

BMJ Open Process evaluation of the Data-driven Quality Improvement in Primary Care (DQIP) trial: quantitative examination of variation between practices in recruitment, implementation and effectiveness

Tobias Dreischulte,^{1,2} Aileen Grant,³ Adrian Hapca,² Bruce Guthrie²

To cite: Dreischulte T, Grant A, Hapca A, *et al.* Process evaluation of the Data-driven Quality Improvement in Primary Care (DQIP) trial: quantitative examination of variation between practices in recruitment, implementation and effectiveness. *BMJ Open* 2018;**8**:e017133. doi:10.1136/bmjopen-2017-017133

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-017133>).

Received 9 August 2017

Revised 3 October 2017

Accepted 1 November 2017



CrossMark

¹Prescribing Support Unit, NHS Tayside, Dundee, UK

²Population Health Sciences Division, Medical Research Institute, University of Dundee, Dundee, UK

³School of Nursing and Midwifery, Robert Gordon University, Aberdeen, UK, Scotland

Correspondence to

Dr Tobias Dreischulte;
t.dreischulte@dundee.ac.uk

ABSTRACT

Objectives The cluster randomised trial of the Data-driven Quality Improvement in Primary Care (DQIP) intervention showed that education, informatics and financial incentives for general medical practices to review patients with ongoing high-risk prescribing of non-steroidal anti-inflammatory drugs and antiplatelets reduced the primary end point of high-risk prescribing by 37%, where both ongoing and new high-risk prescribing were significantly reduced. This quantitative process evaluation examined practice factors associated with (1) participation in the DQIP trial, (2) review activity (extent and nature of documented reviews) and (3) practice level effectiveness (relative reductions in the primary end point).

Setting/participants Invited practices recruited (n=33) and not recruited (n=32) to the DQIP trial in Scotland, UK.

Outcome measures (1) Characteristics of recruited versus non-recruited practices. Associations of (2) practice characteristics and 'adoption' (self-reported implementation work done by practices) with documented review activity and (3) of practice characteristics, DQIP adoption and review activity with effectiveness.

Results (1) Recruited practices had lower performance in the quality and outcomes framework than those declining participation. (2) Not being an approved general practitioner training practice and higher self-reported adoption were significantly associated with higher review activity. (3) Effectiveness ranged from a relative increase in high-risk prescribing of 24.1% to a relative reduction of 77.2%. High-risk prescribing and DQIP adoption (but not documented review activity) were significantly associated with greater effectiveness in the final multivariate model, explaining 64.0% of variation in effectiveness.

Conclusions Intervention implementation and effectiveness of the DQIP intervention varied substantially between practices. Although the DQIP intervention primarily targeted review of ongoing high-risk prescribing, the finding that self-reported DQIP adoption was a stronger predictor of effectiveness than documented review activity supports that reducing

Strengths and limitations of this study

- A major strength of this study is the collection of detailed quantitative data on individual practices' responses to the Data-driven Quality Improvement in Primary Care intervention, which enabled examination of substantial heterogeneity in both the implementation of intervention processes and effectiveness, thereby complementing qualitative case studies examining practitioners' perceptions of practices' responses.
- A limitation of the study is that the sample size was small (33 practices) implying relatively limited power for identifying practice characteristics significantly associated with trial participation, and for examining variation in intervention implementation and outcomes.
- A further limitation of the study is the use of a questionnaire (to measure the adoption of the intervention by general practices), which has not undergone formal psychometric testing.

initiation and/or re-initiation of high-risk prescribing is key to its effectiveness.

Trial registration number NCT01425502; Post-results.

INTRODUCTION

High-risk prescribing in primary care

High-risk prescribing in primary care is a major concern for healthcare systems internationally. Between 2% and 4% of emergency hospital admissions are caused by preventable adverse drug events.^{1 2} The National Institute of Clinical Excellence (NICE) estimated in 2015 that avoidable drug-related admissions in England cost £530million per year³ and in the USA the combined cost of drug-related hospital admissions, emergency department and outpatient visits was estimated at \$19.6billion in 2013.⁴ A large proportion of these admissions are caused by high-risk prescribing of commonly prescribed

drugs, with non-steroidal anti-inflammatory drugs (NSAIDs) and antiplatelets being the main or among the main drugs implicated, causing gastrointestinal, cardiovascular and renal adverse events.⁵⁻⁷

Data-driven Quality Improvement in Primary Care intervention and trial

In the UK, virtually all primary care prescribing is done by general practitioners (GP). The Data-driven Quality Improvement in Primary Care (DQIP) intervention was systematically developed and optimised⁸⁻¹⁰ and comprised three intervention components: (1) professional education about the risks of NSAIDs and antiplatelets via an outreach visit by a pharmacist; (2) financial incentives to review patients at the highest risk of NSAID and antiplatelet adverse drug events, split into a participation fee of £350 and £15 per patient reviewed and (3) access to a web-based IT tool to identify such patients and support structured review. The DQIP intervention was evaluated in a pragmatic cluster randomised controlled stepped wedge trial in 33 practices from one Scottish health board, where all participating practices received the intervention but were randomised to one of 10 different start dates.¹¹ Across all practices, targeted high-risk prescribing fell from 3.7% immediately before to 2.2% at the end of the intervention period (adjusted OR 0.63 (95% CI 0.57 to 0.68), $P<0.0001$). The intervention only incentivised review of ongoing high-risk prescribing, but led to reductions in both ongoing (1.5% at end vs 2.6% preintervention, adjusted OR 0.60 (95% CI 0.53 to 0.67), $P<0.001$) and 'new' high-risk prescribing (0.7% vs 1.0%, adjusted OR 0.77 (0.68 to 0.87), $P<0.001$) and notably, reductions in high-risk prescribing were sustained in the year after financial incentives stopped. In addition, there were significant reductions in emergency hospital admissions with gastrointestinal ulcer or bleeding (from 55.7 to 37.0/10 000 person-years, relative risk (RR) 0.66 (95% CI 0.51 to 0.86), $P=0.002$), and heart failure (from 708 to

513/10 000 person-years, RR 0.73 (95% CI 0.56 to 0.95), $P=0.02$).¹²

Process evaluation of the DQIP intervention

Descriptions of complex interventions in the research literature often lack details about the context in which interventions were delivered and about the processes of intervention delivery and implementation of individual intervention components.¹³⁻¹⁶ Such details are however important to decide whether and how an intervention can be implemented in routine care and to inform future research.^{14 15} We therefore carried out a comprehensive mixed-methods process evaluation alongside the main DQIP trial, based on a cluster-randomised trial process-evaluation framework¹⁷ which we developed drawing on the general literature on process evaluation, published process evaluations and the RE-AIM framework.¹⁸ Our framework¹⁷ emphasises the importance of considering two levels of intervention delivery and response that often characterise cluster-randomised trials of behaviour change interventions (although their relative importance will depend on intervention design). The first is the intervention that is delivered to clusters, which respond by adopting (or not) the intervention and integrating it with existing work. The second is the desired change in care which the cluster professionals deliver to individual patients. Figure 1 shows the application of the framework to the DQIP intervention. In DQIP, the delivery of the intervention to professionals was predefined, intended to be delivered with high fidelity across all practices and under the control of the research team, whereas the intervention delivered to patients was largely at the discretion of practices, who decided whether and how they reviewed patients and whether to stop or otherwise proactively manage high-risk prescribing in those reviewed. This is similar to most health service interventions of this nature, where primary care organisations deliver an intervention to practices (eg, contractual changes, financial incentives,

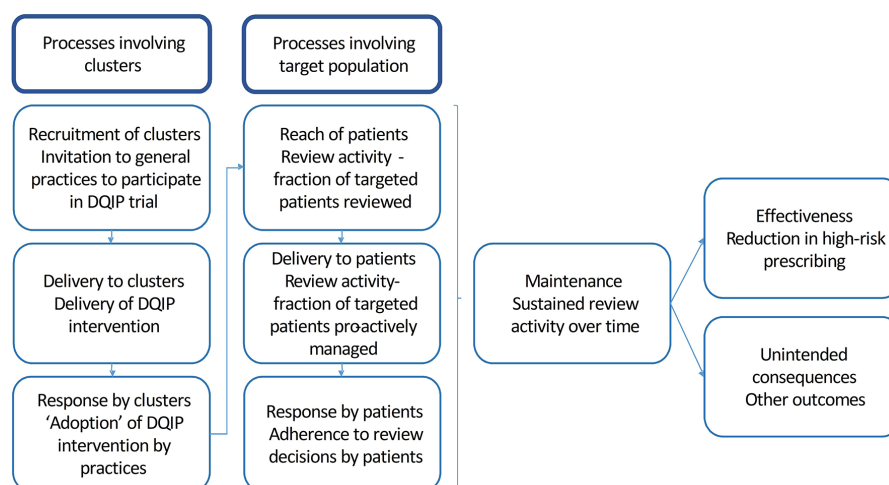


Figure 1 Framework model for designing process evaluations of cluster-randomised controlled trials applied to the Data-driven Quality Improvement in Primary Care (DQIP) trial.

education and so on), but have relatively little control over whether or how practices implement the intended care for patients. We used this framework to structure our parallel process evaluation, mapping data collection to a logic model of how the DQIP intervention was expected to work.¹⁷

Focus of this paper

Understanding cluster factors associated with participation, implementation and effectiveness is important to understand how generalisable the findings are and how the intervention achieves (or fails to achieve) intended outcomes. Nevertheless, published evaluations of complex interventions rarely compare trial and target populations and often only report overall intervention impact. Additionally, process evaluations often largely assess participants' perceptions of how interventions were implemented using qualitative methods, although many processes can usefully be examined quantitatively, providing potentially complementary findings about implementation. In this paper, we report the findings of prespecified quantitative analyses to examine (1) whether and how practices recruited to the DQIP trial differed from non-recruited ones; (2) whether practice characteristics and practice-level response to the intervention delivered to them by the research team (subsequently referred to as adoption) were associated with the way the practice implemented delivering the intended change in care to patients (subsequently referred to as review activity); (3) whether practice characteristics, adoption and review activity were associated with intervention effectiveness at practice level.¹⁷

METHODS

The design and methods of the overall process evaluation have been described previously in the published protocol.¹⁷ In brief, the design was a mixed methods parallel process evaluation which examined a set of predefined processes and their associations with change in high-risk prescribing at practice level. The qualitative element consisted of comparative case studies in 10 of the 33 participating practices, purposively sampled to include a mix of those initially responding and not responding to the intervention by rapidly reducing their high-risk prescribing, as judged by visual inspection of run charts approximately 4 months after practices started the intervention. The case-study analysis of how practices adopted, implemented and maintained the intervention,¹⁹ and a detailed description of the intervention (including examples of materials used) with analysis of which intervention components were more or less active and when this occurred are described separately.^{19 20} This paper reports the quantitative element of the process evaluation, which examined associations between practice characteristics, key intervention implementation processes and change in the high-risk NSAID and antiplatelet prescribing at

practice level. The study was reviewed by the Fife and Forth Valley Research Ethics Committee (11/AL/0251).

DQIP trial study population

The study population comprised all NHS Tayside practices which participated in the DQIP trial. At practice level, the intervention period lasted 48 weeks from each practice's randomised start date. At patient level, all patients registered with participating practices were included if they had at least one gastrointestinal, heart failure or renal risk factor that made them particularly vulnerable to the adverse effects of NSAIDs or antiplatelets (as defined by primary and secondary prescribing outcome measures).¹² Patients' membership in the study cohort was dynamic because the presence of risk factors in individuals varies over time due to changing age, morbidity and coprescribing, and ended with the end of the 48-week DQIP intervention period in each practice, deregistration with a practice participating in DQIP, or death (whichever occurred first). Moving of patients between practices participating in the DQIP trial was accounted for by patients exiting and entering the respective practices' cohorts at the point of registration with a new practice, although this was rare (<0.5% of patients).

Data collection and variable definition

The characteristics of participating and non-participating practices (list size, rurality, socioeconomic deprivation, proportion of older people, postgraduate training status, overall quality and outcomes framework (QOF) clinical performance) were measured using publicly available data.²¹

Adoption. We conceptualised adoption as the work done by practices to prepare for and begin to deliver reviews to patients, and to ensure effective and sustained review. In order to further elicit whether and how practices responded to the intervention delivered to them by the research team in relation to adoption, the GP most involved in the DQIP work (nominated by the practice) completed two surveys, which were based on the four domains of Normalisation Process Theory,²² namely coherence (whether practices differentiate the intervention from existing prescribing improvement work and understand and value intervention processes and aims), cognitive participation (whether practices engage with and plan to take collective action to implement the intervention), collective action (whether practices integrate the intervention into existing work and avoid disruptions to routine care) and reflexive monitoring (whether and how practices re-evaluate the intervention after attempting initial implementation and whether they modify practice processes if necessary). Given that there are no validated measures of these concepts, and constrained by the resources available, we pragmatically developed two instruments for use in an exploratory analysis (see online supplementary appendix 1 and 2).

In brief, a range of items were created with final item selection based on discussion with a panel of eight GPs

and/or normalisation process theory (NPT) experts to confirm content validity,²³ and internal consistency was demonstrated for items targeting the same NPT constructs. The first survey was administered shortly after the educational outreach visit at the start of the intervention in each practice and asked respondents about their degree of agreement with statements relating to coherence (six items) and cognitive participation (five items). The second survey was administered 6–9 months after practices started the intervention and asked respondents about their degree of agreement with statements on collective action (three items), and reflexive monitoring (four items) during intervention implementation. Agreement with the statements was measured on a 7-point Likert scale (1=strongly agree, 4=neither agree nor disagree, 7=strongly disagree), with all statements being worded such that lower levels of agreement (higher numerical scores) indicated lower levels of coherence, cognitive participation, collective action and reflexive monitoring, respectively. Both surveys are provided in the appendix material. For each of the four NPT constructs, we transformed the agreement scale to an ‘achievement’ scale, so that the highest level of agreement with statements favouring each construct (agreement score=1) had the highest achievement score (achievement score=6), and the highest level of disagreement (agreement score=7) had the lowest achievement score (achievement score=0). Practices’ achievement was defined as the proportion of achieved over total maximum achievable scores across all questionnaire items that used these scales for each NPT construct.

Review activity was measured in terms of practice level reach, delivery to the patient and maintenance as specified in the published protocol¹⁷ (see figure 1). The DQIP IT tool provided data on reach (the proportion of patients flagged as needing a review over the 48-week intervention period, for whom the practice used the DQIP tool to record completion of a review), delivery to the patient (the proportion of eligible patients where there was an active change in treatment or proactive follow-up to determine appropriateness after initial records review) and maintenance (defined as reach in the final 24 weeks of the 48-week intervention period).

Effectiveness was measured using the trial primary outcome, which was a composite of nine individual high-risk prescribing measures targeting high-risk NSAID and antiplatelet prescribing. In order to reduce chance effects (eg, in the context of measuring every 8 weeks, those arising from variable time intervals between prescriptions of a high-risk drug like NSAIDs and those relating to the small number of patients measured in some practices at single time-points), we defined effectiveness in each practice as the relative change in mean high-risk prescribing (measured using the trial primary outcome between the baseline 48 weeks before each practice started the intervention and the final 24 weeks of the intervention period).

Statistical analysis

All analyses were conducted at cluster level. In order to inform judgements about generalisability (objective 1), we compared recruited practices to those who were invited but declined participation. In order to examine variation between practices in the intended review activity (objective 2), we used univariate and multivariate linear regression to examine associations between practice characteristics (list size, whether practices were accredited for postgraduate training, proportion of patients aged 75 years or older, proportion of patients living in the most socioeconomically deprived areas and baseline high-risk prescribing) and the four measures of adoption (coherence, cognitive participation, collective action, reflexive monitoring) with each of the three measures of review activity (reach, maintenance and delivery to the patient). In order to examine variation between practices in effectiveness (objective 3), we used univariate and multivariate linear regression to study associations between practice characteristics, adoption and review activity with effectiveness. Given our hypothesis that DQIP would drive review activity, which in turn would drive effectiveness, strong associations between adoption and review activity variables were expected. In order to explore such mediation effects, we therefore considered two multivariate models. The first only considered review activity variables and the second considered both adoption and review activity variables (in addition to practice characteristics). This allowed us to examine changes in the association between review activity and effectiveness after controlling for DQIP adoption. In all models, continuous explanatory variables were dichotomised as ‘high’ or ‘low’ using the median as a threshold. Multivariate modelling was conducted using stepwise elimination of variables that were not significant in each model at the $P=0.10$ level. Collinearity between variables was tested by examining the correlation coefficients between each pair of explanatory variables and variance inflation factors (VIF). Where meaningful collinearity was identified between two explanatory variables, only the variable that yielded the better overall model fit was retained in the final model. To account for the small sample size, model fit was assessed using the adjusted R^2 .²⁴ In sensitivity analyses, we repeated the above analyses without dichotomising continuous variables.

RESULTS

Comparison of recruited vs non-recruited practices

Among recruited practices the median (range) list size was 6167 (1367–11643), the median proportion of registered patients aged 75 years and older was 8.8% (5.9–13.9), and the median proportion of patients living in the most deprived quintile of Scottish postcodes was 2.15% (0.00, 43.8). The median level of high-risk prescribing during the baseline period was 3.44% and varied more than sevenfold between practices (from 0.7% to 5.2%). Table 1 shows that participating practices had on average statistically significantly lower QOF performance than

Table 1 Characteristics of general practices recruited to the DQIP trial compared with eligible practices which declined to participate

Characteristic	Recruited (n=34)*	Eligible but declined (n=32)†	P value
Median (range) list size	6173 (1367 to 11803)	6551 (115 to 12 869)	0.348‡
Median (range) % aged ≥75 years	8.7 (5.7 to 14.1)	8.4 (0.0 to 18.1)	0.875‡
Median (range) % of registered patients resident in the 15% most deprived Scottish postcode areas	2.2 (0.0 to 43.8)	10.9 (0.0 to 61.8)	0.292‡
Median (range) % of registered patients resident in settlements of <3000 inhabitants	30.0 (5.8 to 64.1)	15.3 (4.8 to 41.1)	0.287‡
Contract type (nGMS vs other)	31 (91.2)	32 (100.0)	0.131§
Median (range) % of total QOF points achieved¶	99.4 (83.5 to 100.0)	99.7 (94.5 to 100.0)	0.028*‡
GP postgraduate training accreditation	9 (26.5)	14 (43.8)	0.456§

*The total number of initially recruited practices. One practice dropped out before their allocated DQIP start date.

†Two additional practices which were involved in intervention optimisation and development were not eligible to be invited.

‡Mann-Whitney U test.

§ χ^2 test.

¶The QOF financially rewards primary care practices according to their performance on a range of clinical and organisational indicators, each of which is associated with a number of maximum achievable points, with each point corresponding to a defined payment. QOF scores range from 0 to 1000, with higher scores indicating better performance. The metric reported here represents the average proportion of maximum achievable points actually achieved by respective practices.

GP, general practitioner; nGMS, new general medical services contract; QOF, quality and outcomes framework.

practices who declined participation. Other differences were not statistically significant, although there were also reasonably substantial absolute differences in the proportion of registered patients living in settlements of <3000 inhabitants (30.0% in participating vs 15.3% in non-participating), the proportion of registered patients living in the most deprived Scottish postcode areas (2.2% vs 10.9%) and the proportion of practices accredited for GP training (26.5% vs 43.8%).

Variation in adoption, review activity and effectiveness

Adoption—response rates to questionnaires. All 33 practices in the trial completed the first adoption questionnaire (assessing coherence and cognitive participation before trial implementation) when they started the intervention, and 29 practices (response rate 88%) completed the second questionnaire (assessing collective action and reflexive monitoring during initial implementation) 6–9 months after they started the intervention. All items were completed in returned questionnaires.

Adoption—coherence and cognitive participation. At the start of the intervention in each practice, the median coherence and cognitive participation scores were 77.8% (range 22.2–100.0) and 80.0% (range 26.7–100.0) reflecting that at the point of DQIP initiation, the majority of responding GPs indicated that they and their colleagues understood the work required (coherence questions A1.3 and A1.4), agreed that DQIP was worthwhile (coherence questions A1.5 and A1.6), expected that DQIP would lead to reductions in targeted high-risk prescribing (cognitive participation question A1.7) and that all prescribers in the practice should and would get involved (cognitive participation question A1.8 and A1.9). However, there were more mixed responses

to questions about whether DQIP was different from existing NHS quality improvement work (coherence questions A1.1 and A1.2) (figure 2).

Adoption—collective action and reflexive monitoring. In comparison, at 6–9 months after the start of the intervention in each practice, median levels of agreement with statements relating to collective action (66.7%, range 11.1–100.0) and reflexive monitoring (66.7%, range 16.7–100.0) during initial intervention implementation were somewhat lower. Less than half of practices (n=13) reported frequently using the DQIP IT tool to monitor progress in reducing high-risk prescribing (reflexive monitoring question A2.5), and less than a third (n=9) stated that DQIP had led to changes in practice processes, structures or systems beyond delivering the actual DQIP work (reflexive monitoring question A2.8). Nevertheless, over half of practices agreed that those involved in DQIP had the necessary skill set (collective action, question A2.1, 19 practices), that it was easy to integrate DQIP with existing work (collective action question A2.2, 19 practices) and that patients would respond well to medication changes (collective action question A2.3, 18 practices) (figure 2).

Review activity. The median proportion of patients with high-risk prescribing who had a review documented in the DQIP tool (reach) at any point during the DQIP intervention period was 78.8% (range 0.0%–100.0%), with a similar median reach of 75.0% (0%–100.0%) in the second half of the DQIP intervention period (maintenance). The median proportion of patients flagged as needing a review, who were judged to require proactive follow-up at review (delivery to patient) was 39.7%, ranging from 0.0% to 83.3%.

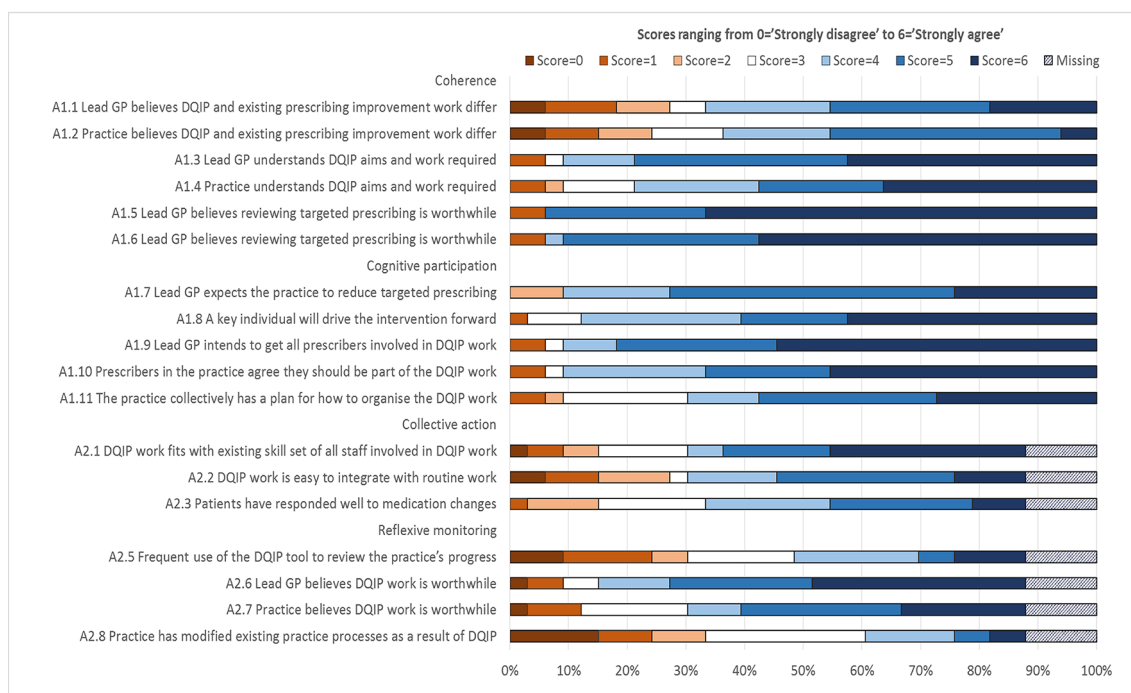


Figure 2 Findings from two adoption questionnaires completed by the general practitioner (GP) leading on Data-driven Quality Improvement in Primary Care (DQIP) in each practice.

Effectiveness. The median effectiveness was a relative reduction of 46.6%, ranging from a relative increase in high-risk prescribing of 24.1% to a relative reduction of 77.2%.

Associations of practice characteristics and adoption with review activity

Table 2 shows the coefficients for the associations of the respective explanatory variables with reach, maintenance and delivery to patient. Since all explanatory variables were dichotomised, the coefficients reflect the mean absolute differences in the outcome variable between the two categories of explanatory variables (eg, reach was 20.7 percentage points lower in large compared with small). In multivariate analyses, being a registered GP training practice was the only practice characteristic significantly associated with review activity (delivery to patient), in that being a training practice was a predictor of a lower proportion of patients having their high-risk prescribing changed or actively followed up (despite the fact that baseline levels of high-risk prescribing were comparable between training and non-training practices (3.8% vs 3.5%) and not significantly associated with review activity). With respect to adoption variables, coherence and reflexive monitoring were significantly associated with a higher proportion of patients with high-risk prescribing reviewed over the 48-week intervention period (reach), and reflexive monitoring was significantly associated with a higher proportion of patients having their high-risk prescribing changed or actively followed up (delivery to patient). Collective action was no longer significantly associated with reach in multivariate analysis after adjustment for coherence and reflexive monitoring

(in the absence of evidence of significant collinearity (all VIFs<2)).

Associations of practice characteristics, adoption and review activity with effectiveness

Figure 3A ranks participating practices by effectiveness (the achieved relative reduction in NSAID or antiplatelet high-risk prescribing) and figure 3B shows the review activity in these practices. Practices with larger relative reductions in high-risk prescribing appeared to generally have reviewed higher proportions of patients (reach) and identified higher proportions of patients for proactive follow-up (delivery to patient) than practices with smaller reductions, although there were exceptions. For example, the practices ranking 22nd and 23rd for effectiveness, had relative reductions in high-risk prescribing of 19.3% and 19.0%, respectively, despite minimal review activity recorded in the informatics tool (and therefore no payment of financial incentives).

In multivariate model 1 (examining the relationship between review activity and effectiveness while controlling for practice characteristics but not adoption variables), only reach remained significantly associated with effectiveness (although maintenance and delivery to patient were also significantly associated with effectiveness in univariate analysis; there was no evidence of significant collinearity (VIF<2)). In multivariate model 2 (when adoption variables were considered additionally), higher coherence and collective action were significantly associated with effectiveness, but reach and other review activity variables were not. Model 2 explained a substantially larger proportion (adjusted $R^2=0.64$) of variation in effectiveness compared with model 1 (adjusted $R^2=0.33$)

Table 2 Associations of practice characteristics and adoption with review activity

Explanatory variables	% (95% CI) difference in reach		% (95% CI) difference in maintenance		% (95% CI) difference in delivery to patient	
	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate
List size (large vs small)	−20.7 (−40.3 to −1.2) P=0.04	NS	−13.7 (−35.3 to 7.8) P=0.20	NS	−16.1 (−30.5 to −1.7) P=0.03	NS
% of patients aged ≥75 years (high vs low)	3.4 (−17.5 to 24.4) P=0.74	NS	3.3 (−18.8 to 25.4) P=0.76	NS	1.3 (−14.3 to 16.8) P=0.87	NS
Approved GP training practice (yes vs no)	−24.9 (−46.6 to −3.2) P=0.02	NS	−23.2 (−46.6 to 0.14) P=0.05	NS	−21.6 (−37.1 to −6.0) P=0.008	−19.8 (−33.1 to −6.6) P=0.005
% of patients in most deprived quintile (high vs low)	−3.0 (−23.9 to 18.0) P=0.77	NS	−2.8 (−24.9 to 19.4) P=0.80	NS	−6.1 (−21.5 to 9.30) P=0.42	−10.1 (−21.8 to 1.6) P=0.09
% baseline high-risk prescribing (high vs low)	4.9 (−16.0 to 25.8) P=0.63	NS	5.8 (−16.3 to 27.8) P=0.60	NS	0.8 (−14.8 to 16.4) P=0.92	NS
% coherence (whether practices understand and value intervention aims and processes) (high vs low)	12.0 (−9.8 to 33.8) P=0.27	16.9 (1.06 to 32.7) P=0.04	8.1 (−15.2 to 31.4) P=0.48	NS	13.0 (−2.82 to 28.8) P=0.10	NS
% cognitive participation (whether practices engage with and plan to implement) (high vs low)	7.2 (−14.4 to 28.8) P=0.50	NS	3.3 (−19.7 to 26.3) P=0.77	NS	8.1 (−7.75 to 24.0) P=0.30	NS
% collective action (whether practices effectively implement in the context of existing work) (high vs low)	25.7 (7.0 to 44.4) P=0.009	NS	30.3 (11.1 to 49.5) P=0.003	NS	13.8 (−0.88 to 28.6) P=0.06	NS
% reflexive monitoring (whether practices evaluated and if necessary modify the implementation) (high vs low)	30.1 (12.2 to 48.0) P=0.002	26.4 (9.9 to 42.9) P=0.003	28.4 (10.6 to 46.2) P=0.003	28.4 (10.6 to 46.2) P=0.003	23.9 (10.9 to 36.8) P=0.001	18.8 (7.2 to 30.3) P=0.003
R ² for multivariate model (adjusted for number of variables*)		0.39 (0.34)		0.28 (0.25)		0.52 (0.46)

*The adjusted R² was calculated as $1 - ((1 - R^2) \times (N - 1) / (N - P - 1))$, where R² is sample R-square, P is number of and n is total sample size. GP, general practitioner; NS, not significantly associated in multivariate model so dropped.

(table 3). The results of the sensitivity analyses were consistent with the primary analyses findings (see online supplementary appendix 3).

DISCUSSION

Summary of findings

In this analysis of quantitative data from the process evaluation of the DQIP trial, lower QOF performance was found to be the only practice characteristic that was statistically significantly associated with willingness to participate in DQIP, consistent with recruited practices not being higher performing in general. Across all practices, there was large variation in both recorded review activity

(reach and maintenance each ranging from 0.0% to 100%, delivery to patient ranging from 0.0% to 83.3%), and effectiveness (ranging from a relative increase of 24.1% to a relative reduction of 77.2%, although 30/33 practices had a reduction of some kind in the targeted prescribing). With respect to factors explaining variation in review activity, higher adoption was significantly associated with higher review activity (coherence with reach, and reflexive monitoring with all three of reach, maintenance and delivery to patient). Being a training practice was significantly associated with lower review activity (delivery to patient) despite comparable levels of baseline high-risk prescribing in training (3.8%) and non-training practices (3.5%). This was somewhat unexpected since

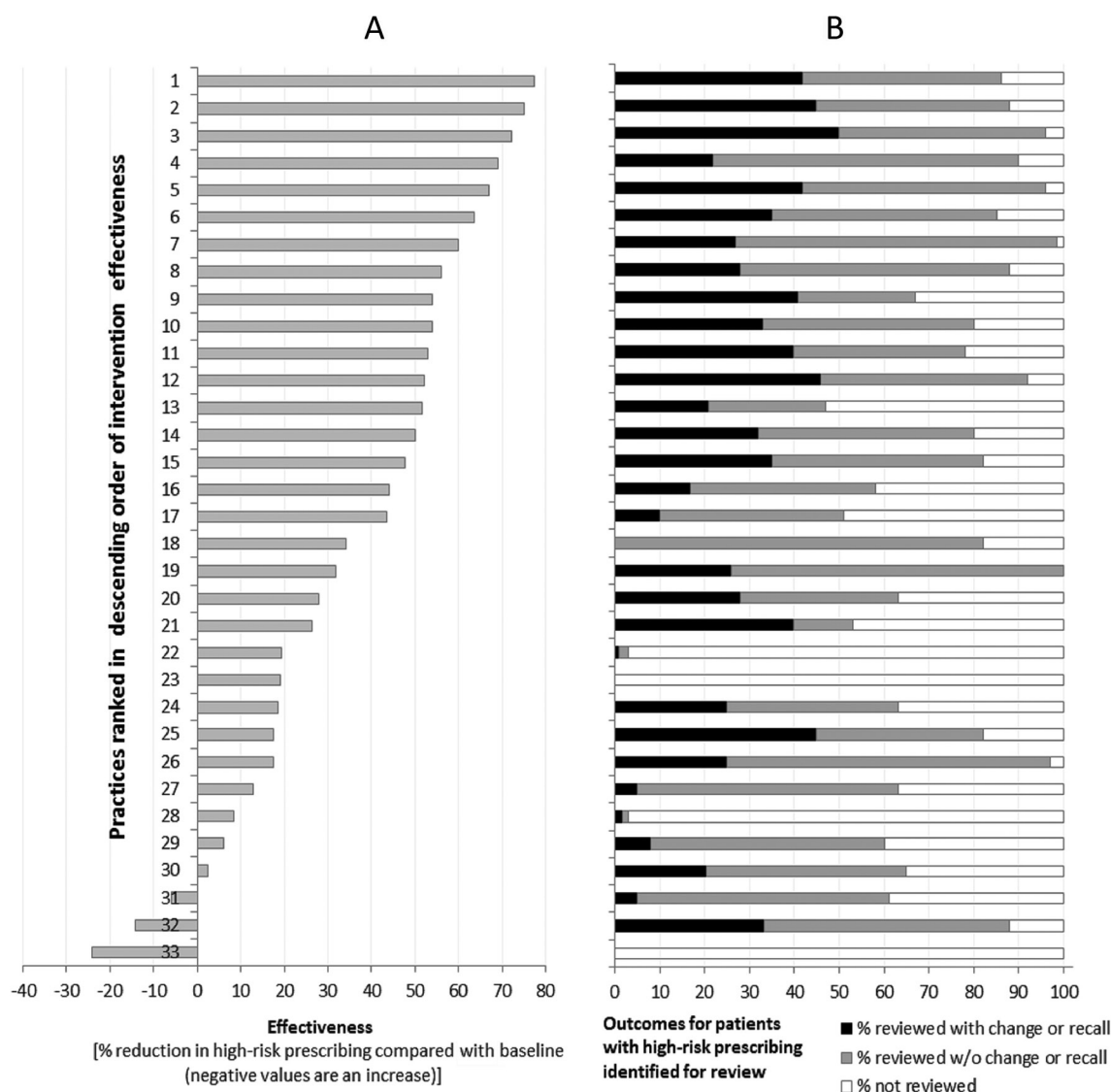


Figure 3 Variation among practices in effectiveness (A) and review activity (B).

the general expectation is that training practices will have higher quality, and further research to examine this would be of interest.

With respect to factors explaining variation in effectiveness, higher baseline high-risk prescribing was associated with larger reductions in high-risk prescribing, which is likely to reflect larger scope for improvement. Higher review activity was also significantly associated with effectiveness after adjustment for practice characteristics, with practice characteristics and review activity together explaining 38.5% of variation in effectiveness. However, when adoption variables were additionally introduced into the model, review activity was no longer independently associated with effectiveness. The final model comprising baseline high-risk prescribing, coherence and collective action as explanatory variables explained a substantive amount (64%) of variation in effectiveness between practices.

Strengths and limitations

A major strength of this study is that we were able to collect detailed quantitative data on how practices responded to

the DQIP intervention (in terms of adoption and review activity) from all participating practices while the intervention was live. This enabled us to examine substantial heterogeneity in both the implementation of intervention processes and effectiveness, thereby complementing qualitative case studies examining practitioners' perceptions of practices' responses. However, our study also has some limitations. First, as with most cluster-randomised trials, the sample size was small (33 practices) reflecting that it is driven by the power calculation for the main trial rather than being driven by the process evaluation. This has a number of implications including risk of biased overestimation of explained variation (which we accounted for by reporting an adjusted R^2 value) and risk of overfitting of models (as a result of which we dropped several marginally significant ($0.05 < P < 0.1$) variables from final models, and our fitted model has at least 10 practices per variable which is adequate in view of simulation studies showing that a minimum of only two SPV may suffice in linear regression for adequate estimation of regression coefficients).²⁴ More problematically, there is relatively limited

Table 3 Associations of practice characteristics, adoption and review activity with effectiveness

Explanatory variables	% (95% CI) difference in effectiveness (univariate)	% (95% CI) difference in effectiveness (multivariate model 1—practice characteristics and review activity only)	% (95% CI) difference in effectiveness (multivariate model 2—all variables)
Practice characteristics			
List size (large vs small)	−12.9 (−31.8 to 5.9), P=0.17	NS	NS*
% of patients aged ≥75 years (high vs low)	3.5 (−15.9 to 22.8), P=0.72	NS	NS†
Training status (yes vs no)	−18.8 (−39.5 to 1.8), P=0.07	NS	NS
% of patients in most deprived postcodes (high vs low)	−7.6 (−26.8 to −11.6), P=0.42	NS	NS
% baseline high-risk prescribing (high vs low)	23.2 (5.8 to 40.6), P=0.01	19.6 (4.0 to 35.3), P=0.02	13.9 (−0.12 to 27.9), P=0.05
Review activity			
% reach (high vs low)	26.5 (9.7 to 43.3), P=0.003	23.6 (7.9 to 39.2), P=0.004	NS
% maintenance (high vs low)	18.9 (0.8 to 37.1), P=0.04	NS	NS
% delivery to patient (high vs low)	24.9 (7.7 to 42.0), P=0.006	NS	NS
Adoption			
% 'coherence' (high vs low)	25.3 (6.9 to 43.7), P=0.009	Not examined	22.2 (7.8 to 36.5), P=0.004
% 'cognitive participation' (high vs low)	3.8 (−16.3 to 23.9), P=0.70	Not examined	NS
% 'collective action' (high vs low)	32.1 (16.7 to 47.6), P<0.001	Not examined	30.1 (17.4 to 42.9), P<0.001
% 'reflexive monitoring' (high vs low)	21.9 (4.1 to 39.6), P=0.02	Not examined	NS
R ² for multivariate model (adjusted for number of variables‡)		0.38 (0.33)	0.68 (0.64)

*Retained in the initial model as P=0.07 (criteria for retaining P<0.1) but dropped to avoid overfitting given the small number of practices.

†Retained in the initial model as P=0.08 (criteria for retaining P<0.1) but dropped to avoid overfitting given the small number of practices.

‡The adjusted R² was calculated as $1 - [(1 - R^2)(N - 1)/(N - P)]$, where R² is sample R-square, P is number of predictors and n is total sample size.

NS, not significantly associated in multivariate model so dropped.

power for identifying practice characteristics significantly associated with trial participation, and for examining variation in intervention implementation and outcomes. Our decision to dichotomise continuous explanatory variables further reduced power, but as described above we believe it is necessary to enhance interpretability for an improvement audience and our findings demonstrate that variables associated with a 20% or larger absolute difference in review activity and effectiveness could be identified as statistically significant. In addition, sensitivity analyses that fitted predictor variables as continuous yielded findings consistent with the primary analyses (see online supplementary appendix 3).

Second, reflecting limited resources and the lack of any existing NPT instruments at the time of the study, we limited examination of practice characteristics to those that could be measured using publicly available data

sources and we took a pragmatic approach to the development and administration of the two NPT-informed adoption surveys. Although face and content validity was confirmed by a panel of NPT experts, time constraints meant that we were unable to conduct psychometric testing. We also pragmatically limited scoring of the questionnaires to the DQIP lead in each practice rather than including all professionals targeted by the intervention. Given evidence of substantial variation in the tendency to issue high-risk prescriptions of NSAIDs between GPs in the same practice,²⁵ it is plausible that perceptions of targeted high-risk prescribing and (by extension) the value of the intervention may differ between different team members. Although the questionnaires also sought perceptions of the participating GP regarding the practice as a whole, future research should aim to elicit how valid such perceptions are. The adoption analysis should

therefore be considered exploratory, but it is interesting that scores on the adoption surveys explained a large proportion of variation in effectiveness, indicating the potential for the use of NPT-based surveys for measuring the adoption of complex interventions by clinical teams. In addition, the survey findings are consistent with reported implementation processes elicited by interviews and observations in 10 case-study practices¹⁹ (see online supplementary appendix 4). While the instruments used here were study specific, a generic instrument to measure NPT implementation processes has since been developed.²⁶

Third, our measure of effectiveness (change from the baseline period to the second half of the intervention period) is vulnerable to secular trends since the analysis did not account for time trends (although there was no evidence of an overall secular trend in the primary analysis¹²). Finally, our measures of review activity are all derived from the DQIP informatics tool and some practices may have carried out reviews or other improvement activity without recording it in the tool. Practices only received payment if they completed a review in the tool, which would be expected to make recording more complete, but several practices with large reductions in prescribing recorded few if any reviews. In the main trial analysis, the intervention was also found to significantly reduce new high-risk prescribing.¹² This is consistent with the intervention having effects beyond those directly attributable to the review of ongoing prescribing, which was the main target of the intervention and process evaluation measurement. Our measurement of 'review activity' is therefore likely to be an incomplete assessment of activity triggered by the intervention. This is consistent with the findings of the case study which identified other work being done in practices beyond that required by the trial team, including work by GPs to minimise the restarting of medicines where there had been an active decision to stop.¹⁹

The findings complement the two accompanying process evaluation papers.^{19 20} Participants perceived that all intervention components delivered to the practice by the research team were active,^{19 20} which is consistent with the findings of overall high adoption and review activity, and the observation that 30/33 practices had at least some reduction in targeted high-risk prescribing. The quantitative findings are also consistent with the findings of the mixed methods case-study evaluation of intervention adoption and implementation of review activity by practices, in that most of the case-study practices sampled for *not* having initially reduced their high-risk prescribing were found to actually have delayed implementation and had reduced their high-risk prescribing by the end of the trial.¹⁹ The case-study evaluation identifies that there are additional improvement processes happening which are not captured by our measures of review activity, highlighting the value of in-depth qualitative examination of implementation in a sample of practices alongside broader quantitative examination in all practices.¹⁹

Implications

In order to reduce patient exposure to high-risk prescribing, the DQIP intervention was designed to use education, informatics and financial incentives to activate GPs and practices to review patients with ongoing high-risk prescribing, where documentation of reviews in the DQIP tool was required to obtain payment. We therefore hypothesised that adoption of the intervention (measured in terms of the four NPT constructs coherence, cognitive participation, collective action and reflexive monitoring) would stimulate review activity (measured in terms of reach, maintenance and delivery to patient) documented in the DQIP tool, which in turn would translate into reductions in high-risk prescribing. We found significant associations between adoption and review activity (in both univariate and multivariate analysis) and between review activity and effectiveness (in univariate analysis and in multivariate analysis after adjustment for practice characteristics only), which generally supports the hypothesised mechanism of action. However, the association between review activity and effectiveness was no longer significant after adoption variables had been introduced into the multivariate model and the inclusion of adoption versus review activity variables in the model explained a substantially larger proportion of variation in effectiveness. This was somewhat surprising, because it suggests that DQIP adoption (as defined and measured in this study) impacted on the outcome in ways beyond those primarily targeted by the DQIP intervention (ie, review and proactive management of *ongoing* high-risk prescribing). Given the relatively small sample size, we did not attempt to conduct a formal mediation analysis²⁷ and these findings should therefore be interpreted with caution. Nevertheless, our results are consistent with the main trial finding of significant reductions in the initiation of high-risk prescribing which was not the primary target of the intervention (in addition to reductions in ongoing prescribing which was the primary target).¹² One limitation of the DQIP intervention was its narrow focus on two classes of medicines (NSAIDs and antiplatelets). The combined findings of the main trial and this process evaluation suggest that it may be possible to extend the scope of targeted high-risk prescribing without risking a rebound in the high-risk prescribing targeted here. Whether this is the case and whether facilitating and incentivising review of other potentially suboptimal care more generally lead to changes in sustained changes in professional behaviour is an interesting area for future research.

CONCLUSION

Despite the large overall reductions in high-risk prescribing observed in the DQIP trial, the findings from this quantitative process evaluation demonstrate substantial variation between practices in terms of their response to the intervention and effectiveness, the latter being partially explained by differences in baseline high-risk

prescribing. Our findings suggest that the overall impact of the DQIP intervention was significantly mediated by its impact on new high-risk prescribing or its re-initiation. The adoption processes coherence (whether practices understand and value intervention processes and aims) and collective action (whether practices effectively implement the intervention in the context of existing work) could be targets for further refinement of the DQIP intervention in order to achieve a more consistent intervention effect and further enhance effectiveness.

Acknowledgements The authors would like to thank all participating practices and the University of Dundee's Health Informatics Centre for data management and anonymisation, the Advisory and Trial Steering Groups and Debby O'Farrell who provided administrative support.

Contributors BG, AG and TD designed the study. AG led the development and initial testing of the adoption questionnaires. TD conducted all statistical analyses supported by BG and AH. TD wrote the first draft of the manuscript with all authors commenting on subsequent drafts.

Funding The study was supported by a grant (ARPG/07/02) from the Scottish Government Chief Scientist Office. The funder had no role in study design, data collection, analysis and interpretation or the decision to publish.

Competing interests None declared.

Patient consent Detail has been removed from this case description/these case descriptions to ensure anonymity. The editors and reviewers have seen the detailed information available and are satisfied that the information backs up the case the authors are making.

Ethics approval The study was reviewed by the Fife and Forth Valley Research Ethics Committee (11/AL/0251).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement There are no data available for sharing since the permissions for data use from the NHS organisations which are the data controllers only allow access in a Safe Haven environment by the research team. Further details of the DQIP intervention components, including the IT tool used, are available from the authors on request.

Open Access 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/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Hakkarainen KM, Hedna K, Petzold M, *et al.* Percentage of patients with preventable adverse drug reactions and preventability of adverse drug reactions—a meta-analysis. *PLoS One* 2012;7:e33236.
- Howard R, Avery A, Bissell P. Causes of preventable drug-related hospital admissions: a qualitative study. *Qual Saf Health Care* 2008;17:109–16.
- National Institute for Health and Care Excellence (NICE). Costing statement: Medicines optimisation - Implementing the NICE guideline on medicines information (NG5). 2015. <https://www.nice.org.uk/guidance/ng5/resources/costing-statement-6916717> (accessed 11 May 2016).
- IMS Institute for health care informatics. Avoidable Costs in U.S. Healthcare. 2013. http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Corporate/IMS%20Institute/RUOM-2013/IHII_Responsable_Use_Medicines_2013.pdf
- Howard RL, Avery AJ, Slavenburg S, *et al.* Which drugs cause preventable admissions to hospital? A systematic review. *Br J Clin Pharmacol* 2007;63:136–47.
- Budnitz DS, Lovegrove MC, Shehab N, *et al.* Emergency hospitalizations for adverse drug events in older Americans. *N Engl J Med* 2011;365:2002–12.
- Leendertse AJ, Egberts AC, Stoker LJ, *et al.* Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. *Arch Intern Med* 2008;168:1890–6.
- Grant AM, Guthrie B, Dreischulte T. Developing a complex intervention to improve prescribing safety in primary care: mixed methods feasibility and optimisation pilot study. *BMJ Open* 2014;4:e004153.
- Dreischulte T, Grant AM, McCowan C, *et al.* Quality and safety of medication use in primary care: consensus validation of a new set of explicit medication assessment criteria and prioritisation of topics for improvement. *BMC Clin Pharmacol* 2012;12:5.
- Guthrie B, McCowan C, Davey P, *et al.* High risk prescribing in primary care patients particularly vulnerable to adverse drug events: cross sectional population database analysis in Scottish general practice. *BMJ* 2011;342:d3514.
- Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006;6:54.
- Dreischulte T, Donnan P, Grant A, *et al.* Safer prescribing—a trial of education, informatics, and financial incentives. *N Engl J Med* 2016;374:1053–64.
- Dopson S, Locock L, Chambers D, *et al.* Implementation of evidence-based medicine: evaluation of the Promoting Action on Clinical Effectiveness programme. *J Health Serv Res Policy* 2001;6:23–31.
- Glasziou P, Meats E, Heneghan C, *et al.* What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008;336:1472–4.
- Hoffmann TC, Glasziou PP, Boutron I, *et al.* Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687.
- Shekelle PG, Pronovost PJ, Wachter RM, *et al.* Advancing the science of patient safety. *Ann Intern Med* 2011;154:693–6.
- Grant A, Dreischulte T, Treweek S, *et al.* Study protocol of a mixed-methods evaluation of a cluster randomized trial to improve the safety of NSAID and antiplatelet prescribing: data-driven quality improvement in primary care. *Trials* 2012;13:1–6.
- Glasgow RE, McKay HG, Piette JD, *et al.* The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management? *Patient Educ Couns* 2001;44:119–27.
- Grant A, Dreischulte T, Guthrie B. Process evaluation of the Data-driven Quality Improvement in Primary Care (DQIP) trial: case study evaluation of adoption and maintenance of a complex intervention to reduce high-risk primary care prescribing. *BMJ Open* 2017;7:e015281.
- Grant A, Dreischulte T, Guthrie B. Process evaluation of the data-driven quality improvement in primary care (DQIP) trial: active and less active ingredients of a multi-component complex intervention to reduce high-risk primary care prescribing. *Implement Sci* 2017;12:4.
- Information Services Scotland. General practice. <http://www.isdscotland.org/Health-Topics/general-practice/> (accessed 12 Nov 2015).
- May C, Finch T. Implementing, Embedding, and Integrating Practices: An Outline of Normalization Process Theory. *Sociology* 2009;43:535–54.
- Lynn MR. Determination and quantification of content validity. *Nurs Res* 1986;35:382–6.
- Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. *J Clin Epidemiol* 2015;68:627–36.
- Guthrie B, Donnan PT, Murphy DJ, *et al.* Bad apples or spoiled barrels? Multilevel modelling analysis of variation in high-risk prescribing in Scotland between general practitioners and between the practices they work in. *BMJ Open* 2015;5:e008270.
- Finch TL, Rapley T, Girling M, *et al.* Improving the normalization of complex interventions: measure development based on normalization process theory (NoMAD): study protocol. *Implement Sci* 2013;8:1–8.
- MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol* 2007;58:593–614.