Sending emails to reduce medical costs? The effect of feedback on general practitioners' claiming of fees

Ole Kristian Aars
Department of Health Management and Health Economics,
University of Oslo,
Norway

Geir Godager
Department of Health Management and Health Economics,
University of Oslo,
Norway
Health Services Research Unit,
Akershus University Hospital, Oslo,
Norway

Oddvar Kaarbøe
Department of Health Management and Health Economics,
University of Oslo,
Norway
Department of Global Public Health and Primary Care and Department of Economics, University of Bergen,
Norway

Tron Anders Moger
Department of Health Management and Health Economics,
University of Oslo,
Norway

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH NETWORK
Working paper 2022:1

Health Economics Research Network at the University of Oslo
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 978-82-7756-272-8

Sending emails to reduce medical costs?

The effect of feedback on general practitioners' claiming of fees

Ole Kristian Aars[a]*, Geir Godager[a, b], Oddvar Kaarboe[a, c], Tron Anders Moger[a]

Abstract:

Audit and feedback is used as a strategy to guide practices of health care professionals towards certain targets. The outcome of interest can be quality improvements, but also ensuring that health care workers adhere to relevant regulations. We conducted a nationwide field experiment in the Norwegian primary care sector to study the behavioral responses from giving general practitioners feedback (GPs) on their claiming of fees. The email-based feedback intervention targeted GPs who most frequently claimed fees for double consultations and provided them with a reminder of the formal regulations for double consultations. The intervention caused a 2-5 percentage point reduction in the use of the double-consultation fee, reducing the yearly health care spending of the Norwegian government by approximately €877 000 (or €1 270 per GP).

---

a: Institute of Health and Society, Department of Health Management and Health Economics, University of Oslo, Norway

b: Health Services Research Unit, Akershus University Hospital, Oslo, Norway

c: Department of Global Public Health and Primary Care and Department of Economics, University of Bergen, Norway

*Corresponding author: email: olekaar@uio.no

"We thank HELFO for providing the data and the Research Council of Norway for funding the projects Modernizing the GP Scheme (project number 288592) and Norwegian Centre for Health Services Research (project number 296114)"

1 Introduction,

Audit and feedback is used as a strategy to align professional practice with professional targets or standards. This is widely used for health care professionals, including general practitioners (GPs). The belief is that they are prompted to modify their practice when given performance feedback showing that their clinical practice is inconsistent with a desirable target (Hysong et al., 2006). The outcome of interest can be quality improvements and/or adherence to relevant regulations when prescribing. The purpose of this paper is to evaluate a field experiment on professional feedback, following a financial audit on Norwegian GPs.

GPs in Norway are remunerated with a combination of capitation (i.e. a fixed amount per patient) and fee-for-service (FFS). Beyond patient co-payments, GPs are reimbursed by HELFO – the government agency tasked with paying health professionals. One of the most used fees is a double-consultation fee (DCF), which can be claimed for consultations with a long duration (lasting more than 20 minutes). In 2019, the DCF was used more than five million times, amounting to €110M. This constituted more than 20% of the total fee reimbursement GPs received from the Norwegian health insurance scheme (HELFO, 2020).

In collaboration with HELFO, we conducted a nationwide field experiment in the Norwegian primary care sector to study the causal effects of giving GPs feedback on their claiming of DCF. The email-based feedback intervention targeted GPs who most frequently claimed fees for double consultations and provided them with a reminder of the formal regulations for double consultations. GPs in our study sample were randomly assigned to a control group (Control) or one of the two intervention groups Mild or Strong. GPs in the two intervention groups would receive different versions of a feedback email, whereas GPs in Control would not receive any feedback email. The feedback emails contained information about the total cost of DCFs reimbursed by HELFO in 2018 and stated that the GP used DCFs more frequently than the average GP. The information received by the two different intervention groups only differed in the heading. Mild received a feedback email with the heading "Information about DCF" and Strong had the heading "Regarding your use of DCF". With HELFO distributing the email, we were able to use HELFO's established infrastructure to conduct the experiment, thus minimizing administrative costs and costs for patients and providers (Ivers et al., 2014), while enhancing the external validity of the study (Harrison et al., 2004). The field experiment enabled us to compare the behavioral responses to the different formulations used when messaging GPs, similar to Bott et al. (2019) who studied the effects of differently worded emails to reduce tax evasion. We also adjusted for seasonal effects and studied the impact of the Covid-19 pandemic on the estimated effects. Finally, and due to the possibility of internal communications among the GPs in a restricted Facebook group for GPs in Norway[2] and coverage of the intervention in the (national) media (Brandtzæg Clausen, 2019; Hafstad, 2019; Storvik, 2019), we analyzed the effect of the intervention on ***Control*** to quantify possible spillover effects. We quantified the causal effect of our field intervention on the monthly claims of the DCF relative to the regular consultation fee, i.e., the fee for a consultation lasting less than 20 minutes, in the 14-month period after the intervention.

The intervention had a statistically significant effect: Both the descriptive analysis of mean differences and the regression models showed a 2-5 percentage point drop in the use of DCF for both mild and strong intervention groups compared to the pre-treatment period. This relatively large short-term effect diminished over time. However, the effect remained statistically significant one year after the

[2] According to Gronseth et al. (2020), a restricted Facebook group for GPs in Norway exists. As of spring 2018, the list included 3,357 members, of whom approximately 50 participated regularly in discussions.

intervention. We also find that the observed difference between the ***Mild*** and ***Strong*** intervention groups is not statistically significant. Our interpretation is that receiving feedback has a larger impact than the specific wording used. The intervention effects in this study have economic significance. The reduction observed in the use of DCF for the intervention groups added up to €877 279 per year or €1 270 per GP in the sample.

We contribute to the literature on field experiments in health care by implementing two relatively minor email interventions to influence clinical practice and generate cost savings. Related studies focusing on antibiotics have shown that information letters (Meeker et al., 2016; Schwartz et al., 2021) and a mystery shopper scheme (Cheo et al., 2020) can reduce prescribing. Laboratory experiments have also been used for studying the behavioral responses caused by disclosing information about providers' performance, and by reminding providers about professional norms. Godager et al. (2016) found that compared to a regime with private information, a regime with performance disclosure was more likely to result in maximum benefits for patients. Experimental results reported by Kesternich et al. (2015) indicate that raising the saliency of professional norms affect patient-regarding preferences and improve health outcomes.

A common approach in the literature is to refer to a recommended clinical practice. In our study, the informational email only contained reference to a consultation fee without benchmarking it to standardized clinical care. More generally, there are other studies on the effects of auditing and feedback on clinical practice establishing the effects as small to moderate (Eccles et al., 2001; Ivers et al., 2012; Jamtvedt et al., 2006). These studies have primarily been concerned with high intensity feedback – i.e., peer-to-peer, interviews, telephone calls, visits, educational components, and seminars. Likewise, the use of reminder messages to reduce radiology referrals has shown to be effective, but without reference to a GPs relative performance (Shojania et al., 2009). Moreover, these studies have primarily been concerned with – implicitly or explicitly – making accurate assessments of diagnoses.

The paper proceeds as follows: Section 2 provides an overview of the Norwegian study setting. The randomization and interventions in the field experiment is presented in Section 3. Data and empirical methods are presented in Section 4. The results from nonparametric and parametric analysis are presented in Section 5, followed by a discussion of the findings, limitations, and implications for GP practice in Section 6.

2 Study setting

Norway has a National Health Service system financed through general taxation. Norwegian health care is organized into primary and secondary health care sectors. The former is the responsibility of municipalities while the latter is the responsibility of the central government. Since 2001, every Norwegian is listed with a GP, who also acts as a gatekeeper to access specialized care. In 2019, there were approximately 4,800 GPs, and only 0.2 % of the inhabitants had opted out of the system (Gaardsrud, 2020). Patients may switch GPs twice a year, and about 3 % of the patients do so annually. Most GPs (85 %) are self-employed and contract with a municipality. All fees and co-payments are set at the national level, without any geographical variations. The fee schedule specifies patient co-payments and fees reimbursed by HELFO - the government agency tasked with paying health professionals who contract with the National Health Service. The fee schedule includes a DCF, a fee that can be claimed in addition to the RCF by GPs when consultations exceed a duration of 20 minutes and can be repeated per each started 15 min.[3] The RCF is

[3] I.e. One DCF can be claimed for a 30 min consultation and two DCFs can be claimed for 35 min.

higher for GPs who have qualified as specialists in general medicine. A typical GP consultation would result in claims for one RCF plus DCFs for long consultations.

Table 1 GP fees for three consultation durations in 2019.

	Consultation duration in minutes		
	0 – 20	21 – 35	36 – 50
RCF [RCF for specialist in general med.]	€16 [€26]	€16 [€26]	€16 [€26]
DCF	0	1*€21	2*€21
Total claim	**€16 [€26]**	**€38 [€48]**	**€59 [€69]**

HELFO is also the financial auditor of GPs and may independently determine sanctions against health professionals that have failed to comply with the relevant regulations. HELFO performs financial audits of health professionals to ensure that the reimbursements to the health professionals are in line with the financial regulations of the public national insurance scheme. They also organize courses for GPs as part of their specialization. Following an audit, HELFO may independently determine sanctions against health professionals that have failed to comply with the relevant rules. The sanctions include instructions to adjust current practice, refund, loss of the right to practice at the expense of the Norwegian government, and reporting to the police. In 2020 (2019), HELFO reported 9 (3) cases to the police, 8 (10) health professionals lost the right to practice at the expense of the Norwegian government, and health care professionals had to return €5.6M (€2.8M) of payments received (Norwegian Directorate of Health, 2020). DCF has been a focus-area for HELFO in the lead up to our intervention, including information to GPs on fee regulations through newsletters and courses and by implementing automating rejection of excess use of DCFs.

3 The field experiment

3.1 Sampling and randomization

The inclusion criteria for the field intervention sample were as follows:

1. GPs had to claim reimbursement from HELFO during the first six months of 2019 (“be active”).
2. GPs had to claim at least 500 RCF during this period.
3. GPs had to rank among the top 700 GPs based on their frequency of DCFs relative to RCFs.

The frequency of DCFs relative to RCFs in our sample varied from 59 % to 173 % during the inclusion period. In comparison, the average relative frequency for all GPs was around 38 %. Nine GPs were dropped since they already were being audited by HELFO, resulting in a study sample of 691 GPs. To avoid contamination between study arms, we made sure that GPs located at the same GP practice address were allocated to the same arm of the experiment[4].

After sampling, GPs were randomly assigned to one of three study arms: ***Control*** (no feedback) or either of two intervention groups ***Mild*** or ***Strong***. ***Mild*** and ***Strong*** received identical feedback emails except for the different headings. GPs in ***Strong*** received feedback email where the heading was “*Regarding your use*

[4] Approximately 49 % of GPs in the sample belonged to the same GP-practice, of which 26 % were sharing practice with one other GP in the sample, while 23 % were in practices with 2-4 other sample GPs.

of DCF". In contrast, GPs in ***Mild*** received feedback emails where the heading was simply: "*Information about DCF*". Hence, we followed Bryan et al. (2013) and used active language to address the reader as "you" in the ***Strong*** arm and passive language in the ***Mild*** arm.

3.2 The feedback intervention

In the email body, the first paragraph stated the total amount (in NOK) that HELFO reimbursed for DCFs in 2019. The second paragraph gave details about the information campaign "Do you know" that was meant to increase the GP's awareness about claiming fees (HELFO, 2019). The third paragraph stated actively that "you" were receiving the email since statistics showed that "you" had used DCFs significantly more that the average GP. The fourth paragraph provided information about where the documentation for the claim in the former paragraph was taken from and where the GP could find more information about his or her claiming of fees. The fifth paragraph stated that the email was for information and guidance that the GP need not answer. It also provided an email address, a phone number, and a reference number to be used if the GP would like to contact HELFO. The last paragraph provided a link to the "Do you know" campaign and a link to the HELFO newsletter.

An overview of the process for arriving at the control and intervention groups is provided in Flow Chart 1.

[Flow chart 1 about here]

4 Data and methods

4.1 Variable definitions

Data on the use of DCFs and RCFs per month from Jan 2017 to Nov 2020 were extracted for each GP in the three study arms. We could then compute the percentage of DCF relative to RCF henceforth denoted by ***%DCF***. ***%DCF*** is the main outcome variable in our empirical analyses. Using the indexes *ijt* to represent the fee *j* claimed by GP *i* in month *t*, our outcome variable ***%DCF*** is defined in Equation (1).

$$\%DCF_{it} \stackrel{\text{def}}{=} 100 * \frac{\sum_j DCF_{ijt}}{\sum_j RCF_{ijt}} \qquad (1)$$

In the descriptive analysis, within group differences in the ***%DCF*** claims over time periods after vs. before the intervention were examined by means of Wilcoxon signed rank tests. Overall differences in the ***%DCF*** claims across the intervention groups were examined by Kruskal-Wallis tests. Tests were performed both as a simple comparison before vs. after the intervention and as before the intervention vs. three time periods after the intervention (0-4 months, 5-9 months, 10-14 months) to study whether effects were reduced over time. Non-parametric tests were used due to skewness in the ***%DCF*** claims in the sample.

4.2 Model specification

We specify linear regression models with random GP specific effects. To examine how effects change over time, our specification let post intervention effects vary by time period. We control for seasonal effects by using dummy variables for months and studied the impact of the Covid-19 pandemic on the estimated effects by adding a dummy variable equal to 1 from March 2020 onwards. Our model is specified as:

$$\%DCF_{it} = \beta_0 + \boldsymbol{\beta}_1 * \boldsymbol{Int}_i * \boldsymbol{Post}_t + \boldsymbol{\beta}_2 * \boldsymbol{Month}_t + \beta_3 * Covid_t + u_i + \varepsilon_{it}, \qquad (2)$$

where $\boldsymbol{Int_i}$ is a vector of dummy variables for the three arms. $\boldsymbol{Post_t}$ is a vector of dummy variables for the periods Sep 2019-Jan 2020 (0-4 months after intervention), Feb-Jun 2020 (5-9 months after) and Jul-Nov 2020 (10-14 months after). $\boldsymbol{Int_i} * \boldsymbol{Post_t}$ is the interaction between the two, giving rise to nine combinations, three terms per arm.[5] $\boldsymbol{Month}$ is a vector of eleven dummy variables Feb,…, Dec with January being the reference category. $Covid$ is a dummy variable taking the value 1 starting from March 2020. The latter two were included to assess the robustness of the intervention effect by taking into account changes in activity across seasons and following the pandemic. Finally, u_i is a GP specific random effect, while ε_{it} is a noise term.

Note that in our specification, we assume no difference in DCF use between the three arms prior to the intervention (observations score 0 on all dummies in $\boldsymbol{Int_i}$ and $\boldsymbol{Post_t}$ variables). This assumption is reasonable given that GPs are randomized to one of the three arms. The assumption was also supported by testing for difference using observations prior to Sep 2019 only (both by Kruskal-Wallis and by using a regression model with dummy variables for the study arms). Also note that the model will enable us to study effects in the control arm over time and hence the possible impact of information leaks via personal communication between GPs, in Facebook groups, and to the media: Effects in the control arm are represented by the three coefficients in $\boldsymbol{\beta_1}$ that are assigned to ***Control***. Similarly, the three coefficients in $\boldsymbol{\beta_1}$ that are assigned to ***Mild*** (and the three coefficients in $\boldsymbol{\beta_1}$ that are assigned to ***Strong***) will reflect the effects of the ***Mild*** (***Strong***) intervention in each time period post intervention compared to pre intervention.

The distribution of ***%DCF*** is skewed in the sample. Using a gamma generalized linear mixed model with log link yielded very similar p-values for the regression coefficients, thus we opted to keep the linear model to get absolute instead of relative effects from the independent variables. The small differences in results across model specifications is likely to be a result of the sample size (691 GPs contributing to 27,304 observations).

The choice between a fixed or random effect model will in general involve a trade-off between robustness (fixed effects) and efficiency (random effects). A random effect model provides efficient slope estimates when the random effects are uncorrelated with the regressors, and a random effect model is preferred when this assumption is met. Since all regressors are deterministic in the case at hand, the random effect assumptions are not restrictive. Our choice was also supported by a Hausman test (p-value of 0.13 for the model above), and there were negligible differences between the coefficients from the fixed and random effect models. In two sensitivity analyses, we first added to the model a random effect for practice to capture dependence in doctors' behaviors within the same practice and variation between practices in ***%DCF*** claims. However, this did not change the results. The variation between practices was non-significant when including variation between GPs. Second, GPs with low activity in regular consultation fees will either have lower activity in general or have shorter total follow-up time. Removing the 50 GPs with the lowest activity in regular consultation fees during the study period did not alter the results.

[5] These nine coefficients correspond to the nine first coefficients in **Table 3**.

5 Results

5.1 Descriptive results

Table 2 present descriptive results. The arms are approximately balanced, considering the large standard deviations for the fees per month per GP. As is clear, the unadjusted effects seem stable through the follow-up time period. There is a small reduction in ***%DCF*** after the intervention in the control arm; however, this is not significant at the 5 % level for any of the time periods after compared to before the intervention according to Wilcoxon signed rank tests (p>0.2 for all). For the mild and strong intervention, however, the reductions in ***%DCF*** for all time periods after the intervention compared to before are statistically significant (p<0.01 for all). There are also significant differences between the control and intervention groups in ***%DCF*** for each time period after intervention vs. before (p=0.01 for time period 0-14 months vs. before, Table 2).

Table 2 Descriptive statistics. Mean (SD) monthly claims for RCF and DCF and Mean (SD) of %DCF for the three arms.

Study arm	Variable	Before intervention:	After: 0-4 mths (p<0.001)*	5-9 mths (p=0.027)*	10-14 mths (p=0.025)*	0-14 mths (p=0.01)*
Control N=230	***%DCF***	70.5 (16.7)	69.6 (20.9)	68.6 (20.4)	69.5 (20.8)	68.8 (20.2)
	#DCF	136 (53)	142 (63)	131 (62)	139 (65)	137 (63)
	#RCF	193 (60)	206 (81)	194 (81)	207 (89)	200 (79)
Mild N=230	***%DCF***	70.2 (18.0)	65.9 (21.1)	66.7 (22.2)	67.6 (23.4)	66.2 (19.5)
	#DCF	144 (49)	148 (58)	142 (59)	148 (65)	226 (81)
	#RCF	209 (65)	230 (87)	222 (86)	227 (94)	145 (56)
Strong N=231	***%DCF***	70.3 (16.7)	65.8 (19.9)	66.3 (22.0)	66.8 (21.8)	66.2 (19.6)
	#DCF	147 (61)	152 (72)	141 (72)	147 (75)	225 (78)
	#RCF	211 (70)	234 (83)	213 (84)	220 (84)	148 (70)

*The p-values refer to Kruskal-Wallis tests on differences across intervention and control arms in the %DCF at 0-4 months after vs. before, 5-9 months after vs. before, and 10-14 months after vs. before.

Figure *1* presents the ***%DCF*** by arm and months. In September 2019 – the first month GPs would have been able to alter their claiming of fees – we see a distinct reduction in the use of DCFs for all three study arms. The largest reduction in DCF claims is observed in the two intervention groups, Strong and Mild.

[Figure 1 about here]

Figure 2 presents a histogram of relative frequencies of GPs in each of ten deciles for ***%DCF*** claims in August and September 2019. The intervention was implemented in the end of August 2019. When comparing the distribution in September with August, one can see clearly how the probability mass was shifted to the left. For example, we see that the relative frequency of GPs with ***%DCF >*** 90% was reduced from 17.9 % in August 2019 to 13.1 % in September 2019. We also see that the relative frequency of GPs with ***%DCF*** between 40 and 49 was more than doubled, from 4.9 % in August 2019 to 11.8 % in September the same year.

[Figure 2 about here]

5.2 Regression analysis

Table 3 presents the estimation results from two regression models; the unadjusted model showing the effects of each intervention group at different time periods and an adjusted model where dummy variables are added for months and the Covid-19 outbreak.

Table 3: Estimation results from linear regression models. N=691 GPs contributing to 27,304 observations in total. Dependent variable %DCF

	Unadjusted analysis		**Adjusted analysis**	
Variable:	**Coefficient (95% CI):**	**p-value**	**Coefficient (95% CI):**	**p-value**
Before intervention	Reference		Reference	
Control 0-4 months after	-1.05 (-2.08, -0.02)	0.05	-1.04 (-2.10, 0.03)	0.06
Control 5-9 months after	-1.14 (-2.19, -0.08)	0.03	0.58 (-1.08, 2.24)	0.49
Control 10-14 months after	-0.38 (-1.46, 0.71)	0.49	1.14* (-0.79, 3.08)	0.25
Mild 0-4 months after	-4.01 (-5.01, -3.00)	<0.001	-3.98 (-5.02, -2.93)	<0.001
Mild 5-9 months after	-3.97 (-4.99, -2.93)	<0.001	-2.25 (-3.89, -0.61)	0.01
Mild 10-14 months after	-4.34 (-5.39, -3.26)	<0.001	-2.82 (-4.73, -0.90)	<0.001
Strong 0-4 months after	-4.96 (-5.95, -3.96)	<0.001	-4.93 (-5.96, -3.88)	<0.001
Strong 5-9 months after	-3.79 (-4.83, -2.75)	<0.001	-2.06* (-3.71, -0.41)	0.01
Strong 10-14 months after	-3.86 (-4.93, -2.80)	<0.001	-2.34* (-4.26, -0.41)	0.02
Dummies for months			YES	
Before Covid-19			Reference	
After Covid-19			-1.96 (-3.54, -0.37)	0.02
Constant	70.52 (69.23, 71.80)	<0.001	71.50 (70.08, 72.91)	<0.001
R2	0.0047		0.0085	
Share variance due to GP random effect	56%		55%	

*Marks significant differences across time periods within the study arms at the 5% level.

There is a borderline significant effect of around one percentage point reduction in %DCF in the control arm 0-4 months after the intervention in both unadjusted and adjusted analyses. However, this seems to disappear after 5 months in the adjusted analyses.

There is a statistically significant drop in ***%DCF*** for both ***Mild*** and ***Strong*** intervention arms – and substantially higher than for the control group. While the ***Mild*** and ***Strong*** intervention groups also have an apparent reduction over time following intervention after adjusting for seasonal effects and Covid-19, the reduction is significant only for the ***Strong***. Significance follows from tests of the $\boldsymbol{Int_i} * \boldsymbol{Post_t}$ interaction, using 0-4 months after the intervention as reference for each of the study arms. For ***Strong*** in the adjusted analyses, for example, the initial effect is a reduction of 4.93 percentage points in ***%DCF***. The size of the effect is reduced over time but remains significant; that is, a medium- and a long-run effect exists, respectively, of 2.06 and 2.34 percentage points, 5-9 months and 10-14 months after the intervention. Switching reference category for the $\boldsymbol{Int_i} * \boldsymbol{Post_t}$ interaction will also show significant differences for each time period between controls after the intervention and for both the mild and strong intervention groups. There are no significant differences between the mild and strong interventions, however. The large variation in the outcome is apparent from the low R2-values. Due to collinearity, it is not feasible to further separate exogenous calendar time trends from the effects of the intervention.

6 Discussion and conclusion

We conducted a nationwide field experiment in the Norwegian primary care sector to study the behavioral responses of GPs receiving feedback on their claiming of fees. The interventions focused on the mode of treatment and whether the use of fees was aligned with regulatory and not clinical guidelines.

We found statistically significant effects of a simple feedback intervention via email. The effects are relatively large and long-lasting, and observable in both intervention groups through the 14-month follow-up period. We also find a short-run borderline significant effect in the control group, most probably caused by sharing of information in a closed Facebook groups and media coverage that followed the intervention. While our findings inform the debates around drawing causal conclusions from randomized control trials, the known challenges concerning the upscaling of experiment results into real world settings remain (Al-Ubaydli et al., 2021; Al-Ubaydli and List, 2015; Banerjee et al., 2017; Deaton, 2020; Harrison, 2021, 2011).

There are multiple channels through which the intervention caused behavioral responses for the ***Mild*** and ***Strong*** intervention groups. The intervention highlighted the existence of HELFO's auditing measures. The mere existence of an audit may influence a GP's belief about the likelihood that the auditor has knowledge about his or her income-generating activities. This may affect the decision to claim the DCF. This information is given to both intervention groups, but not to the control group and the GPs that learned about the audit through the media leak. However, the mild intervention group would only learn that HELFO possessed this knowledge if they opened the feedback email, as the passive language used in the heading did not reveal that the email contained individual information of the GPs use of the DCF. Having said so, the mild intervention group might have learned about the individual information given in the email through the media leak, as the media reported that this information was given in the email, and that the division director of HELFO confirmed that the receivers of the emails had used the DCF significantly more than the average GP (Storvik, 2019). Hence the shift in the claiming of DCFs can be understood as being a resulting of shift in beliefs.

We conclude by noting that the effects of the field experiment constituted a significant reduction in reimbursement of the DCF, hence contributing to savings for Norwegian taxpayers. The reduction observed in ***%DCF*** among the mild and strong intervention groups in the study sample add up to €877 279 per year or €1 270 per GP. How big the savings would be if the field experiment were to be scaled up is difficult to estimate due to issues of external validity and scalable policies (Al-Ubaydli and List, 2015;

Banerjee et al., 2017). One issue that might lower the effect of a large-scale implementation is that honest GPs might feel they are unfairly treated due to the audit and respond by being less inclined to follow the financial regulations of the public National Insurance Scheme (Houser et al., 2012; Hu and Ben-Ner, 2020).

References

Al-Ubaydli, O., Lee, M.S., List, J.A., Mackevicius, C.L., Suskind, D., 2021. How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling. Behavioural Public Policy 5, 2–49. https://doi.org/10.1017/bpp.2020.17

Al-Ubaydli, O., List, J., 2015. On the Generalizability of Experimental Results in Economics: With A Response To Camerer, in: Handbook of Experimental Economic Methodology. https://doi.org/10.3386/w19666

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017. From proof of concept to scalable policies: Challenges and solutions, with an application, in: Journal of Economic Perspectives. American Economic Association, pp. 73–102. https://doi.org/10.1257/jep.31.4.73

Bott, K.M., Cappelen, A.W., Sørensen, E.O., Tungodden, B., 2019. You've Got Mail: A Randomized Field Experiment on Tax Evasion. Management Science 66, 2801–2819. https://doi.org/10.1287/MNSC.2019.3390

Brandtzæg Clausen, V., 2019. Reagerer på at 450 fastleger får brev fordi de bruker for mye tid med pasienter [Letters to 450 GPs regarding excess consultation length causes reactions]. TV2.

Bryan, C.J., Adams, G.S., Monin, B., 2013. When cheating would make you a cheater: Implicating the self prevents unethical behavior. Journal of Experimental Psychology: General 142, 1001–1005. https://doi.org/10.1037/a0030655

Cheo, R., Ge, G., Godager, G., Liu, R., Wang, J., Wang, Q., 2020. The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment. Health Economics Review 2020 10:1 10, 1–19. https://doi.org/10.1186/S13561-020-00290-Z

Deaton, A., 2020. Introduction: Randomization in the Tropics Revisited, a Theme and Eleven Variations. Randomized Control Trials in the Field of Development 29–46. https://doi.org/10.1093/oso/9780198865360.003.0002

Eccles, M., Steen, N., Grimshaw, J., Thomas, L., McNamee, P., Soutter, J., Wilsdon, J., Matowe, L., Needham, G., Gilbert, F., Bond, S., 2001. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: A randomised trial. The Lancet, 357, 1406–1409. https://doi.org/10.1016/S0140-6736(00)04564-5

Gaardsrud, P.Ø., 2020. Styringsdata for fastlegeordningen, 4 kvartal 2019 [GP management data, 4th quarter 2019]. Norwegian Directorate of Health, Oslo.

Godager, G., Hennig-Schmidt, H., Iversen, T., 2016. Does performance disclosure influence physicians' medical decisions? An experimental study. Journal of Economic Behavior and Organization 131, 36–46. https://doi.org/10.1016/J.JEBO.2015.10.005

Gronseth, I.M., Malterud, K., Nilsen, S., 2020. Why do doctors in Norway choose general practice and remain there? A qualitative study about motivational experiences. Scandinavian Journal of Primary Health Care 1–8.

Hafstad, A., 2019. Helfo-kampanjen – en bommert [The Helfo campaign - off target]. Dagens Medisin.

Harrison, G.W., 2021. Field experiments and public policy: festina lente. Behavioural Public Policy 5, 117–124. https://doi.org/10.1017/bpp.2020.28

Harrison, G.W., 2011. Randomisation and Its Discontents. Journal of African Economies 20, 626–652. https://doi.org/10.1093/JAE/EJR030

Harrison, G.W., List, J.A., Burks, S., Camerer, C., Carpenter, J., Gerking, S., Isaac, R.M., Krueger, A., Mcmillan, J., Ortmann, A., Plott, C., Reiley, D., Rutström, E.E., Wilcox, N., 2004. Field Experiments. Journal of Economic Literature XLII, 1009–1055.

HELFO, 2020. Helserefusjon 2019- Statistikk og nøkkeltall for legeområdet [Health care claims 2019 - statistics and data on doctors].

Houser, D., Vetter, S., Winter, J., 2012. Fairness and cheating. European Economic Review 56, 1645–1655. https://doi.org/10.1016/j.euroecorev.2012.08.001

Hu, F., Ben-Ner, A., 2020. The effects of feedback on lying behavior: Experimental evidence. Journal of Economic Behavior and Organization 171, 24–34. https://doi.org/10.1016/j.jebo.2019.12.019

Hysong, S.J., Best, R.G., Pugh, J.A., 2006. Audit and feedback and clinical practice guideline adherence: Making feedback actionable. Implementation Science 1, 9. https://doi.org/10.1186/1748-5908-1-9

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J.M., Odgaard-Jensen, J., French, S.D., O'Brien, M.A., Johansen, M., Grimshaw, J., Oxman, A.D., 2012. Audit and feedback: Effects on professional practice and healthcare outcomes. Cochrane Database of Systematic Reviews 2012. https://doi.org/10.1002/14651858.CD000259.pub3

Ivers, N.M., Grimshaw, J.M., Jamtvedt, G., Flottorp, S., O'Brien, M.A., French, S.D., Young, J., Odgaard-Jensen, J., 2014. Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. Journal of general internal medicine 29, 1534–1541. https://doi.org/10.1007/S11606-014-2913-Y

Jamtvedt, G., Young, J.M., Kristoffersen, D.T., Ann, M.O., Oxman, A.D., 2006. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. BMJ Quality & Safety 15, 433–436. https://doi.org/10.1136/qshc.2006.018549

Kesternich, I., Schumacher, H., Winter, J., 2015. Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? Journal of Public Economics 131, 1–11. https://doi.org/10.1016/J.JPUBECO.2015.08.009

Meeker, D., Linder, J.A., Fox, C.R., Friedberg, M.W., Persell, S.D., Goldstein, N.J., Knight, T.K., Hay, J.W., Doctor, J.N., 2016. Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices: A Randomized Clinical Trial. JAMA 315, 562–570. https://doi.org/10.1001/JAMA.2016.0275

Norwegian Directorate of Health, 2020. Årsrapport 2019 [Annual report 2019].

Schwartz, K.L., Ivers, N., Langford, B.J., Taljaard, M., Neish, D., Brown, K.A., Leung, V., Daneman, N., Alloo, J., Silverman, M., Shing, E., Grimshaw, J.M., Leis, J.A., Wu, J.H.C., Garber, G., 2021. Effect of

Antibiotic-Prescribing Feedback to High-Volume Primary Care Physicians on Number of Antibiotic Prescriptions: A Randomized Clinical Trial. JAMA Internal Medicine. https://doi.org/10.1001/JAMAINTERNMED.2021.2790

Shojania, K.G., Jennings, A., Mayhew, A., Ramsay, C.R., Eccles, M.P., Grimshaw, J., 2009. The effects of on-screen, point of care computer reminders on processes and outcomes of care. Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.CD001096.pub2

Storvik, A.G., 2019. Leger reagerer på Helfo-brev om tidsbruk [GPs react to Helfo-letter regarding consultation length]. Dagens Medisin.

Flow Chart 1: Process for sampling, randomization, and intervention

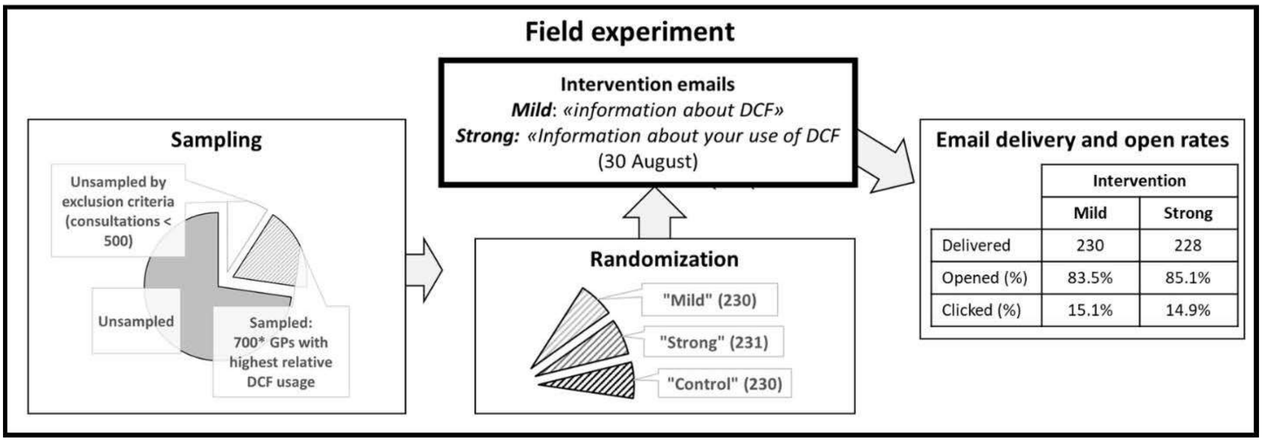

Figure 1: %DCF by study arm and month. N=691 GPs contributing to 27,304 observations in total

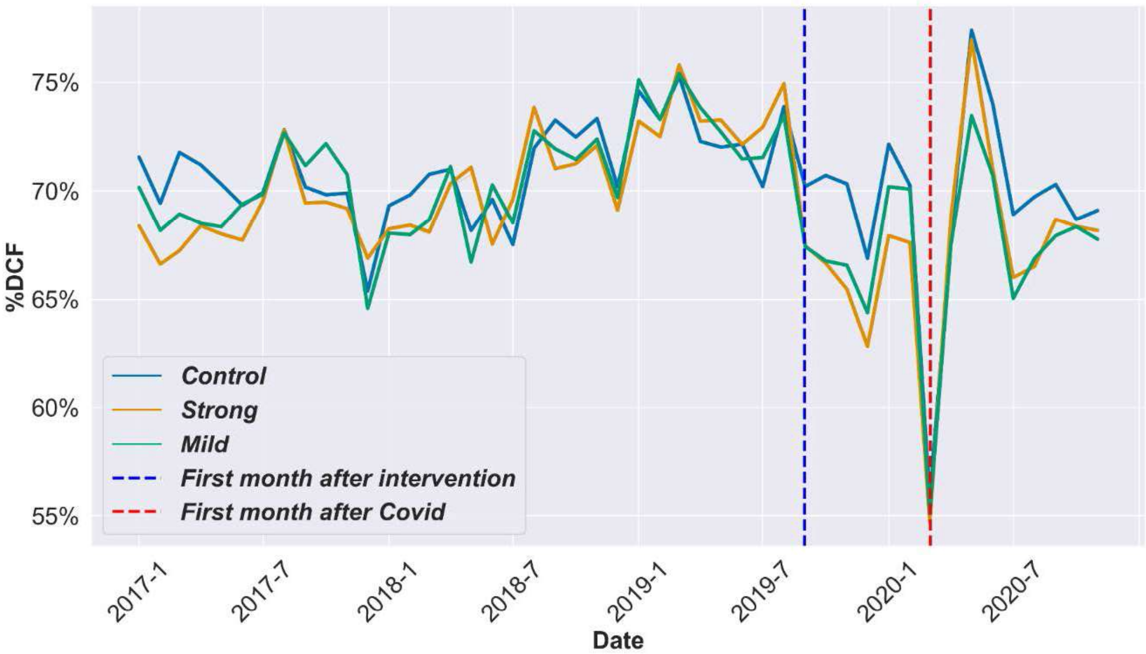

Figure 2: Histogram of relative frequencies in the months pre- and post- intervention. N=691 GPs contributing to 27,304 observations in total