The effectiveness of repeating a social-norm feedback intervention to high prescribers of antibiotics in general practice: A national regression discontinuity design

Michael RATAJCZAK[1, 2]

Natalie GOLD* D.Phil[1, 3]

Simon HAILSTONE[1]

Tim CHADBORN PhD[1]

[1]Public Health England, Wellington House, 133-155 Waterloo Road, London, SE1 8UG, United Kingdom

*Corresponding author: Natalie Gold, natalie.gold@phe.gov.uk, 07967 839 044

Short running title: repeating a social-norm feedback intervention

[2]Present affiliation: Department of Linguistics and English Language, Lancaster University, Lancaster, LA1 4YL, United Kingdom
[3]Also affiliated with: Faculty of Philosophy, Radcliffe Humanities, Woodstock Road, Oxford, OX2 6GG, United Kingdom

Structured synopsis

Objectives

Unnecessary antibiotic prescribing contributes to antimicrobial resistance. A randomized controlled trial in 2014-5 showed that a letter from England's Chief Medical Officer (CMO) to high prescribing General Practitioners (GPs), giving feedback about their prescribing relative to the norm, decreased antibiotic prescribing. The CMO sent further feedback letters in succeeding years. We evaluate the effectiveness of the repeated feedback intervention.

Methods

Publicly available databases were used to identify GP practices whose antibiotic prescribing was in the top 20% nationally (the intervention group). In April 2017, GPs in every practice in the intervention group (n = 1,439) were sent a letter from the CMO. The letter stated that, "the great majority of practices in England prescribe fewer antibiotics per head than yours". Practices in the control group received no communication (n = 5,986). We used a Regression Discontinuity Design to evaluate the intervention because assignment to the intervention condition was exogenous, depending on a 'rating variable'. The outcome measure was the average rate of antibiotic items dispensed from April 2017 to September 2017.

Results

The GP practices who received the letter changed their prescribing rates by -3.69% [95% CI = -2.29, -5.10]; $p < 0.001$, representing an estimated 124,952 fewer antibiotic items dispensed. The effect is robust to different specifications of the model.

Conclusions

Social-norm feedback from a high-profile messenger continues to be effective when repeated. It can substantially reduce antibiotic prescribing at low cost and at national scale. Therefore, it is a worthwhile addition to antimicrobial stewardship programmes.

Introduction

Antimicrobial resistance (AMR) is a growing public health threat.[1] It is estimated that 700,000 people die of resistant infections every year and that this number could rise to 10 million lives a year by 2050, with common procedures, such as surgery and chemotherapy, becoming too dangerous to perform.[2] One cause of the problem is poor antimicrobial stewardship, with large quantities of antibiotics being prescribed to patients who do not need them.[2, 3]

In the UK, about 80% of antibiotics are prescribed in primary care.[4] There is significant variation in prescribing between General Practitioner (GP) practices, which cannot be explained by practice demographics, indicating that many unnecessary antibiotics are being prescribed in primary care.[5-7] As a part of its AMR stewardship strategy, the UK Government aims to eventually give all health and care providers feedback in a format that is useable and relevant to support good practice.[8]

Giving GPs feedback on their practice's performance compared to other practices can decrease prescribing.[9] In September 2014, we ran a randomised controlled trial (RCT) in which the Chief Medical Officer (CMO) sent a feedback letter to 3,227 GPs stating that their practice was prescribing antibiotics at a higher rate than 80% of practices in its local area. As well as the social norm information, the letter presented three specific, feasible actions that the recipient could take to reduce unnecessary prescriptions of antibiotics: giving patients advice on self-care, offering a delayed prescription, and talking about the issue with other prescribers in his or her practice. The letter was accompanied by a copy of the "Treating your infection" leaflet (a part of the TARGET Antibiotic Toolkit),[10] which aims to facilitate communication between prescriber and patient, and increase the patient's confidence to self-care. Between October 2014 and March 2015, the practices in the intervention group dispensed 3.3% fewer antibiotic items relative to the control group. The rate of dispensing antibiotics differed significantly in every month in the study period, with no evidence of a trend. Therefore, the feedback letter was sent to the control group at the beginning of April 2015.

The rise in AMR is an issue of international concern,[1, 2] so the trial had an international impact. The intervention has already been adopted by CMOs in Australia,[11] Northern Ireland,[12] and Canada,[13, 14] and France is planning to follow suit.[15]

Following on from the success of the RCT, the CMO for England has sent annual feedback letters to GPs whose practices are in the top 20% of prescribers, each winter flu season. However, there is a question of whether a repeated feedback intervention continues to be effective.[16] Because the CMO's feedback was targeted at a specific segment of practitioners—the top 20%—it is possible to use a regression discontinuity design to evaluate the subsequent feedback intervention.

Method

Intervention: Letters to High Antibiotic Prescribers

The CMO sent her feedback letter on antibiotic prescribing rates to GPs in April 2017. The feedback letters were sent to 6,318 individual GPs in 1,439 different GP practices with high antibiotic prescribing rates. Practices were allocated to the intervention arm if they were in the top 20% of prescribers for the twelve months prior to the end of the intervention (Oct 2015 –Sep 2016), as judged using a prescribing indicator which divides Antibacterial Items by STAR-PU (Specific Therapeutic group Age-sex Related Prescribing Units) weightings. This means that the prescribing rate is adjusted to take into account some of the demographics of the GP practice.

The letters differed slightly from the 2014 trial:[9] NHS Local Areas no longer exist, so each practice's prescribing was compared to that of practices in England; and additional guidance was presented in a box on the right-hand side, instead of in the middle of the letter. There were six different letters (Supplementary Data: Figures S1-S6). The letters were tailored according to GP practice prescribing rate, change in prescribing over time, and whether they were previously sent feedback. Specifically, the top 11-20% practices with increasing prescribing rates were told 'The great majority (80%) of practices in England prescribe fewer antibiotics per head than yours. Most other practices have reduced their prescribing rates since 2013/14 but yours has increased' (Letter 1: Figures S1-S2); the rest of the top 20% were simply told 'The great majority (80%) of practices in England prescribe fewer antibiotics per head than yours' (Letter 2: Figures S3-S4). Finally, the top 10% of practices were told 'The great majority (90%) of practices in England prescribe fewer antibiotics per head than yours' (Letter 3: Figures S5-S6). When a practice had received a previous letter, the CMO noted, 'I have written to your practice previously', so there were two variants of each letter, A and B. Table 1 shows the key differences between the letter variants. As in the 2014 trial, the letter was accompanied by a copy of the TARGET "Treating your infection" leaflet (Supplementary Data: Figure S7).

Regression Discontinuity Design

We decided to use Regression Discontinuity Design (RDD) to analyse the effects of our intervention instead of more conventional analytical approaches, such as interrupted time series (ITS) and difference-in-difference design. RDD is a standard approach for evaluating interventions.[17] In an RDD, the assignment of participants to the intervention versus the control condition is exogenous, depending on whether a numerical 'rating variable' falls above or below a certain threshold. Participants scoring above the threshold are assigned to one group, such as the treatment, whereas those scoring below the threshold are assigned to another group, for example, the control. The assumption behind RDD is that, in a window around the threshold (the 'bandwidth'), observations on each side are on average identical in terms of all pre-treatment variables. There is a 'local randomization' in this window. There is also a discontinuity in the probability of treatment at the threshold between the treatment group and the control. Therefore, if there is a corresponding discontinuity in the intercept of the regression of the outcome variable on the predictor at the threshold level, the change in outcome is attributed to the effect of the intervention.[18] Critically, RDD can provide evidence for the causal effect of an intervention because, controlling for the value of the rating variable, it is possible to account for unobserved differences between the treatment and the control group.[19-21]

RDD a quasi-experimental evaluation method that can be used in cases where it would be impractical or unethical to assign participants to different groups,[17] for example, the impact of prisons on recidivism, and the impact of health insurance on improving service utilisation.[22, 23] In the present study, it is unethical to experimentally assign GP practices to different experimental groups since we have reason to believe that the treatment is effective, given the evidence from our previous trial.[9] Therefore we evaluate the intervention using a sharp RDD, a variation of the RDD where the rating variable perfectly predicts treatment allocation.[24] This means that the probability of treatment changes from 0 to 1 at the threshold. The rating variable was the prescribing indicator (average antibiotic prescribing rate adjusted for STAR-PU) and the threshold was 1.14825 antibiotic items dispensed per STAR-PU, the cut-off point separating the 20% of highest prescribers from the remaining GP practices. We used R,[25] specifically function call RDestimate of package rdd,[26] to build our RDD models.

Bandwidth Size

One of the critical steps in RDD is selecting the bandwidth size around the threshold to create the localised sample.[27] This is because narrower bandwidths increase comparability between cases on each side of the threshold while decreasing the statistical power, whereas wider bandwidths increase power at the cost of decreasing internal validity by including cases further away from the threshold. Hence it is necessary to select the appropriate bandwidth, which finds the balance between precision and power. The Imbens-Kalyanaraman algorithm provides a data driven, asymptotically optimal bandwidth for RDD.[27] The calculated bandwidth is tailored to specific features of the RDD setting. Using package rdd,[26] in R,[25] we calculated the optimal bandwidth size to create the localised sample around the threshold that minimised bias and optimised precision. The optimal bandwidth differed for different models considered. For the most parsimonious model (Table 2) the optimal bandwidth was 0.5234 antibiotic items dispensed per STAR-PU. Thus, practices within 0.5234 points of the threshold in either direction were included in the local linear regression analyses. The number of practices within this bandwidth (n = 6,524), used to calculate the most plausible estimate of the Local Average Treatment Effect (LATE), provided sufficient power to detect a statistically significant effect at $p < 0.001$.

Data

The data for this study came from the Public Health England data warehouse (originally collected by the NHS Business Services Authority and NHS Digital).[28, 29] These data consist of all GP practices in England who were sent the April 2017 CMO feedback letter (n = 1,439), as well as the remaining practices who served as the control group (n = 5,986). To account for differences in prescribing due to seasonality, we defined our baseline measure, which we used to determine the percentile of prescribers that each practice was in, as mean-STAR-PU-adjusted prescribing rate from October 2015 to September 2016. The outcome measure was mean-STAR-PU-adjusted rate of antibiotic items dispensed during the intervention period, April 2017 to September 2017.

Data Preparation

We calculated our baseline and outcome measures using the same quarterly Antibacterial Items/STAR-PU weightings that we used to select the practices for letter allocations. For the analysis we coded all the practices that received the letter (20% of highest prescribers) as the intervention group. The remaining practices, which did not receive the letter because they were in the 21 – 100% of prescribers, were the control group.

We kept the GP practices in the central 99% of the records of both baseline and outcome measures. This means that we removed practices that were classified as outliers due to their extremely high or low rates of prescribing (n = 133), or due to extremely large/small patient populations (n = 57). We also removed practices due to instances of missing prescribing data between October 2015 to September 2017 (n = 231).

Since the CMO letters were sent only to the highest 20% of prescribers, we aimed to set the RDD cut-off threshold at the 5th quintile. However, the threshold was adjusted to 1.1812 from 1.1813 to account for the removal of some practices (n = 9) during data preparation because of practice changes over the intervention period or because of lack of STAR-PU data for the duration of the intervention period. Overall, we removed 430 GP practices. This resulted in the total sample of 6,995 GP practices, 1,378 of which received the letter in April 2017 (treatment group), and 5,617 of which did not (control group).

Results

Table 2 and Figure 1 demonstrate the results of the simplest model, where the baseline measure (average prescribing rate adjusted for STAR-PU during the period October 2015 to September 2016) is regressed on the outcome measure (average prescribing rate adjusted for STAR-PU during the six month of the intervention period). The RDD estimates can be interpreted as weighted average effects of treatment across all GP practices. The LATE estimate, which is calculated using the optimal bandwidth, shows that the letter intervention resulted in a significant discontinuity ($\beta = -0.016$, $SE = 0.003$, $z = -5.141$, $p < 0.001$); the GP practices who received the letter changed their prescribing rates by approximately -3.69% [95% CI = -2.29, -5.10]. However, it is important to acknowledge that deprivation levels could vary between the local areas covered by GP practices in the treatment and the control groups. This variation could affect prescribing rates of GP practices. Thus, we adjusted the estimates for this potential variation in deprivation using deprivation indices.[30] We found that the intervention effect was not sensitive to the inclusion of a deprivation measure (Table 3). Moreover, although the effect size varies depending on the bandwidth used, overall the effect remains robust to halving or doubling the bandwidth and therefore including a different number of observations for both the unadjusted and adjusted estimates. This indicates that our findings are likely to be relatively reliable and valid.

We also built an RDD with a categorical variable specifying GP practices' allocation to one of six letters or the control group, to examine whether the intervention was effective across all letter types. We present the estimates for each letter type, both unadjusted and adjusted for deprivation, in (Table 4). The four letters sent to the top 11-20% (Letters 1A, 1B, 2A, and 2B) were estimated to be relatively effective at reducing antibiotic prescribing (1A: -5.21% [95% CI = -1.98, -8.45]; 1B: -4.75% [95% CI = -2.96, -6.54]; 2A: -2.32% [95% CI = -0.45, -4.18]; 2B: -3.81% [95% CI = -1.86, -5.76]), with the letters sent to those practices whose prescribing was increasing being most effective. However, for the practices in the top 10% (Letters 3A and 3B) our estimates of the intervention in the general population were not clear. This is because the effects of the intervention for the two letter types were not statistically significant (3A: -2.13% [95% CI = 0.76, -5.02]; 3B: 1.47% [95% CI = 7.24, -4.30]).

Based on a predicted 3.69% reduction in antibiotic prescribing for the intervention group over a six-month period (Table 2), an estimated 124,952 fewer antibiotic items were dispensed during the study as a result of the letter intervention. Specifically, we estimated that 3,386,243 antibiotic items would have been dispensed without the intervention by the GP practices in the intervention arm, whereas 3,261,291 antibiotic items were found to have been dispensed with the intervention.

To estimate the effect on direct prescribing costs for the public sector, we considered the basic cost of a drug as used in primary care during the intervention period, specifically the average net ingredient cost (NIC) of £8.29 per antibiotic item.[31] The NIC does not take into account dispensing costs, fees or prescription charge income. Thus, we assumed that 10.2% of antibiotics (12,745) incurred a £8.60 prescription charge (in line with the overall exemption rate for April to September 2017), which we deducted from the cost estimate.[31-33] We also added the £1.25 per item professional fee payable to pharmacy contractors by NHS England in 2017.[32, 33] The cost of printing and mailing to 6,318 GPs in 1,439 practices, which was deducted from the savings estimate to produce a net savings estimate, was £5,527. This equates to £0.04 per prescription prevented during the study period. Overall, the predicted reduction in the number of prescribed antibiotics during the intervention is estimated to have generated net savings of £1,076,908 (£991,842 if adjusted for practice deprivation) in NIC and dispensing costs for the public sector.

Discussion

RDD analysis shows that the 2017 CMO feedback letter reduced antibiotic prescribing by 3.69%, which is similar to the 3.3% reduction in prescribing that was found in the 2014 RCT.[9] The finding that repeated feedback continues to have an effect on antibiotic prescribing is consistent with other trials that have used repeated feedback, in medicine and in other policy areas.[34, 35] A systematic review finds that repeated feedback is more effective than single instances in healthcare settings.[36] However, these studies tend to analyse the repeated feedback as a single intervention. To our knowledge, this is the first study that investigates whether a successful antibiotic feedback intervention can be as effective when repeated. Our results suggest that the social-norm feedback intervention can be successfully implemented multiple times.

There were many anti-microbial stewardship interventions happening in the UK in 2017-8. For instance, the Quality Premium gave financial incentives for reducing inappropriate prescribing; the TARGET toolkit was developed, a suite of resources that health-care workers can use to support patients to self-care; the Antibiotic Guardian campaign encouraged professionals to pledge to prescribe appropriately; and the Keep Antibiotic Working campaign was a patient facing advertising campaign.[37] Therefore, it may not be a surprise that antibiotic prescribing in primary care is decreasing.[37] However, the RDD is a quasi-experimental method of evaluation and the difference between the groups within the bandwidth is whether or not they received a letter, so we can attribute the 3.69% difference in prescribing between groups to the letter.

The intervention decreased prescribing amongst practices in the top 11-20%, especially those whose prescribing had been increasing in the year beforehand; this decrease was observed both in practices that had received a letter in the previous year and in practices that had not received a letter. However, it did not have a significant effect on practices in the top 10%. There are two types of possible explanation for this difference. First, there may have been something specific to the practices in the top 10%. Our adjustments accounted for the age and sex of the practices' patient population (STAR-PU) and their deprivation. However, there may have been other relevant factors that affect antibiotic prescribing, for instance prevalence of Chronic Obstructive Pulmonary Disorder or smoking, which were not captured in our models. Second, there may have been something about the message, which was more extreme for the top 10%. Telling GPs that their practice was in the top 10% of prescribers may have made their prescribing seem so disproportionate that they felt they could not do anything about it. The forceful message may have provoked psychological reactance, leading to negative attitudes towards the message and the generation of counterarguments, resulting in a lower behavioural intention to comply with the message.[38-40] Alternatively, while highly credible sources are more persuasive the more discrepant the receiver finds the message, moderately credible sources may be less persuasive when they send a highly discrepant message. If the CMO is only perceived as a moderately credible source, then the highly discrepant message received by the top 10% of prescribers may have been less persuasive than the less discrepant message received by the top 20%.[41, 42]

The intervention has already been used successfully in Australia, Ireland, and Canada, and France is planning to follow suit.[11-15] The countries that are implementing the intervention are fairly similarly culturally, they are all Western Educated Industrialized Rich Democracies (WEIRD).[43] It is possible that the effect of descriptive norms and therefore the international applicability of the intervention might vary between cultures. The strength of social norms in countries is thought to vary along a spectrum that ranges from 'tight' (have many strong norms and a low tolerance of deviant behaviour) to 'loose' (have weak social norms and a high tolerance of deviant behaviour), and the UK, Australia, and France are in the middle of the spectrum.[44, 45] We might anticipate that the feedback letter intervention would be least effective in the countries with the loosest cultures (Estonia, Hungary, and Ukraine) and most effective in countries with the tightest cultures (India, Malaysia, Pakistan, Singapore, and South Korea). However, this is very speculative because the way that primary care is delivered, and that antibiotic prescribing occurs, may differ in these countries and because the UK has lower antibiotic prescribing than many countries.[46]

Our study has several limitations. First, the STAR-PU data is calculated on a quarterly rather than a monthly basis. Therefore, it is not currently possible to detect the exact extent of the discontinuity in a specific month of an intervention. The strength of the intervention may differ over the course of the outcome, but this may be masked by the three-month average of the outcome variable. Thus, we can conclude that there was a difference in prescribing over the entire evaluation period, but we are not able to detect time trends, if they exist. Future studies, conditional on fine-grained longitudinal data, could utilise ITS to address that.

Second, the RDD tends to be less precise than a comparable RCT. An RDD sample has to be at least 2.4 times greater than that of an equivalent randomised trial in order to achieve the same level of precision.[24] Moreover, the precision of an RDD is sensitive to the distribution of scores around the threshold. The non-parametric method of calculating the bandwidth uses observations close to the threshold.[27] Therefore, although our data-driven bandwidth choice is claimed to be optimal, it ignores the observations outside of the bandwidth. This is why non-parametric methods tend to be less precise than parametric methods for a given sample.[24] In the present study, we accounted for this by looking at the estimates of different RDD models given different bandwidths.

In conclusion, our results suggest that it could be worthwhile for antimicrobial stewardship programmes to incorporate regular social-norm feedback into their activities, as a part of their strategy. We found that, on average feedback was effective, but it was not effective for the top 10% of prescribing practices, and future research could investigate why. Individual-level feedback is likely to be even more effective, but data by prescriber are not yet centrally available in England. The effectiveness of social-feedback may also generalize to other domains, including antibiotic prescribing that occurs outside of primary care and other areas of medicine where evidence suggests that there may be high levels of inappropriate prescribing or over-use of clinical tests.

Funding
Public Health England.

Transparency declaration
Natalie Gold affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained. There are no conflicts of interest to declare.

Research Ethics

This study used publicly available data to evaluate an intervention, which was not an RCT, so research ethics were not required.

Data Availability Statement

The raw data are publicly and freely available from Fingertips, https://fingertips.phe.org.uk/. The cleaned dataset is available from the corresponding author upon reasonable request.

Acknowledgments

We would like to thank Professor Dame Sally Davies for her support in this project. Rohan Arambepola and Dr. Simon Taylor provided useful comments on the manuscript. The study was funded by Public Health England, where the authors were affiliated when doing this work. No other pharmaceutical company or agency paid for the work. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. The paper was presented at the 14th Annual Meeting of the Society for Behavioural Medicine, Reference 87.

References

[1]World Health Organization 2014. *Antimicrobial resistance: global report on surveillance.* https://apps.who.int/iris/bitstream/handle/10665/112642/9789241564748_eng.pdf?sequence=1

[2]The Review on Antimicrobial Resistance 2016. *Tackling Drug-Resistant Infections Globally: final report and recommendations*. https://amr-review.org/sites/default/files/160525_Final%20paper_with%20cover.pdf

[3]Davies SC, Fowler T, Watson J *et al*. Annual Report of the Chief Medical Officer: infection and the rise of antimicrobial resistance. *Lancet* 2013; **381**: 1606-9.

[4]Public Health England 2015. *Behaviour change and antibiotic prescribing in healthcare settings: Literature review and behavioural analysis*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/774129/Behaviour_Change_for_Antibiotic_Prescribing_-_FINAL.pdf

[5]Hawker JI, Smith S, Smith GE *et al*. Trends in antibiotic prescribing in primary care for clinical syndromes subject to national recommendations to reduce antibiotic resistance, UK 1995-2011: analysis of a large database of primary care consultations. *J Antimicrob Chemother* 2014; **69**: 3423-30.

[6]Public Health England 2014. *English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR)*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/362374/ESPAUR_Report_2014__3_.pdf

[7]Smieszek T, Pouwels KB, Dolk FCK *et al*. Potential for reducing inappropriate antibiotic prescribing in English primary care. *J Antimicrob Chemother* 2018; **73** Suppl 2: ii36-43.

[8]HM Government 2014. *UK five year antimicrobial resistance (AMR) strategy 2013-2018: annual progress report and implementation plan.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/385733/UK_AMR_annual_report.pdf

[9]Hallsworth M, Chadborn T, Sallis A, *et al*. Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *Lancet* 2016; **387**: 1743–52.

[10]Royal College of General Practicioners 2019. *TARGET Antibiotic Toolkit.* https://www.rcgp.org.uk/clinical-and-research/resources/toolkits/target-antibiotic-toolkit.aspx

[11]Australian Government, Department of Health 2018. *Nudge vs Superbugs: A behavioural economics trial to reduce the overprescribing of antibiotics.* https://www1.health.gov.au/internet/main/publishing.nsf/Content/1598E371849EEA83CA2582B000062D9C/$File/Nudge%20vs%20Superbugs%20-%20A%20behavioural%20economics%20trial%20to%20reduce%20the%20overprescribing%20of%20antibiotics%20June%202018.pdf

[12]Public Health Agency 2017. *Surveillance of Antimicrobial Use and Resistance in Northern Ireland, Annual Report.* https://www.publichealth.hscni.net/sites/default/files/AMR_annual_report_final_0.pdf

[13]Government of Canada 2019. *Evidence synthesis summary: interventions to address antimicrobial use.* https://www.canada.ca/en/public-health/services/publications/drugs-health-products/interventions-address-antimicrobial-use.html

[14]Ontario Agency for Health Protection and Promotion 2019. *Providing Antibiotic Prescribing Feedback to Primary Care Physicians: The Ontario Program To Improve AntiMicrobial USE (OPTIMISE).* https://www.clinicaltrials.gov/ct2/show/study/NCT03776383?show_desc=Y#desc

[15]Behavioural Insights Team and UK Science & Innovation Network in France 2019. *Influencing behaviours through science: UK-France perspectives.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/799398/FCO-BIT_conference_report_02MAy19.odt

[16]Gould IM, Lawes T. Antibiotic stewardship: prescribing social norms. *Lancet* 2016; **387**:1699-701.

[17]HM Treasury 2011. *The Magenta Book: Guidance for evaluation.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

[18]Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *J Educ Psychol* 1960; **51**: 309-17.

[19]Lee DS, Lemieux T. Regression Discontinuity Designs in Economics. *J Econ Lit* 2010; **48**: 281-355.

[20]MDRC 2013. *The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation*. https://www.mdrc.org/sites/default/files/validity_precision_comparative_interrupted_time_series_design.pdf

[21]Gelman A, Imbens G. Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *J Bus Econ Stat* 2019; **37**: 447-56.

[22]Mitchell O, Cochran JC, Mears DP *et al*. Examining Prison Effects on Recidivism: A Regression Discontinuity Approach. *Justice Q* 2017; **34**: 571-96.

[23]Palmer M, Mitra S, Mont D *et al.* The impact of health insurance for children under age 6 in Vietnam: A regression discontinuity approach. *Soc Sci Med* 2015; **145**: 217-26.

[24]MDRC 2012. *A Practical Guide to Regression Discontinuity*. https://www.mdrc.org/sites/default/files/regression_discontinuity_full.pdf

[25]R Core Team 2018. *R: A Language and Environment for Statistical Computing.* http://www.R-project.org/

[26]Dimmery D 2016. *Package 'rdd'*. https://cran.r-project.org/web/packages/rdd/rdd.pdf

[27]Imbens G, Kalyanaraman K. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. Rev Econ Stud 2012; **79**: 933–59.

[28]NHS Business Services Authority 2019. *Prescribing data.* https://www.nhsbsa.nhs.uk/prescription-data/prescribing-data

[29]NHS Digital 2019. *GP and practice related data.* https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data
[30]National Statistics 2015. *English indices of deprivation*. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015
[31]NHS Digital 2018. *Prescriptions Dispensed in the Community - Statistics for England, 2007-2017*. https://files.digital.nhs.uk/14/832F9A/PDitC_2017_Appendix_Tables.xls
[32]NHS Business Services Authority 2017. *Amendments to the drug tariff April 2017*. https://www.nhsbsa.nhs.uk/sites/default/files/2017-04/April_2017.pdf
[33]NHS Business Services Authority 2017. *Amendments to the drug tariff September 2017*. https://www.nhsbsa.nhs.uk/sites/default/files/2017-08/September%202017.pdf
[34]Guthrie B, Kavanagh K, Robertson C *et al*. Data feedback and behavioural change intervention to improve primary care prescribing safety (EFIPPS): multicentre, three arm, cluster randomised controlled trial. *BMJ* 2016; **354**: i4079.
[35]Tiefenbeck V, Tasic V, Schöb S *et al*. Long-lasting effects or Short-Term Spark? On the Persistence of behaviour Change induced by Real-Time feedback on Resource Consumption. In: *Research Papers of Twenty-Fourth European Conference on Information Systems (ECIS), İstanbul, Turkey, 2016*. Research Paper 84. Association for Information Systems, Atlanta, Georgia, USA.
[36]Ivers N, Jamtvedt G, Flottorp S *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012; **6**: 1-229.
[37]Public Health England 2018. *English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR)*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/759975/ESPAUR_2018_report.pdf
[38]Brehm JW. *A theory of psychological reactance*. Oxford: Academic Press, 1966.
[39]Dillard JP, Shen L. On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs* 2005; **72**: 144–68.
[40]Steindl C, Jonas E, Sittenthaler S *et al*. Understanding Psychological Reactance: New Developments and Findings. *Z Psychol* 2015; **223**: 205–14.
[41]Aronson E, Turner JA, Carlsmith JM. Communicator credibility and communication discrepancy as determinants of opinion change. *J Abnorm Soc Psychol* 1963, **67**, 31-6.
[42]Pornpitakpan C. The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *J Appl Soc Psychol* 2004, **34**, 243-81.
[43]Henrich J, Heine SJ, Norenzayan A. Most people are not WEIRD. *Nature* 2010; **466**: 29.
[44]Gelfand MJ, Harrington JR. The Motivational Force of Descriptive Norms: For Whom and When Are Descriptive Norms Most Predictive of Behavior? *J Cross Cult Psychol* 2015; **46**: 1273-78.
[45]Gelfand MJ, Raver JL, Nishii L *et al.* Differences Between Tight and Loose Cultures: A 33-Nation Study. *Science* 2011; **332**, 1100-4.
[46]World Health Organization 2018. *WHO report on surveillance of antibiotic consumption: 2016-2018 early implementation.* https://www.who.int/medicines/areas/rational_use/who-amr-amc-report-20181109.pdf?ua=1

Table 1. Key differences between the six letter variants

	Previous letter	**No previous letter**
	Paragraph 2 starts: I have written to your practice previously and…	
Top 20%, prescribing not increasing *Header*: Your practice is amongst the 20% highest prescribers of antibiotics nationally *Between Paragraph 1 and Paragraph 2:* The great majority (80%) of practices in England prescribe fewer antibiotics per head than yours. Most other practices have reduced their prescribing rates since 2013/14 but yours has increased.	1A (Figure S1)	1B (Figure S2)
Top 20%, prescribing increasing *Header:* Your practice is amongst the 20% highest prescribers of antibiotics nationally *Between Paragraph 1 and Paragraph 2:* The great majority (80%) of practices in England prescribe fewer antibiotics per head than yours.	2A (Figure S3)	2B (Figure S4)
Top 10% *Header:* Your practice is amongst the 10% highest prescribers of antibiotics nationally *Between Paragraph 1 and Paragraph 2:* The great majority (90%) of practices in England prescribe fewer antibiotics per head than yours.	3A (Figure S5)	3B (Figure S6)

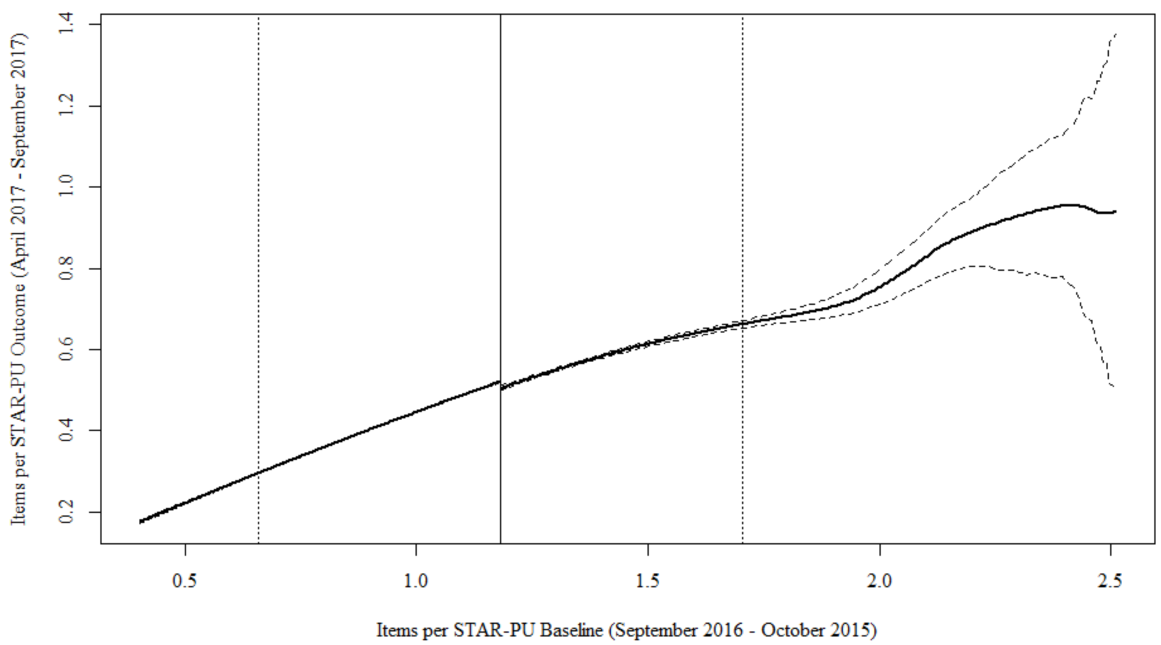
Table 2. Difference in prescribing rates between treatment and control

	Bandwidth	Number of observations	Estimate [95% CIs]	*Standard error*	*z*-value	*P-value*	Effect [95% CIs]	Predicted Change in Dispensed Items
LATE*	0.523	6524	-0.016 [-0.010, - 0.023]	0.003	-5.141	<0.001	-3.69% [-2.29, -5.10]	-124952 [-77544, -172698]
Half-bandwidth	0.262	4294	-0.015 [-0.007, -0.023]	0.004	-3.832	<0.001	-3.42% [-1.67, -5.17]	-115810 [-56550, -175069]
Double-bandwidth	1.047	6991	-0.014 [-0.008, -0.020]	0.003	-4.725	<0.001	-3.19% [-1.86, -4.51]	-108021 [-62984, - 152720]

	F-statistic	Numerator degrees of freedom	Denominator degrees of freedom	*P-value*
LATE*	2185	3	6520	<0.001
Half-bandwidth	461	3	4290	<0.001
Double-bandwidth	4205	3	6987	<0.001

* LATE = Local Average Treatment Effect and represents the optimal bandwidth.

Figure 1. Discontinuity based on the most parsimonious model



Note. The solid vertical black line represents the discontinuity threshold between the practices that received the letter (right side of the line), versus the practices that did not (left side of the line). The two dotted lines represent the LATE bandwidth thresholds. The solid black regression line represents the point-based estimates of the RDD, whereas the two broken lines on either side of it represent 95% confidence intervals. Only 22 GP practices in the 10% of highest prescribers were outside of the optimal bandwidth thresholds, indicating that the bandwidth thresholds are unlikely to be the reason as to why the effects of the intervention were not detected among these practices.

Table 3. Deprivation adjusted difference in prescribing rates between treatment and control

	Bandwidth	Number of observations	Estimate [95% CIs]	*Standard error*	*z*-value	*P-value*	Effect [95% CIs]	Predicted Change in Dispensed Items
LATE*	0.302	4814	-0.015 [-0.008, -0.022]	0.004	-4.105	<0.001	-3.40% [-1.78, -5.02]	-115132 [-60275, -169989]
Half-bandwidth	0.151	2546	-0.016 [-0.006, -0.026]	0.005	-3.081	0.002	-3.53% [-1.29, -5.78]	-119534 [-43683, -195725]
Double-bandwidth	0.604	6744	-0.015 [-0.009, -0.021]	0.003	-5.041	<0.001	-3.46% [-2.17, -4.81]	-117164 [-73481, -162878]
	***F*-statistic**	**Numerator degrees of freedom**	**Denominator degrees of freedom**	***P-value***				
LATE*	530.6	4	4809	<0.001				
Half-bandwidth	69.76	4	2541	<0.001				
Double-bandwidth	2107.79	4	6739	<0.001				

* LATE = Local Average Treatment Effect and represents the optimal bandwidth. Adjusted estimates are altered by the inclusion of a deprivation covariate (IMD-2015-based measure of population weighted average of the Lower Layer Super Output Areas covered by each practice).[30]

Table 4. Deprivation adjusted difference in prescribing rates between treatment and control by letter

Letter type (n)	LATE*	Bandwidth	Number of observations	Estimate [95% CIs]	*Standard error*	*z*-value	*P-value*	Effect [95% CIs]	Predicted Change in Dispensed Items
1A	Unadjusted	0.523	6524	-0.023 [-0.009, -0.037]	0.007	-3.157	0.002	-5.21% [-1.98, -8.45]	-17303 [-6576, -28063]
(n = 127)	Adjusted	0.302	4814	-0.026 [-0.011, -0.041]	0.008	-3.322	<0.001	-5.84% [-2.39, -9.28]	-19395 [-7937, -30820]
1B	Unadjusted	0.523	6524	-0.021 [-0.013, -0.029]	0.004	-5.205	<0.001	-4.75% [-2.96, -6.54]	-40416 [-25186, -55647]
(n = 352)	Adjusted	0.302	4814	-0.019 [-0.011, -0.028]	0.004	-4.398	<0.001	-4.41% [-2.44, -6.37]	-37523 [-20761, -54200]
2A	Unadjusted	0.523	6524	-0.010 [-0.002, -0.018]	0.004	-2.433	0.015	-2.32% [-0.45, -4.18]	-16822 [-3263, -30308]
(n = 290)	Adjusted	0.302	4814	-0.009 [-0.0003, -0.018]	0.005	-2.026	0.043	-2.10% [-0.07, -4.13]	-15227 [-508, -29946]
2B	Unadjusted	0.523	6524	-0.017 [-0.008, -0.025]	0.004	-3.830	<0.001	-3.81% [-1.86, -5.76]	-28383 [-13856, -42910]
(n = 289)	Adjusted	0.302	4814	-0.014 [-0.004, -0.023]	0.005	-2.898	0.004	-3.14% [-1.02, -5.27]	-23392 [-7599, -39260]
3A	Unadjusted	0.523	6524	-0.009 [0.003, -0.022]	0.007	-1.442	0.150	-2.13% [0.76, -5.02]	-12445 [4440, -29330]
(n = 250)	Adjusted	0.302	4814	-0.007 [0.007, -0.021]	0.007	-0.932	0.352	-1.51% [1.67, -4.68]	-8822 [9757, -27344]
3B	Unadjusted	0.523	6524	0.007 [0.032, -0.019]	0.013	0.500	0.617	1.47% [7.24, -4.30]	2190 [10784, -6405]
(n = 70)	Adjusted	0.302	4814	0.006 [0.031, -0.020]	0.013	0.448	0.654	1.31% [7.04, -4.42]	1951 [10486, -6583]

* LATE = Local Average Treatment Effect calculated using the optimal bandwidth for a given RDD. Adjusted estimates are altered by the inclusion of a deprivation covariate (IMD-2015-based measure of population weighted average of the Lower Layer Super Output Areas covered by each practice).[30] n = number of practices that received each letter.