Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial

Martin Eccles, Nick Steen, Jeremy Grimshaw, Lois Thomas, Paul McNamee, Jennifer Soutter, John Wilsdon, Lloyd Matowe, Gillian Needham, Fiona Gilbert, Senga Bond

Summary

Background Radiological tests are often used by general practitioners (GPs). These tests can be overused and contribute little to clinical management. We aimed to assess two methods of reducing GP requests for radiological tests in accordance with the UK Royal College of Radiologists’ guidelines on lumbar spine and knee radiographs.

Methods We assessed audit and feedback, and educational reminder messages in six radiology departments and 244 general practices that they served. The study was a before-and-after, pragmatic, cluster randomised controlled trial with a 2×2 factorial design. A random subset of GP patients’ records were examined for concordance with the guidelines. The main outcome measure was number of radiograph requests per 1000 patients per year. Analysis was by intention to treat.

Findings The effect of educational reminder messages (ie, the change in request rate after intervention) was an absolute change of −1.53 (95% CI −2.5 to −0.57) for lumbar spine and of −1.61 (−2.6 to −0.62) for knee radiographs, both relative reductions of about 20%. The effect of audit and feedback was an absolute change of −0.07 (−1.3 to 0.9) for lumbar spine of 0.04 (−0.95 to 1.03) for knee radiograph requests, both relative reductions of about 1%. Concordance between groups did not differ significantly.

Interpretation 6-monthly feedback of audit data is ineffective but the routine attachment of educational reminder messages to radiographs is effective and does not affect quality of referrals. Any department of radiology that handles referrals from primary care could deliver this intervention to good effect.

Lancet 2001; 357: 1406–09

Introduction

General practitioners (GPs) can overuse radiological tests, particularly lumbar spine and knee radiographs. Such tests are frequently of little clinical use. Guidelines for use of these investigations are in the UK Royal College of Radiologists’ publication Making the best use of a radiology department. However, few studies have been done of interventions designed to change GPs’ behaviour. Although these studies showed that GPs altered their use of radiological tests, they were badly designed, used inappropriate analysis, had short duration of follow-up, or omitted cost considerations. Grol and Lomas have summarised the theory of how to change doctors’ behaviour, and Oxman and colleagues have reviewed the effectiveness of interventions. Specific prompts at the time of consultation are a powerful strategy and have been shown to alter GPs’ behaviour—eg, when referring patients for infertillity investigations—but the effect of the widely-used strategy of audit and feedback is not so certain.

We assessed two methods (audit and feedback, and educational messages) of reducing GPs’ requests for radiological tests in accordance with the UK Royal College of Radiologists’ guidelines. Our hypothesis was that either intervention alone would be more effective than a control and that both interventions together would be more effective than either alone.

Methods

Study design

The study was based in six radiology departments in the north-east of England and Scotland and in GPs’ surgeries (practices) that referred patients exclusively to them. The study was a before-and-after, pragmatic, cluster randomised controlled trial, with a 2×2 factorial design—practices were the units of randomisation and practice size, was done by the study statistician (NS) with computer-generated random-number tables. The study was approved by the West Midlands Multi-site Research Ethical Committee.

Procedures

A group of GPs and consultant radiologists wrote referral guidelines and educational messages for lumbar spine and knee radiographs (based on the Royal College of Radiologists’ guidelines and the Royal College of General Practitioners’ back-pain guidelines). The referral guidelines were sent by post to all GPs. Each practice was then randomly allocated to: audit and feedback, or control; and educational messages or control.

Feedback was prepared by the research team from routine data provided by the radiology departments. It covered the previous 6 months and was sent to GPs at the start of the intervention period and 6 months later. Feedback contained the number of requests for lumbar spine and knee radiographs made by the whole practice compared with requests made by all GPs in the study.
Educational messages were attached to the reports of every knee or lumbar spine radiograph requested during the 12-month intervention. Lumbar spine radiographs, “in either acute (<6 weeks) or chronic back pain, without adverse features, radiograph is not routinely indicated”; knee radiographs, “in adults with knee pain, without serious locking or restriction in movement, radiograph is not routinely indicated”. In three radiology departments, messages were attached by reporting software, and in two, coloured stickers were attached to radiographs by hand. In one department, messages were added by reporting software after an operator pressed a preprogrammed key for each radiograph. Rates of message attachment were checked regularly.

The main outcome was the number of each radiograph requested per 1000 patients registered with every practice per year for 2 years; the second year was the intervention period. The research team knew which practices received which intervention but had no personal contact with them. Furthermore, the outcome measure was objective and recorded by radiology departments, hence there was no possibility of bias.

To assess concordance with the referral guidelines, we examined the medical records of patients from a random subset of practices for whom radiographs were requested in the intervention period. Requests for radiographs were classified as concordant with the guidelines, or not. Only GPs who had requested at least one radiograph were sampled.

Statistical analysis
With historical data from a radiology department and standard methods for a quantitative variable, we calculated that 55 practices in each group would have 80% power to detect a 15% difference in radiograph requests, assuming a type-I error of 0.05. At that stage we had no information about practice size. Previous studies had used the ratio of radiograph requests after intervention to the number of requests preceding it as a dependent variable that took some account of practice size. We were able to estimate the SD of this ratio and based the sample size on detecting a difference of 0.15—a reduction in requests of 15%. However, a ratio is a very crude way of correcting for practice size. We decided that a more appropriate analysis could be undertaken if we could collect accurate information about practice size. This was done as part of the study. From these data, we calculated that a factorial design with four groups required 220 practices. With no interaction effect, the estimate of main effects would be based on a comparison of two groups of 110 practices and have 80% power to detect an 11% difference in radiograph requests. To detect a 12.5% difference in concordance with 80% power, 5% significance, and assuming an intraclass correlation of 0.15, five requests per condition were needed from 162 doctors (81 in each group).

We analysed request rates with the multilevel modelling package MLwiN (version 1.0), assuming that rates were normally distributed. Variation between practices and between years were modelled as random effects. Treatment effects (audit and feedback, and reminder messages) were included as fixed effects. Because rate estimates are more precise for larger practices, a weighted least-squares procedure (with practice-list size as the weight) was used to assess the models. We also analysed concordance of requests and guidelines by MLwiN with requests nested within GPs, which were nested within practices. Requests were deemed either concordant with the guidelines or not. We assumed a binomial error structure and used a logit link function. The log (odds) of a request being appropriate was assumed to vary randomly across GPs and practices. We then included fixed treatment effects in the models as above. Both analyses were by intention to treat. Although the study was a cluster randomised trial, the main analysis was undertaken on summary statistics corresponding to the unit of randomisation, and clustering was not a problem. Two participating departments were only able to provide data at the practice level. We were therefore unable to undertake a GP-level analysis that included all sites.

Results
247 practices were enrolled, three of which dropped out (figure 1). To assess concordance with the guidelines, data were abstracted from 1693 patients’ records of 162 GPs in 48 practices. The audit and feedback intervention was delivered to all eligible GPs according to study design. Attachment rate of educational reminder messages was close to 100%, or was 100%, in departments in which messages were attached electronically; was 100% in departments in which messages were attached by hand; and around 40%
in that in which an operator pressed a key to add the message.

Table 2 shows the rates of radiograph requests. Reminder messages reduced request rates more than the control intervention (paper copies of guidelines only). Audit and feedback also had an apparent effect. However, when we modelled the data and added sources of random variation, the effect disappeared. Some of the reduction in the audit and feedback group is probably attributable to regression to the mean. Rates were most reduced in practices that received both interventions, but the number of requests at baseline were significantly larger for this group than others (p<0.05). We modelled the data assuming that all changes in table 2 were attributable to intervention and showed that audit and feedback significantly reduced requests for knee radiographs (p<0.05). However, because of baseline imbalance we think it is inappropriate to draw conclusions from this analysis.

We analysed practice rates with the weighted least-squares procedure. Estimates of intervention effects are expressed as radiograph requests per 1000 patients and are based on a comparison of the practices that got a particular intervention with those that did not. The effect of educational reminder messages was an absolute change of $-1.53$ (95% CI $-2.5$ to $-0.57$) for lumbar spine radiographs and of $-1.61$ ($-2.6$ to $-0.62$) for knee radiograph requests; these estimates are both relative reductions of about 20%. The effect of audit and feedback was an absolute change of $-0.07$ ($-1.3$ to $0.9$) for lumbar spine and $0.04$ ($-0.95$ to $1.03$) knee radiograph requests. Relative reductions were about 1% (knee) and almost no change (lumbar spine). For both types of radiograph, interaction between interventions was not significant—ie, there was no increased effect of receiving both interventions.

Table 3 shows concordance of the subset of radiographs with guidelines. Sample size was 162 doctors and 788 requests for lumbar spine radiographs (mean 5·2 per doctor) and 905 knee radiograph requests (mean 6·0 per doctor). There were no significant differences between groups. Requests from doctors who had received audit and feedback were no more likely to be appropriate than requests from other doctors. The odds ratio for lumbar spine radiographs was 0·75 (95% CI 0·52–1·07) and for knee was 0·82 (0·50–1·33). For doctors who had received educational reminder messages, the equivalent values were 0·95 (0·63–1·67) and 1·36 (0·86–2·23).

### Discussion

Our results have shown that the routine attachment of educational reminder messages to radiographs can reduce the number reminder messages by 20%. Any department of radiology that takes referrals from primary care could deliver this intervention. Monthly feedback of audit data does not reduce radiograph requests.

Educational reminder messages are a response to information overload, and are easy to deliver and help to implement guidelines. Tierney and McDonald, looking at delivering preventive care, showed that reminders are more effective if received when decisions are made than at another time, although both are effective. They also concluded that it was easier to deliver reminders away from the time of decision and that this might offset a small effectiveness difference. Our results show that reminder prompts delivered after a decision is made can have a useful effect. The prompts probably influence later decisions and this depends on their lasting from one decision to the next. This strategy is likely to be most effective for frequent decisions. Unfortunately, from these data we cannot calculate the minimum frequency or how long the effect might last.

Our results suggest that, for primary-care radiology referrals at least, audit and feedback does not change behaviour. Thomson-O’Brien and colleagues said of audit and feedback used alone, “when it is effective the effects appear to be small to moderate”, and had serious concerns about the studies they examined. Many of their concerns (eg, inappropriate level of randomisation or analysis) would cause an overestimation of effect size. Kerry and colleagues, in a well designed study, showed a positive effect of audit and feedback. However, although significant, the effect was small. Thus there does not seem to be a major role for audit and feedback as a strategy to change behaviour.

Studies of radiology use in primary and secondary care have used unreliable methods. Our results are reliable and unbiased because of the cluster randomised design, intention-to-treat analysis, and allowance for variation at appropriate levels (doctor and practice). Randomised controlled trials have long been recognised as appropriate for clinical trials. Many implementation-research studies have been non-randomised and have not allowed for important effects, such as clustering. Appropriately-designed studies, such as ours, are needed to evaluate interventions.

Ideally, appropriateness is assessed while party to the consultation leading to the request. From our notes-based review, we could only assess concordance with guidelines; for lumbar spine radiographs the rates of concordance were similar to those reported by others. A 12.5% improvement in concordance (estimated in our sample size calculation) from base rates of 40% (spine) and 25% (knee) would give odds ratios of 1·66 and 1·80, respectively. Audit and feedback effects were much smaller than these, and educational reminder messages for lumbar spine radiographs were also unlikely to have an effect of 12.5%. However, the odds ratio of 1·36 for knee radiographs was a 6% increase in concordance (from 25%
to 31%). Although this increase was not significant, the upper CI of 2.23 was an increase in concordance of 17% (from 25% to 42%). Thus, there was possibly an effect on concordance with guidelines, for knee radiographs, close to the target of 12.5%.

We assumed that avoided radiographs were not replaced with other action, but patients who were not referred for radiograph might have been managed with an alternative strategy, such as referral for an outpatient appointment or a different imaging procedure. However, such effects on secondary care referral were not shown in a study of requests for biochemical tests,19 in which doctors who responded best to reminders had the lowest referral rates. Our study has shown that reminders can substantially reduce GPs' use of radiographic services. Our straightforward, reproducible interventions could be used with other services and in other settings.

Contributors
Martin Eccles was the main investigator and report writer, and helped manage and run the study. Senga Bond was responsible for management and running, and helped write the report. John Wilson helped with management, report writing, and specialist input. Nick Steen did the statistical analysis and helped write the report. Paul McNamee helped to write the report. Jeremy Grimshaw helped with management and writing the report. Lloyd Matowe, Jennifer Soutter, and Lois Thomas helped with management and running, and helped write the report. Fiona Gilbert and Gillian Needham provided specialist input.

Acknowledgments
We thank the radiology departments and general practices for commitment and enthusiasm, and Norma Cardill for secretarial support. We thank the radiology departments and general practices for management and running, and helped write the report. Fiona Gilbert and Gillian Needham provided specialist input.

References
22 Tierney WM, McDonald CJ. Delayed feedback of physician performance versus immediate reminders to perform preventive care: effects on physician compliance. Med Care 1986; 24: 659–66.