



Variation in responsiveness to warranted behaviour change among NHS clinicians: novel implementation of change detection methods in longitudinal prescribing data

Alex J Walker,¹ Felix Pretis,^{2,3} Anna Powell-Smith,¹ Ben Goldacre¹

¹The DataLab, Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Oxford OX2 6GG, UK

²Department of Economics, University of Victoria, Victoria, BC, Canada

³Institute for New Economic Thinking, Oxford Martin School, University of Oxford, Oxford, UK

Correspondence to: B Goldacre ben.goldacre@phc.ox.ac.uk (or @bengoldacre on Twitter ORCID 0000-0002-5127-4728)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2019;367:l5205 <http://dx.doi.org/10.1136/bmj.l5205>

Accepted: 27 July 2019

ABSTRACT

OBJECTIVES

To determine how clinicians vary in their response to new guidance on existing or new interventions, by measuring the timing and magnitude of change at healthcare institutions.

DESIGN

Automated change detection in longitudinal prescribing data.

SETTING

Prescribing data in English primary care.

PARTICIPANTS

English general practices.

MAIN OUTCOME MEASURES

In each practice the following were measured: the timing of the largest changes, steepness of the change slope (change in proportion per month), and magnitude of the change for two example time series (expiry of the Cerazette patent in 2012, leading to cheaper generic desogestrel alternatives becoming available; and a change in antibiotic prescribing guidelines after 2014, favouring nitrofurantoin over trimethoprim for uncomplicated urinary tract infection (UTI)).

RESULTS

Substantial heterogeneity was found between institutions in both timing and steepness of change. The range of time delay before a change was implemented was large (interquartile range 2-14 months (median 8) for Cerazette, and 5-29 months

(18) for UTI). Substantial heterogeneity was also seen in slope following a detected change (interquartile range 2-28% absolute reduction per month (median 9%) for Cerazette, and 1-8% (2%) for UTI). When changes were implemented, the magnitude of change showed substantially less heterogeneity (interquartile range 44-85% (median 66%) for Cerazette and 28-47% (38%) for UTI).

CONCLUSIONS

Substantial variation was observed in the speed with which individual NHS general practices responded to warranted changes in clinical practice. Changes in prescribing behaviour were detected automatically and robustly. Detection of structural breaks using indicator saturation methods opens up new opportunities to improve patient care through audit and feedback by moving away from cross sectional analyses, and automatically identifying institutions that respond rapidly, or slowly, to warranted changes in clinical practice.

Introduction

Medicine is characterised by the development of new interventions, and new information on existing interventions. This progress requires that clinical practice changes in response to updated evidence on effectiveness, safety, and cost. The diffusion of innovation is a longstanding area of research, originating with 1950s work on agriculture¹ and antibiotics.² Previous work has largely focused on narrative descriptions, discussing the nature of the innovation (its relative advantage, compatibility, and complexity to implement); the channels through which the innovation is communicated; and the so-called social system that is involved in implementing the innovation.³ Previous quantitative work has relied on the manual characterisation of individuals and organisations as either adopting, or not adopting, a new intervention.^{1,2} Typically, the rate of adoption is variable over time, starting with a small number of initial early adopters followed by a large number of institutions rapidly adopting the change, and then followed by a slower rate while so-called laggards adopt the change over a longer period.³

Diffusion of innovations has received some attention in healthcare⁴ but research so far has primarily focused on case studies,⁵ narrative descriptions of clinicians' responses to change in guidance,⁶ interviews,⁷⁻⁹ and theoretical frameworks.¹⁰ Previous quantitative work assessing implementation of new practices has typically relied on measuring change at the level of a whole population, using techniques such as

WHAT IS ALREADY KNOWN ON THIS TOPIC

Implementation of new evidence is critical in a well performing healthcare system

Speed of diffusion of innovations in healthcare is thought to vary but previous work has focused on small samples, narrative descriptions, or cross sectional analysis at one time point

WHAT THIS STUDY ADDS

In two example measures of clinical behaviour (based on changes in contraceptive and antibiotic prescribing), substantial variation was detected between general practices in the timing and slope of change in clinical practice in England

The detection method can automatically and robustly detect the timing and magnitude of changes in clinical behaviour across thousands of individual institutions

This method creates new opportunities to improve patient care through audit and feedback, by moving away from cross sectional analyses and automatically identifying institutions who respond rapidly, or slowly, to warranted changes in clinical practice

interrupted time series analysis^{11 12} or static measures of variation in care at one point in time through atlases of variation and regression analyses.¹³⁻¹⁹

Assessing variation between institutions in timing of implementation for new clinical behaviours requires a systematic and robust method to identify when institutions have made a change. As it is not feasible to manually review thousands of time series charts to determine when meaningful change has occurred, this review must be done computationally. Statistical methods for the detection of structural change (known as break detection) provide a robust method of detecting the timing of changes in time series data without imposing an intervention or change date a priori.²⁰ These techniques have previously been applied to a diverse range of applications, including economic and climate modelling.^{21 22}

We therefore set out to determine how clinicians vary in the timing of their response to new guidance. To achieve this objective, we repurposed and adapted statistical break detection techniques based on indicator saturation for use in medical time series data. Here, we report the deployment of these methods to assess variation in speed of adoption for two examples of warranted change in clinical practice: firstly, the move from branded to generic versions of the oral contraceptive, desogestrel, in 2012, saving the health service about £10m (€11.1m; \$12.3m) a year²³; and secondly, the change from trimethoprim to nitrofurantoin as the firstline antibiotic for treating uncomplicated urinary tract infection (UTI) at various time points after 2014.

Methods

Data

The monthly prescribing datasets, published by the NHS Business Services Authority, contain one row for each treatment and dose, in each prescribing organisation in NHS primary care in England, describing the number of prescriptions issued and the total cost. To extract data on standard general practices, we limited them to institutions with setting code 4: general practices,²⁴ excluding all other organisations, such as dentists, prisons, and walk-in centres. We excluded data for a measure in any practice where the time series had more than half of its values missing. Missing values were either caused by small numbers leading to months where the denominator was 0 or by the practice not being open for part of the time series (eg, due to closing). We also excluded practices where prescribing did not vary during follow-up time because practices where the proportion of prescriptions stayed constant throughout the sample cannot have any change points.

Desogestrel measure

We measured the total proportion of desogestrel prescriptions that were prescribed as the branded Cerazette. A decrease in this proportion would correspond to an improvement in this measure. The time series for this measure ran from October 2010 to

December 2015. This timing was chosen to centre the data on the time period surrounding the expiry of the Cerazette patent in December 2012. Before the patent expiry, it was still possible to prescribe desogestrel generically but all dispensing would be of branded Cerazette.

Trimethoprim/nitrofurantoin measure

We measured the proportion of trimethoprim prescriptions as a proportion of total trimethoprim and nitrofurantoin prescriptions. A decrease in this proportion would correspond to an improvement in this measure. The time series for this measure ran from June 2013 to June 2018. This timing was done to centre the data on the time period surrounding the following interventions: the change in antibiotic prescribing guidance in October 2014; followed by the introduction of a “quality premium” financial incentive, which was announced in October 2016 and implemented in April 2017.

Detection

We used trend indicator saturation,²⁰ a modified version of indicator saturation,²⁵ in each practice's time series to determine any statistically significant change in prescribing behaviour.²⁵ We formulated the detection of breaks as a model selection problem where a time series regression model of the prescribing behaviour is saturated with a full set of step functions interacted with a linear time trend. We selected over these break functions at every point in time, removing all non-significant breaks at a chosen level of significance (in this case, $P=0.000001$) to tightly control the false positive rate. Step shift (or cliff-like) changes in behaviour can be approximated by a single breaking trend with a high coefficient on the slope while gradual, smooth transition behaviour²⁶ can be approximated through a series of multiple broken linear trends with smaller slope coefficients.

To assess whether the methods for break detection were operating as expected, graphs of the time series for each individual practice were manually inspected, and plotted along with the fitted regression model and detected changes. One hundred randomly sampled graphs from each time series were inspected in detail by two blinded researchers independently to ensure that the automatically detected break points overall reflected a true change in prescribing behaviour, with each giving a narrative description of any issues raised. All remaining graphs were rapidly reviewed to check for gross errors in automated detection.

Indicators of change

We generated three indicators to describe the response in prescribing behaviour of each practice.

- **Timing:** the timing of a change in behaviour is measured as the start of the steepest negative (downward) shift in a time trend of prescribing behaviour during the time series. This measure

captures how long it takes a practice to begin to show a substantial change in behaviour in relation to a stimulus (in these examples, a medicine patent expiry and a change in clinical guidance).

- **Magnitude:** the magnitude of change describes the extent to which each practice reduces the prescribing of the non-favoured drug treatment. This measure is calculated by subtracting the proportion of unfavourable prescribing at the end of the study time series from the proportion of unfavourable prescribing at the start time of the first detected change.
- **Slope:** the steepness of the detected changes measures the pace of change per month within a practice (sudden or gradual) once change has begun.

Multiple break points might be detected in one practice: we therefore limited the model to report the steepest contiguous segment contributing at least 50% to the total level change.

Data sharing

Data management was carried out using SQL (in Google BigQuery), Python, and R. Break detection was implemented using the R package *gets*.²⁰ Complete code and data are provided online on Github (https://github.com/ebmdatalab/change_detection/releases/tag/0.1), and code is also available as a python library (https://pypi.org/project/change_detection/).

Patient and public involvement

We run OpenPrescribing.net, an openly accessible data explorer for all NHS England primary care prescribing data, which receives a large volume of user feedback from professionals, patients, and the public. This feedback is used to refine and prioritise our informatics tools and research activities. Patients were not formally involved in developing this specific study design.

Results

Data

A total of 8078 practices were included in the study overall; 259 practices were excluded from the desogestrel analyses and 398 from the UTI antibiotics analysis because of incomplete time series. One practice was removed from the desogestrel measure because of every value being 1.0. Practices were dropped mainly because of missing values, which is typically a consequence of low prescribing volume. Excluded practices were typically much smaller: mean patient list size for excluded practices was 1861 for the desogestrel measure and 3408 for the UTI antibiotic measure (while the national mean list size was 7078).

Detection

Figure 1 shows examples of practice time series for the desogestrel measure, illustrating the three indicators of change. The timings of detected breaks (the steepest substantial negative shift) are marked as a vertical dashed blue line. The segments over

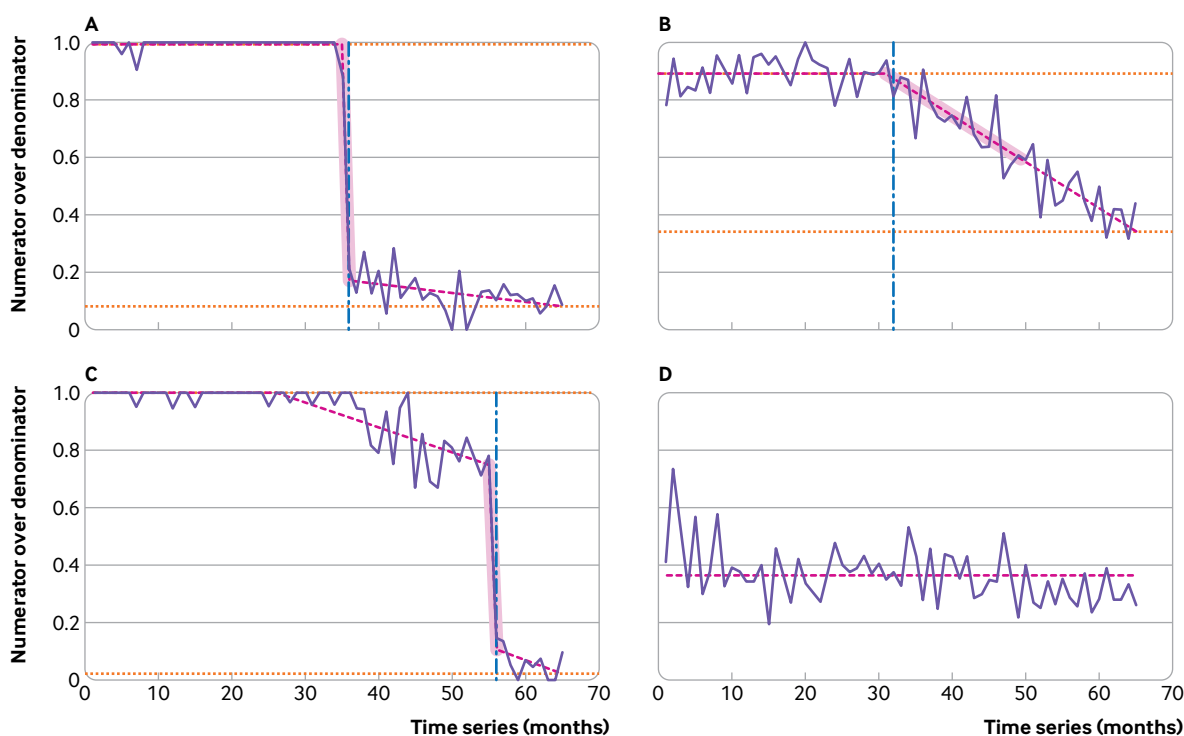


Fig 1 | Examples of practice time series and measured attributes for desogestrel change (from branded to generic). Proportion of Cerazette relative to total desogestrel prescribed of four representative general practices is shown in solid purple lines. Fitted model and detected breaks using trend indicator saturation are shown in pink dashed lines. Commencement of the largest negative shift is marked with a vertical dashed blue line; additional breaks are indicated by changes in the slope of the pink dashed line. The measured slope is highlighted in pink shaded areas, and the pre-break level and final level at the end of the sample are indicated by horizontal orange dotted lines

which the average slope is calculated are shaded in the figure. The magnitude of change is calculated as the difference between the horizontal dotted orange lines. Figure 1A shows a practice where a steep, cliff-like change is detected, followed by a change to a more gradual decline while figure 1B shows a single gradual detected change. Figure 1C shows a practice where an early gradual change is detected followed by a steeper change: as above, for our descriptive analysis, we report timing, slope, and magnitude for the break point contributing to the largest change in practice. For the practice in figure 1D, no changes were detected that reached the necessary significance level ($P=0.000001$).

During the process of manual inspection of 200 randomly selected graphs, a bug was found and fixed whereby if the initial variance of the time series was very low (eg, if a practice prescribed 100% branded Cerazette for many months initially) the technique would become hypersensitive to change, leading to inappropriate detection. We fixed this problem by tweaking one of the parameters of the change detection algorithm away from the default (the maximum size of the block partitioning²⁰). The algorithm was otherwise found to be operating as expected: of 200 time series reviewed, we found two cases of suboptimal detection and four cases of arguable/borderline suboptimal detection. The time series examined, and manual checking datasheet, can be seen in supplementary files A and B.

Indicators of change

Table 1 summarises the detected heterogeneity in prescribing behaviour across all practices, for both measures, with summary statistics over the three estimated measures. For the desogestrel and UTI antibiotic measures, 1711 (22%) and 1380 (18%) practices showed no significant downward changes, respectively.

Timing measurements

For both measures, heterogeneity was considerable between practices in the timing of their largest response

to the warranted change in practice. The top panels of figure 2 and figure 3 show the distribution of the largest detected changes for each measure. Changes were detected across the whole range of the time series. Practices tended to respond more quickly, and with less variation, for the desogestrel measure than the UTI antibiotic measure. For the desogestrel measure, the largest peak in detected changes occurred a few months after expiry of the Cerazette patent. In contrast, relatively few changes were detected in the months following the UTI antibiotic guidance change, with the peak in detected changes not occurring until after the announcement of the quality premium financial incentive.

Slope measurements

The slope of the detected change was also highly variable between general practices (second panels in fig 2 and fig 3), especially for the desogestrel measure, which showed a greater than 10-fold difference in the slope of change between the practice at the 25th centile and the 75th centile (table 1). For the desogestrel measure, the steepness of the change was substantially greater following expiry of the Cerazette patent, indicating that those practices changing later typically did so more rapidly. The mean slope of the detected change for the trimethoprim/nitrofurantoin measure was generally much lower, indicating slower change in practice; the mean slope only substantially increased following implementation of the quality premium financial incentive in April 2017. The relation between timing and slope of change is illustrated in supplementary figures S1 and S2.

Magnitude measurements

The level of heterogeneity in the magnitude of change was less than that for timing or slope of the change (third panels in fig 2 and fig 3). Heterogeneity was variable over time for the desogestrel measure (fig 2) but uniform over time for the trimethoprim/nitrofurantoin measure (fig 3).

Discussion

Summary

The indicator saturation method was successfully implemented to detect meaningful changes in clinical practice. Among general practices in the English health system, we described substantial heterogeneity in the timing and slope of warranted changes in clinical practice following changes in price and clinical guidance on two commonly prescribed treatments: an oral contraceptive and the choice of antibiotic for UTI.

Interpretation

The changes measured in this study were highly warranted from a cost effectiveness or clinical perspective, as illustrated by the fact that most practices eventually showed a substantial change in clinical practice. However, the distribution of the measures of timing, slope of change and, to a lesser extent, magnitude, showed high variation and skewness. While a large proportion of practices showed a significant shift

Table 1 | Summary of detected changes in prescription behaviour for two prescribing measures, across all general practices (move from branded to generic versions of desogestrel in 2012 and change from trimethoprim to nitrofurantoin as the firstline antibiotic for treating uncomplicated urinary tract infection (UTI) at various time points after 2014)

Prescribing measure	Timing (months from intervention)	Steepness (% change per month)	Magnitude (% change)
Desogestrel			
Median	8	8.9	65.9
Interquartile range	2-14	2.0-28.3	43.7-85.0
Mean	6.7	18.7	61.7
Standard deviation	12.5	24.2	27.6
UTI antibiotics*			
Median	18	2.1	37.7
Interquartile range	5-29	1.0-7.5	28.4-47.0
Mean	15.8	6.9	37.9
Standard deviation	14.3	12.4	14.7

*For the UTI measure, timing is in relation to the first intervention (October 2014, the change in Public Health England guidance).

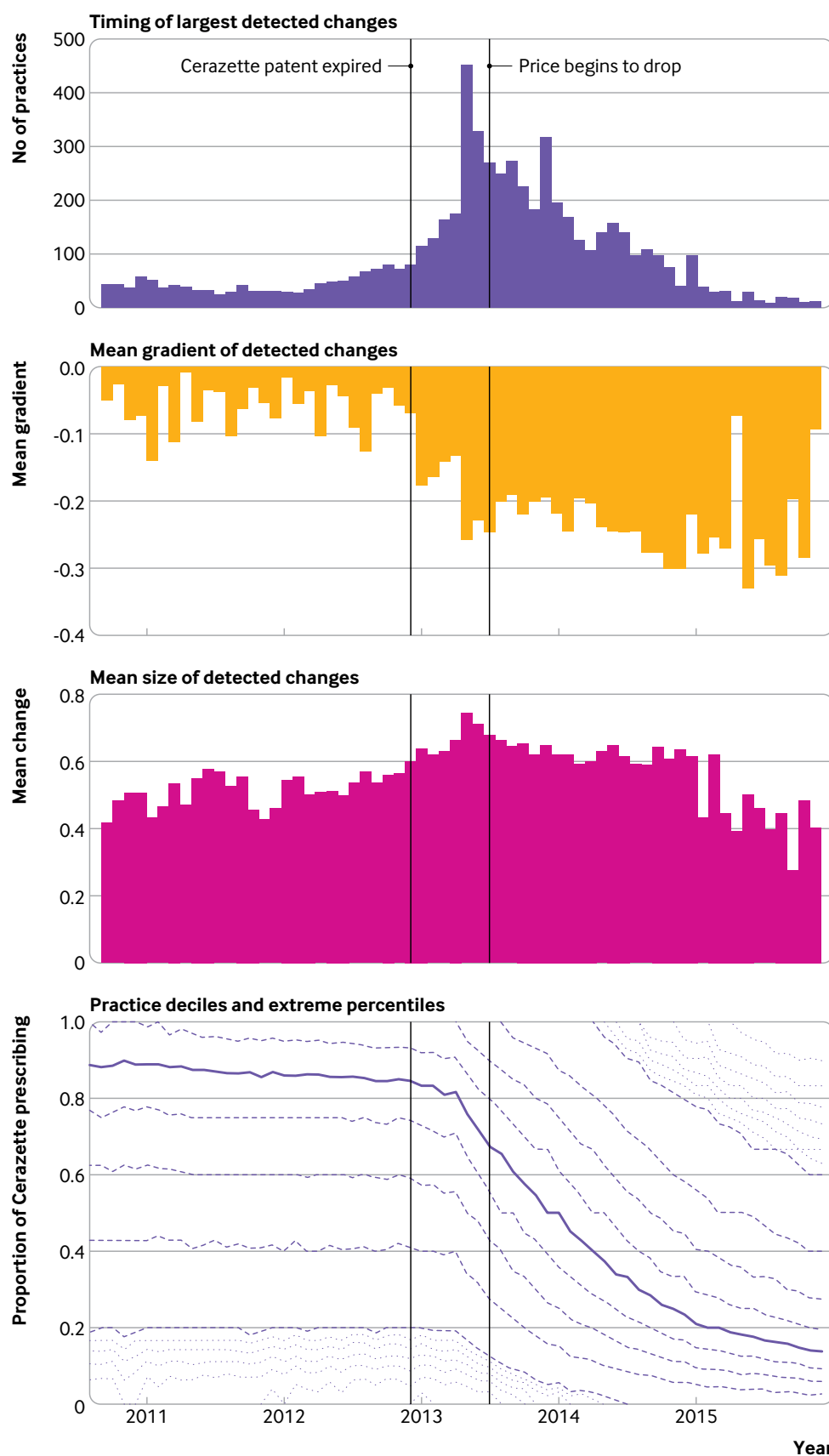


Fig 2 | Response of general practices to the patent expiry of Cerazette and subsequent price change for desogestrel. Top panel=number of practices with their largest detected downward change in each month. Second panel=mean slope of the detected change for all practices changing in that month. Third panel=mean magnitude of the detected change for all practices changing in that month. Bottom panel=median Cerazette prescribing as a proportion of all desogestrel prescribing (solid line), along with deciles (dashed lines) and extreme percentiles (dotted lines)

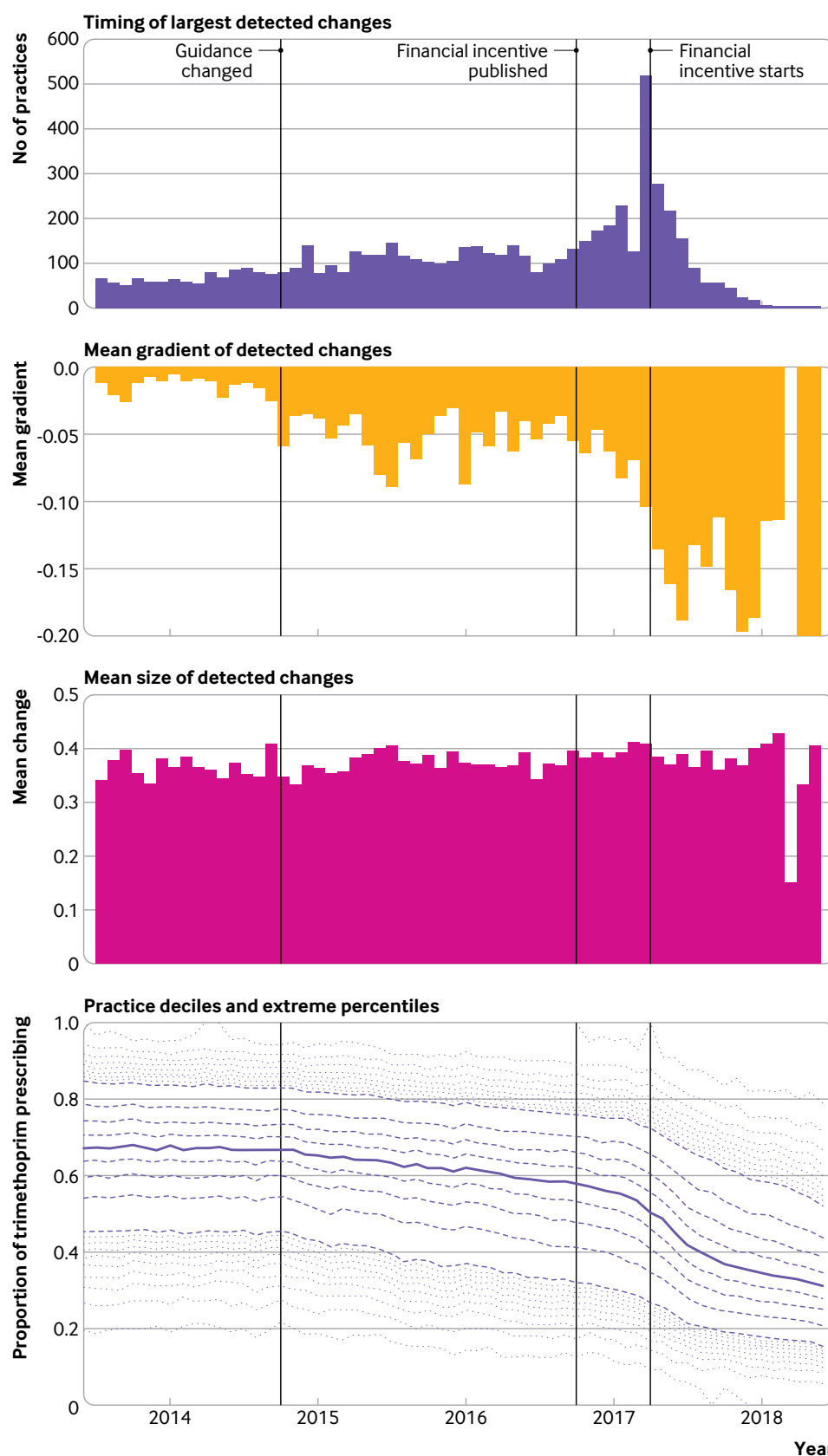


Fig 3 | Response of general practices to a change in antibiotic prescribing guidance (from trimethoprim to nitrofurantoin for uncomplicated urinary tract infection). Top panel=number of practices with their largest detected downward change in each month. Second panel=mean slope of the detected change for all practices changing in that month. Third panel=mean magnitude of detected change for all practices changing in that month. Bottom panel=median trimethoprim prescribing as a proportion of all total trimethoprim and nitrofurantoin prescribing (solid line), along with deciles (dashed lines) and extreme percentiles (dotted lines)

away from branded Cerazette in early 2013, a quarter did not show their most substantial change for 14 months (February 2014), with the slowest 10% changing at least a further 6 months later (September 2014), exposing the health system to substantial avoidable costs.

The spread of timing of changes was more pronounced for the trimethoprim/nitrofurantoin measure, with a quarter of practices not making their largest change until 29 months after the guidance was released and 10% not changing until at least 32 months after the release, exposing patients to suboptimal care. The slower dissemination of the antibiotic guidance could be because the guidance was less clear, with some clinical judgment involved, rather than “always prescribe the generic,” as was the case with desogestrel.

This variation between individual general practices in how they responded to a new warranted change in clinical practice was not limited to the timing of when the change began; variation was also seen in the slope of the change, or how rapidly that change was implemented after the change began. For example, the highest quarter of practices for slope of response reduced their proportion of branded Cerazette prescribing swiftly, by at least 26% in one month while the lowest quarter of practices for slope of response reduced branded prescribing gradually, by less than 2% per month.

We also saw some indication (fig 2, top and second panels) that practices implementing a change late tended to do so more rapidly than those who noticed the need for change earlier. This effect is perhaps due to an increased sense of urgency for practices that have noticed later. Regardless of the heterogeneity in timing and slope of change, the relative uniformity in the magnitude of change suggests that once practices implement a change, they are able to do so effectively, with most practices ultimately implementing a large change in practice.

Strengths and weaknesses

Our data cover the complete prescribing data for all practices in England, not just a sample. The underlying data are highly accurate as they are based on prescription pharmacy claims used for very high tariff transactions within the health service, with all parties motivated to ensure complete and correct information. We accounted for variation in the prevalence of underlying conditions by measuring the proportion of “all” prescribing that is “undesirable” rather than, for example, the crude volume of “undesirable” prescribing (that is, we measured Cerazette as a proportion of all Cerazette and generic desogestrel prescribing and trimethoprim as a proportion of total trimethoprim and nitrofurantoin prescribing).

The indicator saturation approach to detect breaks successfully detected change in prescribing behaviour and appears to be flexible across two different applications: the desogestrel measure had one unambiguous time point, after which prescribing generically was simply preferable; the nitrofurantoin/trimethoprim guidance, in contrast, was communicated to clinicians through various different

routes at different times, and was a change in practice that required ongoing clinical judgment, because prescribing nitrofurantoin rather than trimethoprim might not always be correct for all patients.

Findings in context

To our knowledge, this is the largest study conducted on diffusion of change in medical practice, by a substantial margin. The largest previous study monitored 95 practitioners in Denmark, covering a population of 490 000 citizens, compared with our study covering a population of 55 million.⁵ This Danish study assessed only one crude outcome metric (time to first prescription of a new antibiotic) whereas we were able to harness novel computational methods to automatically detect more detailed changes in clinical practice, across many institutions (about 8000 practices), and for more complex and generalisable clinical behaviours than a first ever prescription of a new medicine.

The previous absence of computational techniques, such as indicator saturation, explains why most previous work on diffusion of change is either small scale or focused purely on narrative descriptions (as discussed in the Introduction): without automation, it is extremely labour intensive to manually categorise whether, and when, a large number of institutions have modified their clinical practice in response to a warranted change, across a large number of patients.

Policy implications

We can identify two sets of policy implications from this work: the fact that substantial heterogeneity was detected in response to warranted changes in clinical practice; and the potential for better metrics and feedback to clinicians through the application of break detection methodology to clinical data.

Variation in speed of implementation

For both of the prescribing measures studied in this analysis, we observed substantial heterogeneity in timing and slope of warranted change but almost all practices ultimately showed substantial changes in clinical practice. In lay terms, most practices changed their behaviour but some changed much later than others; and some practices showed rapid, coordinated change, while others changed only gradually.

This heterogeneity is problematic: it exposes health systems to substantial avoidable costs and exposes patients to suboptimal clinical care. Although expecting all practices to respond immediately and adopt optimal prescription behaviour might be unrealistic, the fact that some practices changed both early and rapidly suggests that rapid timely change is possible. Further work is required to explore the reasons for some practices being slow to implement prescribing changes. We have previously written on the importance—and comparative neglect—of systems to disseminate knowledge to clinicians and patients, and social structures to audit and assess the implementation of warranted changes in practice.^{6 27}

Novel applications of indicator saturation

The automation of change detection also presents new opportunities for better use of data in audit and feedback on clinical practice, which has been shown in systematic review data to solicit modest but cost effective improvements in clinical practice.²⁸ Such audits currently rely on a static snapshot of clinical practice. Indicator saturation methods raise the potential for more sophisticated metrics—for example, describing whether an individual clinician or institution tends to respond rapidly or slowly to changes in price, evidence, or safety across a range of different elements of clinical practice. This in turn could improve the targeting of resources to support those who are responding slowly across a range of warranted changes.

Automated change detection also permits new approaches to interrogate which interventions are most impactful at soliciting change in clinical practice, both in terms of timing for initial change and rapid coordinated change. For example, in figure 3, the financial incentive is clearly associated with the largest number of practices initiating change in one month.

These new methods might also help to distinguish between warranted and unwarranted variation in care, itself an ongoing challenge for all work on variation in clinical practice: specifically, whether an observed variation is driven by variation in patients' clinical needs and preferences (warranted variation) or variation in their clinicians' knowledge, preferences, and service availability (unwarranted variation). A clinician presented with evidence that they are currently an outlier for a new desired change in clinical practice might argue that their patients are unusual and warrant clinical decisions that deviate from best practice guidelines. However, if indicator saturation methods show previous warranted changes in clinical practice that were ultimately implemented by this clinician, but three years later than their peers, then this is stronger evidence that current deviation from best practice is driven by the clinician's knowledge or choices, rather than their patients' needs or preferences.

Lastly, the potential to automate detection in timing and slope of change using indicator saturation presents an immediate opportunity to produce automated metrics on timing of change for individual clinicians and institutions. OpenPrescribing.net is an openly accessible service for detailed exploration of NHS England prescribing data by practice and by month, run by our team, with 14 000 unique users each month. We are currently developing novel measures driven by indicator saturation to describe whether practices and clinical commissioning groups overall tend to implement warranted changes in clinical practice earlier or later than their peers, for deployment and impact evaluation in our large pool of users.

Redeploying this method elsewhere

This method is highly flexible in terms of the type and quality of data that it can be applied to. Both noisy data

and data with missing values can be used, as shown by the examples described in this study. Here, we showed time series with a data point once per month, with a total of about 60 data points, but data of any length and frequency can also be used. In this study, we excluded practices where more than half of the data points were missing because it was unlikely that meaningful changes would be detected and because the false positive rate could be higher, but missing data can be handled flexibly according to the specific use case.

We chose a conservative P value of 0.000001 to ensure a low false positive rate and to increase confidence in the detected breaks actually reflecting underlying changes. In much longer time series (eg, in a sample of 1000 observations), we would expect the probability of a false positive to be $1000 \times 0.000001 = 0.001$ changes detected spuriously on average. However, simulation with small samples (<200) shows that the false positive rate of trend indicator saturation can lie above the chosen P value.²⁰ Consequently, for time series with more data points, a higher P value might be more appropriate.

We implemented break detection in this study using the R package *gets*; it can also be implemented in the econometric package *PcGive*.²⁹ Although not demonstrated in our examples, because the break detection approach is based on regression modelling, it is straightforward to include additional covariates such as seasonal cycles, autoregressive lags to capture persistence, or additional static explanatory variables.²⁰

We chose to present summary statistics for the largest detected change because it represented the most important and coordinated change. However, our break detection approach could be used in different ways for different clinical and research problems—for example, focusing on the first detected change in clinical practice or the first change to reach a prespecified threshold, depending on specific needs.

Conclusions

In our study, variation in the speed with which individual NHS general practices responded to two examples of warranted changes in clinical practice was substantial. Indicator saturation methods open up substantial new opportunities to improve clinical practice by better identifying, understanding, and reducing unwarranted variation in care.

Contributors: BG conceived the study. AJW, FP, AP-S, and BG designed the methods. AJW collected and analysed the data with input from FP, AP-S, and BG. AJW, FP, and BG drafted the manuscript. All authors contributed to and approved the final manuscript. BG supervised the project and is guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: No specific funding was sought for this analysis. Work on OpenPrescribing is supported by the Health Foundation (reference 7599); the National Institute for Health Research (NIHR) Biomedical Research Centre, Oxford; and by an NIHR School of Primary Care Research grant (reference 327). FP is supported by a grant from the Robertson Foundation. Funders had no role in the study design, collection, analysis, and interpretation of the data; in the writing of the report; and in the decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; BG has received research funding from the Laura and John Arnold Foundation, Wellcome Trust, Oxford Biomedical Research Centre, NHS National Institute for Health Research School of Primary Care Research, Health Foundation, and World Health Organization; and receives personal income from speaking and writing for lay audiences on the misuse of science. AJW is employed on BG's grants for the OpenPrescribing project. There are no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: No ethical approval was required.

Data sharing: This study uses exclusively open, publicly available data. Complete code and data are provided online on Github (https://github.com/ebsdatalab/change_detection/releases/tag/0.1), and code is also available as a python library (https://pypi.org/project/change_detection/).

The lead author affirms that the 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 (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- Ryan B, Gross N. Acceptance and diffusion of hybrid corn seed in two Iowa communities. *Res Bull (Int Comm Northwest Atl Fish)* 1950;29:1.
- Coleman JS, Katz E, Menzel H. *Medical innovation: a diffusion study*. Bobbs-Merrill Co, 1966.
- Rogers EM. *Diffusion of innovations*. 5th ed. Simon and Schuster, 2003.
- Berwick DM. Disseminating innovations in health care. *JAMA* 2003;289:1969-75. doi:10.1001/jama.289.15.1969
- Steffensen FH, Sørensen HT, Olesen F. Diffusion of new drugs in Danish general practice. *Fam Pract* 1999;16:407-13. doi:10.1093/fampra/16.4.407
- Crocker R, Walker AJ, Goldacre B. Why did some practices not implement new antibiotic prescribing guidelines on urinary tract infection? A cohort study and survey in NHS England primary care. *J Antimicrob Chemother* 2019;74:1125-32. doi:10.1093/jac/dky509
- Jones MI, Greenfield SM, Bradley CP. Prescribing new drugs: qualitative study of influences on consultants and general practitioners. *BMJ* 2001;323:378-81. doi:10.1136/bmj.323.7309.378
- Prosser H, Almond S, Walley T. Influences on GPs' decision to prescribe new drugs-the importance of who says what. *Fam Pract* 2003;20:61-8. doi:10.1093/fampra/20.1.61
- Prosser H, Walley T. New drug uptake: qualitative comparison of high and low prescribing GPs' attitudes and approach. *Fam Pract* 2003;20:583-91. doi:10.1093/fampra/cmg516
- Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *J Clin Epidemiol* 2005;58:107-12. doi:10.1016/j.jclinepi.2004.09.002
- Curtis HJ, Walker AJ, Goldacre B. Impact of NICE guidance on tamoxifen prescribing in England 2011-2017: an interrupted time series analysis. *Br J Cancer* 2018;118:1268-75. doi:10.1038/s41416-018-0065-2
- Walker AJ, Curtis HJ, Goldacre B. Impact of Chief Medical Officer activity on prescribing of antibiotics in England: an interrupted time series analysis. *J Antimicrob Chemother* 2019;74:1133-6. doi:10.1093/jac/dky528
- Wennberg JE. *Tracking medicine: a researcher's quest to understand health care*. 1st ed. Oxford University Press, 2010.
- Wise J. Health atlas shows large variations in care in England. *BMJ* 2010;341:c6809. doi:10.1136/bmj.c6809
- Curtis HJ, Crocker R, Walker AJ, Richards GC, Quinlan J, Goldacre B. Opioid prescribing trends and geographical variation in England, 1998-2018: a retrospective database study. *Lancet Psychiatry* 2019;6:140-50. doi:10.1016/S2215-0366(18)30471-1
- Curtis HJ, Walker AJ, Mahtani KR, Goldacre B. Time trends and geographical variation in prescribing of antibiotics in England 1998-2017. *J Antimicrob Chemother* 2019;74:242-50.
- Walker AJ, Curtis HJ, Bacon S, Crocker R, Goldacre B. Trends and variation in prescribing of low-priority treatments identified by NHS England: a cross-sectional study and interactive data tool in English primary care. *J R Soc Med* 2018;111:203-13. doi:10.1177/0141076818769408
- Curtis HJ, Dennis JM, Shields BM, et al. Time trends and geographical variation in prescribing of drugs for diabetes in England from 1998 to 2017. *Diabetes Obes Metab* 2018;20:2159-68. doi:10.1111/dom.13346
- Walker AJ, Curtis HJ, Bacon S, Crocker R, Goldacre B. Trends, geographical variation and factors associated with prescribing of gluten-free foods in English primary care: a cross-sectional study. *BMJ Open* 2018;8:e021312. doi:10.1136/bmjopen-2017-021312
- Pretis F, Reade J, Sucarrat G. Automated general-to-specific (GETS) regression modeling and indicator saturation for outliers and structural breaks. *J Statistical Software* 2018;86:1-44. doi:10.18637/jss.v086.i03
- Pretis F, Mann ML, Kaufmann RK. Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *Clim Change* 2015;131:705-18. doi:10.1007/s10584-015-1391-5
- Pretis F, Schneider L, Smerdon JE, et al. Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation. *J Econ Surv* 2016;30:403-29. doi:10.1111/joes.12148
- Curtis HJ, Goldacre B. OpenPrescribing: normalised data and software tool to research trends in English NHS primary care prescribing 1998-2016. *BMJ Open* 2018;8:e019921. doi:10.1136/bmjopen-2017-019921
- NHS Digital. Number of patients registered at a GP practice. 2017. <https://digital.nhs.uk/Patients-registered-at-a-GP-practice-GP-data-hub> (accessed 3 May 2017).
- Castle J, Doornik J, Hendry D, et al. Detecting location shifts during model selection by step-indicator saturation. *Econometrics* 2015;3:240-64. doi:10.3390/econometrics3020240
- Terasvirta T. Specification, estimation, and evaluation of smooth transition autoregressive models. *J Am Stat Assoc* 1994;89:208-18.
- Walker AJ, Bacon S, Curtis H, Crocker R, MacKenna B, Goldacre B. Six months on: NHS England needs to focus on dissemination, implementation and audit of its low-priority initiative. *J R Soc Med* 2019;112:4-5. doi:10.1177/0141076818808429
- Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;(6):CD000259. doi:10.1002/14651858.CD000259.pub3
- Doornik JA, Hendry DF. *Empirical Econometric Modelling using Pc-Give: Volume I*. London: Timberlake Consultants Press, 2018.

Web appendix: Supplementary figures

Web appendix: Supplementary file A

Web appendix: Supplementary file B