there were fewer basic examinations where diagnostics were routinely being performed (see Figure 1). The volume of other types of treatment changed only marginally.

While treatment variations in response to altered incentives have previously been reported (Brocklehurst et al., 2013; Chalkley & Listl, 2018), the type and extent of variation observed in the present paper may still appear intriguing. It is relevant to note that facilitative problems have been reported with respect to the regulatory changes examined in this study. A recently published report pointed at misalignments in treatment codes and care delivery, that is dentists were unable to receive a remuneration for a filling that needed replacement unless the patient was classified as having active disease (Ministry of Health Denmark, 2016). Apparently, dentists categorized patients in the yellow (moderate) risk group in order to get any remuneration. This was reported to have given rise to approximately 20% of yellow risk group categorizations. Also, the use of the initially introduced criteria for diagnosing gingivitis seemed to be affected by inaccuracy, resulting in a relatively large number of patients who were diagnosed with mild gingivitis (yellow risk category). After revision of the guideline, the criteria for mild gingivitis were changed so that individuals with bleeding gums at 15% of sites or less could be classified as green instead of yellow, leading to a 7%-point decrease in patients classified within the yellow category (down from a baseline of 33% before the revision).

While a considerable proportion of patients might have been misclassified, this may have happened for various reasons. Dentists may have exploited the high-risk classes by increasing demand for their services for a financial gain, taking into account that patients classified as “green” are inclined to visit the dentist less often. If such behaviour was also linked to changes in clinical activity, it may be questioned whether such changes were in the best interests of the patient. In addition to provide incentives, other factors might prevent movement towards a more individual-risk-based approach to recalls for dental check-ups. First, assessing risk is not a trivial task (Cagetti et al., 2018), especially for high-risk patients (Twetman, Fontana, & Featherstone, 2013). There have been doubts regarding the usefulness of popular risk assessment tools (Clough, Shehabi, & Morgan, 2016). Second, given the scarcity of conclusive evidence regarding fixed vs. variable recall intervals (see above), it could be hypothesized that dental care professionals may be reluctant to extend recall intervals for risk of causing harm or potential (perceived) threats of legal action for alleged malpractice. Some dentists may believe that the prolonged recall intervals for the “green” category may be too long. Third, demand-side influences may also play a role. For example, patients may have developed preferences for a frequency of services that they have become familiar with over a longer period of time or dentists might feel pressured by patients to categorize them into the yellow risk group instead of the green one because patients might have expectations or preferences to be seen semi-annually as they have been before the reform. Following this logic, the “yellow” and “red” risk classes might support both physician profit and (perceived) patient care.

By and large, rationally explainable distortions in classifications of patients into risk groups due to dentist or patient interests or preferences may constitute only one possible explanation for the observed treatment patterns. It is not always clear how best to roll out and sustain innovations to health care systems while taking into account the perspective of all relevant stakeholders and, at the same time, achieve the desired changes in health care providers’ behaviour (Greenhalgh, Robert, Macfarlane, Bate, & Kyriakidou, 2004). Given their standards, beliefs and expectations, dentists and patients may react differently than anticipated by health policymakers, leading to unpredictable consequences (possibly caused by, e.g., utility maximization or rational/irrational behavior).
Our study has limitations. As the calculations are based on unique patient-level and population-representative data composed of all 72 million treatments carried out in Denmark from 2012 to 2016, statistical analyses of this study do not suffer from pitfalls related to sampled data such as selection bias, entirely eliminating the need of extrapolating statements to a larger population. Also, the multifactorial nature of the data allowed an exact matching of patients and treatments, giving rise to the application of fixed effect methods and thus the modeling of unobserved inter-patient heterogeneity. In total, the statistical approach used is highly robust and not very prone to error and bias. However, our study suffers from the fact that detailed individual risk groupings were not available for scientific evaluation. Despite performing some robustness checks, using proxy variables for dental disease risk (socioeconomic status, age, previous treatment experience), we could not perform more detailed analyses within actual classifications of risk groups. Another weakness of our study was the lack of suitable outcome measures – while the DHA regulations require dentists to monitor the number of carious and missing teeth as well as the number of teeth with fillings, these data are limited to very narrow age groups, hence they do not provide sufficient longitudinal character and rendering the usefulness of the data for purposes of assessing oral health outcomes non-applicable to the present study (see Appendix for details). Note that all treatment codes were changed with the introduction of the regulatory changes (only exception: code for preventive treatment). While this means that, by design of the regulatory changes, there was no possibility for dentists to be slow in changing towards the new taxonomy of treatment codes, data entry mistakes could still be a relevant issue. However, given the absence of reliable reference values for dental diagnostics or treatment needs, the extent of such coding issues is difficult to determine.

Successful implementation of healthcare reforms is a function of the dynamic interaction between evidence, context and facilitation (Cohen et al., 2015). While it is often hard to translate research findings into modified provider behaviour (Grimshaw et al., 2001), the problems encountered throughout the studied reform have mostly been contextual and facilitative. Moreover, implementing health system and practice changes through incentives is highly complex and many various influencing factors can determine success or failure of payment reforms (for a more detailed review of the relevant literature see the introduction section above). Therefore, the question arises how future implementations of similar systems can try to prevent such and similar problems. For example, a sensible amendment could be the use of a monitoring system such as regular checks and inspections of dentists' assessment of oral disease risk scores by an independent regulatory institution. This way, the use of risk categories might be better aligned with actual disease prevalence rates. Another conceivable scenario would be to connect a risk-classification system and a Pay-for-Performance (P4P) implementation based on the powerful assumption that individuals and organizations are motivated to perform better by incentives (Witter et al., 2013). Literature regarding previous implementations of P4P systems in dentistry is sparse, but it has been suggested that P4P in dentistry may not be a viable option before progress is achieved in the development of reliable indicators for quality of dental care (Voinea-Griffin et al., 2010). Other conceivable options include introductions of risk group quotas based on available evidence to regulate the share of patients in each group. This approach was used as an ad-hoc solution in the actual reform by the DHA in 2016 (Sundhedsstyrelsen, 2017). In addition, the existing literature about program implementation suggests it could be sensible to precede large-scale implementations with localized, controlled implementation trials to estimate possible ramifications (Bauer et al., 2015) as adaptations to implementations of health care reforms are not uncommon (Escoffery et al., 2018). This would allow for a closely supervised roll-out during which changes can be implemented smoothly and in a co-productive manner (considering the perspectives of all relevant stakeholders) in advance of large scale (and expensive) implementation. Not least, it is important to bear in mind that health care payment reform may be shaped by social and learning processes that can affect all stakeholders involved (Conrad et al. 2016).

5 CONCLUSION

The findings presented in this study provide novel evidence that dentists change their behavior in reaction to regulatory changes to remuneration systems. The implementation of such changes therefore requires careful planning because a smooth transition to achieve the implementation's goals is challenging. Future research is warranted for further examination of how to prevent inadvertent repercussions of similar reforms and find sensible ways of implementing changes that facilitate a better alignment of supply and need. As one of the first large-scale implementations of risk-based treatment strategies, intended to improve population health and to contain health care costs, this evaluation is relevant to policymakers in all disciplines of health care.
ACKNOWLEDGEMENTS
We thank the Danish Health Authority and Anne Mette Matthiesen for data provision and useful comments. We also thank the contributors to the ADVOCATE project: the ADVOCATE Scientific Advisory Board—Stephen Birch, Martin Chalkley, Roger Ellwood, Ekatarina Fabrikant, Jeffry Fellows, Christopher Fox, Frank Fox, Dympna Kavanagh, John Lavis, Roger Matthews, Mariano Sanz, Paula Vassalo, and Sandra White; the ADVOCATE General Assembly — Robert Bryce, Lisa Bege Christensen, Gail Douglas, Kenneth Eaton, Gerard Gavin, Geert van der Heijden, Jochen Walker, Stefan Listl, Renske van der Kaaden, Melinda Madlena, Karen O’Hanlon, Andrew Taylor, Helen Whelton, Noel Woods; the ADVOCATE Ethics Advisory Board—Mary Donnelly, Jon Fistein, Peter Kurz, Evert-Ben van Veen, and Agnes Zana; the ADVOCATE project coordinator, Maria Tobin; and the coworkers of the ADVOCATE project.

CONFLICT OF INTEREST
The authors declare that they have no competing interests.

FUNDING
This research was conducted as part of the ADVOCATE project (Added Value for Oral Care). The ADVOCATE project has received funding from the European Commission’s Horizon 2020 research and innovation program under grant agreement 635183: http://advocateoralhealth.com.

ORCID
Frank Gabel https://orcid.org/0000-0001-5401-0624
Olivier Kalmus https://orcid.org/0000-0002-7255-6484
Kasper Rosing https://orcid.org/0000-0002-0024-6436
Anna-Lena Trescher https://orcid.org/0000-0003-0975-8830
Stefan Listl https://orcid.org/0000-0002-8176-3397

REFERENCE


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.