Author's response to reviews

Title: A pragmatic cluster randomised controlled trial of a Diabetes REcall And Management system: the DREAM Trial [ISRCTN32042030].

Authors:

Martin P Eccles (martin.eccles@ncl.ac.uk)
Paula M Whitty (p.m.whitty@ncl.ac.uk)
Chris Speed (chris.speed@ncl.ac.uk)
Ian N Steen (nick.steen@ncl.ac.uk)
Alessandra Vanoli (alessandra_vanoli@yahoo.it)
Gillian C Hawthorne (gillian.hawthorne@newcastle-pct.nhs.uk)
Jeremy M Grimshaw (jgrimshaw@ohri.ca)
Linda J Wood (linda.wood@diabetes.org.uk)
David McDowell (david.mcdowell@prowellness.com)

Version: 2 Date: 19 October 2006

Author's response to reviews: see over
A pragmatic cluster randomised controlled trial of a Diabetes REcall And Management system: the DREAM Trial [ISRCTN32042030].

Thank you for your correspondence about this paper. Below I list the concerns raised by the three helpful referees’ reports and our responses to them.

Reviewer 1.
Major Compulsory Revisions
Reviewer’s comment 1. Introduction: I was surprised that the introduction focuses the review by Shojania et al as this was published after the DREAM trial started. The introduction to the original protocol more clearly explains why the authors chose this approach in the context of when the study started. I would have thought it more appropriate to present the comments relating this study to the review in the discussion section. Otherwise one would have to ask why the authors chose this intervention when the review found that the categories of intervention of “provider reminders” and “audit and feedback” were not particularly effective.

Response: We don’t agree with the reviewer’s suggestion that we do not mention in the introduction information that has emerged during the course of the study (and will potentially be known to readers). However we agree that we could better locate this in relation to the reasons for our choice of study intervention. To this end we have added the following text to the introduction after the second sentence:

“At the time of setting up the DREAM trial, computerised central recall systems for patients and their family doctors had been supported by the evidence from a 1999 systematic review (ref: Griffin S, Kinmonth AL. Diabetes care: the effectiveness of systems for routine surveillance for people with diabetes [Cochrane Review]. The Cochrane Library. Issue 4 ed. Oxford: Update Software; 1999.). However, the evidence base on which these conclusions were based was limited to that from patient- rather than practice-randomised trials, in selected practice samples, and without economic evaluation. Thus the effectiveness of an area wide, patient focussed, structured recall and management system (in terms of process of care, patient outcome and economic impact) remained unknown.”
Reviewer’s comment 2. Data collection: I would be concerned that the different time frames for collecting data pre and post intervention would bias results – particularly as it gives more time for process measures of care to be carried out in the intervention period.

Response: The important design feature is that the data collection periods were the same for intervention and control groups; this was not a source of bias when estimating the effects of the intervention. The different time frames for collection data pre and post intervention might have presented analytic difficulties had we analysed change in those measures of process directly but by using analysis of covariance we ensured that such difficulties did not arise.

Reviewer’s comment 3. The intervention: needs to be described more clearly – how does it work at patient level – does a GP have to go on to the register to be prompted or does it occur automatically once a patient is seen for any reason?

Response: We have extended the description of the intervention by replacing the first three sentences of the previous version with this fuller description: “In the enhanced structured and recall management system, a ‘circle of information exchange’ was established between the participating general practices and the database. The central database system identified when patients were due for review and generated a letter to the patient asking them to make an appointment for a review consultation. The rules for generation of review letters were adapted for each PCT area. In one PCT, the system acted as a prompting system for annual review, and patients were identified 11 months after their last diabetes appointment. In the other two PCTs, patients who had missed annual reviews were identified by searching for patients who had not had a diabetes appointment for 14 months or more. At the same time, the central database generated a letter to the practice stating that the patient should be making a review appointment in the near future. The letter to the practice included a ‘structured management sheet’ (to be held in the patient’s record) to capture an agreed minimum data set to be collected during the consultation (Box 1). This management sheet also contained relevant prompts, tailored to a patient’s known clinical or biochemical values, derived from locally adapted, national evidence based guidelines (see Additional File 1).

When the patient was seen in the practice, the primary care professional (usually the practice nurse) completed the management sheet and returned a copy for entry onto the central register within a designated period of time. This circle of information was broken if the patient did not visit the general practice as planned or the general practice did not return the management sheet to the central register. If this happened, the central register would print reminder letters and further structured management sheets at the next routine database search by the diabetes register facilitator (usually at least weekly).

In addition to this cycle based on annual reviews, routine ongoing structured management sheets were produced every time a patient in an intervention practice was identified by the diabetes register facilitator on the register database. For example, when data was inputted on the database for any reason, the system would print a structured management sheet updated for any new data and relevant management prompts, and this would be sent to the relevant practice.
The trial intervention ran for 15 months, commencing on 1 April 2002 and ending on 30 June 2003. The letters to patients to invite them for annual review were only commenced in October 2002 (delayed to overcome concerns about the accuracy of patient details on the database up to this point). The enhanced system was also capable of producing patient letters to accompany routine ongoing structured management sheets for practices but because of difficulties operating this element of the software it was not possible to run this feature during the lifetime of the trial. This was a deviation from the published protocol.”

Reviewers comment 3 (continued). The authors also suggest that the system was not fully operational for the whole time of the intervention period – need more detail. This relates to a need for the authors to consider treatment fidelity – the results of the interviews with practice-based informants are not presented.

Response: The trial was designed as a pragmatic trial (and the title of the paper has been amended to highlight this). The results reflect those that could be expected to be achieved by any healthcare administrative structure introducing a similar system. However, the reviewer is correct that the patient prompt system was never fully operational and we identify this in the paper. We can speculate that this might have resulted in a larger effect so the following phrase highlighting this element has been added to the ‘limitations of methods’ part of the discussion: ‘While the intervention was in place for the planned 15 months, the full intervention ran for only 9 months and the intended patient intervention was never fully operational.’

The results of the interviews with practice based informants are reported in the results section on Economic Data – we have identified this more clearly in the text.

Reviewer’s comment 4. I was surprised at the high proportion of people on diet alone and the low proportion on insulin – this should be addressed in the discussion as it may reflect a bias in the register. Also proportions on aspirin seem very small.

Response:
There are several possible reasons for this, which are now included in an additional paragraph in the discussion, as follows:
“There is some suggestion of under-recording on the registers, with an apparently low proportion of people on insulin and on aspirin. This is mirrored by an apparently high proportion of people on diet alone. However, there are several other possible explanations for these findings: low aspirin rates could be due to acquisition of aspirin by patients themselves ‘over the counter’ rather than on prescription (common in the UK). Self-reported aspirin use was higher than on the registers. People being treated for their diabetes solely by hospital, who are more likely to be treated on insulin and less likely to be on diet alone, were excluded from the study. However, while we have the same rates for diet alone from the register and from self-report, self-reported insulin use was considerably higher than on the registers. This suggests that insulin use was under-recorded on the registers but equally so for both intervention and control groups.”

To provide the data for these comments, we have included an additional table in the results (a new Table 3) , providing the self-reported medication data.

Our final point is that whilst we agree that the low proportions may reflect an inaccuracy we don’t agree that it reflects a bias in the register.
Reviewer’s comment 6. I am aware of a publication relating to an adjacent region regarding district wide diabetes care. (1) Though this was not a randomized controlled trial, I would have thought it would have been referred to, particularly as the process measures of care are better than in the intervention group in this study – why might this be the case?

Response: There are many papers that report local process measures. We could pick any number of them to compare with our data to but chose to leave the reader to make their own comparisons. The main point of the discussion is to deal with the interpretation of the impact of the intervention and we have focussed on that. It is also the case that the area that the reviewer’s suggested paper refers to is highly atypical in being the geographical area served by the diabetologist who subsequently went on to become England’s “Diabetes Tsar”.

Reviewer’s comment 7. Overall, I felt the discussion was weak and that the results of this study should be placed more in context with other international studies and the limitations such as the low response rate should be addressed more comprehensively.

Response: We find the reviewer’s comment a little difficult to understand. The results of the study are placed in the context of the most comprehensive and up-to-date review of QI interventions in diabetes care.

The implications of the response rate are picked up in the discussion of patient-reported outcomes referred to below.

The discussion has also been expanded in several areas as detailed earlier.

Reviewer’s comment 7 continued: There also needs to be discussion relating to the patient-based outcomes.

Response: We are grateful to the reviewer for spotting this omission. We have expanded the discussion to cover patient-reported outcomes, including the reference to the observed ICCs and the possible implications of response rates to the questionnaire survey for all of the self-reported data, with the following sentences:

“We showed no significant difference in patient-reported outcomes between intervention and control groups. The observed clustering in the outcome scores was smaller than that assumed in the sample size calculation and, as we achieved the desired sample size, the lack of significant changes in patient outcomes is unlikely to be due to a lack of power. We do however have to consider the possibility of non-response bias for all the self-reported data with a response rate of 51%, even though there was no difference between intervention and control group response rates or on sociodemographic variables.’

Minor Essential Revisions
Reviewer’s comment 1. Why did one of the PCTs withdraw?

Response: We have added the following sentence to the methods:

“Several factors led to the withdrawal of this PCT. Despite our having appropriate administrative approval, when the trial began it became apparent that the administrative authority did not have the co-operation necessary for all of the GPs to
participate in the trial. Consequently we had to enrol individual practices directly (rather than via the PCT), which resulted in fewer practices enrolling and our being at risk of not achieving our required sample size. We recruited a further PCT to address this problem, However, the original PCT then suspended involvement with the diabetes register and their practices had to be excluded from the study. This was a deviation from the published protocol.”

Reviewer’s comment 2. Study patients: How does the register deal with defaulters from care?

Response: The extended description of the intervention within the methods explains how the system operates when the ‘cycle of information exchange’ is broken (which includes defaulters from care).

Reviewer’s comment 3. Is there any idea of how many attend specialist services as well as GPs. The authors refer to shared care, but make no further mention of it so I presume there isn’t a formal shared care arrangement with the specialty services?

While there were fields in the register to indicate ‘sole GP care’ and ‘shared care’, there was no single field to identify those patients receiving specialist care only (which we required to eliminate them from the analysis). This was by design by the software developer, who had not wished to suggest that patients could receive no GP care. This information had to be imputed from four fields titled “care by”, “nature”, “sub-nature” and “owner”. The information about who was responsible for the care of the patient was not recorded in a consistent way between the two registers. Each register supplied us with an algorithm for determining which patients were being treated solely in hospital. These patients were then omitted from the analysis.

There was no formal shared care system in place in the PCTs covered by the registers. However, one of the authors (GH) working in the area at the time of the study reports that less than 20% of patients were attending specialist services as well as GPs. We have added the following sentence to the methods:

‘At the time of the study, approximately 20% of patients received shared care.’

4. Which version of the DCS questionnaire was used?

Response: We are grateful to the reviewer for highlighting this, as the reference in our submitted paper was incorrect (ref 12), and has now been amended. We were supplied with the version of the questionnaire used by the author.

5. In the analysis section the authors state that this paper will be reporting results based on this model – are they also planning to report results in a different way elsewhere?

Response:
This comment refers to the final paragraph of the ‘analysis’ section: “In addition to the above analyses that were pre-specified, because of large systematic differences between the two registers a further model was fitted which included a register effect. In this paper we report estimates of the effect of the intervention based on this model.” As part of the analytic strategy we analysed the data using a number of alternative models. In the version of the paper originally submitted we included estimates of the effects of the
intervention based on just one of the models. In response to comments by reviewer 3 we now provide an additional electronic table that will enable an interested reader to compare the results from all the models.

6. The participant flow sheet should include a record of numbers of participants per cluster as well as cluster numbers (2)

Response: We have amended Figure 1 to include the mean number of patients per cluster.

7. Deviations from the protocol should also be addressed.

Response: The two deviations from the original study protocol are now identified in the text.

8. The economic analysis is important but is limited by an inability to link costs with outcomes – I would have thought that would have been possible given that the authors claim that the intervention is effective in some areas. However, it is important that they give the figures and the discussion addresses the limitations of the economic analysis.

Response: The reviewer is suggesting that we produce a cost effectiveness ratio for each successful measure of effectiveness. We have not done this, instead, choosing to remain with a cost-consequences or profile approach. Our approach is well recognised and described in standard economic texts (Drummond MF, Sculpher M, O’Brien BJ, Stoddart GL and Torrance GW (2005) Methods of Economic Evaluation in Health Care (3rd Edition). Oxford University Press, Oxford.). We have added the Drummond et al reference to support our use of the cost-consequences approach in this study. It would also not have been appropriate in DREAM to derive cost-effectiveness ratios because the difference in costs due to the intervention was neither statistically significant nor economically significant. Moreover, even if appropriate, we would not have been able to derive a meaningful cost-effectiveness ratio because our economic evaluation was 'partial' (see Drummond et al) since the costing of the intervention was incomplete due to the lack of data on the costs to the intervention practices. We have reflected this as a limitation of our methods in the discussion by adding the following sentence:

‘Our assessment of costs incurred by the practices was limited and so we have only suggested a (hypothetical illustration….)’

There are more general arguments rehearsed in favour of a cost consequences (or alternatively called a balance sheet) approach in health economics, especially for cost benefit analysis as compared with cost utility analysis (eg using cost per QALY)(McIntosh E, Donaldson C, Ryan M. Recent advances in the methods of cost-benefit analysis in healthcare. Matching the art to the science. Pharmoeconomics 1999; 4: 357-367.). There has also been a recent high profile paper criticising the approach to economic evaluation that reduces everything to cost per QALY, and advocated (again) a cost-consequences approach (Coast J. Is economic evaluation in touch with society’s health values? BMJ 2004;329;1233-1236.).

9. The authors refer to a significant lowering of mean cholesterol – this should state a significant difference between control and intervention groups.
Response: The relevant sentence in the discussion has been amended to read “We did, however, show a modest and statistically significant lowering of serum cholesterol of 0.15 mmol/l in the intervention group compared to the control group.”

9. Continued: Also, lower cholesterol levels are more likely to relate to statin prescribing than just dietary advice in this population – this could be discussed.

Response: As we are now reporting the patient-reported medication data, which showed no impact on statin prescribing, we have even less basis to suggest that this impact could have been due to statin prescribing. The relevant sentence in the discussion has been expanded as follows:

“As the impact of the intervention on medication, including lipid lowering therapy, was unclear from the register-derived data and negative from the questionnaire data, it is possible that this effect may be due to the increased delivery of dietary advice – one of the four areas of improvement in provider adherence to recommended care.”

Reviewer 2.
Major compulsory revisions
1. The analysis uses cross sectional values for A1c, BP, LDL, etc. For some of these variables, it may be interesting to see if the predictors were related, not only to values cross-sectionally, but to change in these values over time. The authors may consider such analysis, and could either note why they decided not to do such analysis, or else mention whether the results confirm or disconfirm some of the stated findings of the study.

Response: The analyses reported adjusted for baseline values as explained in the response to the statistical reviewer below and now also highlighted in the text of the paper. They were thus analyses of change over the time of the intervention period.

2. On page 5, the ICC values are quite large. This may be worth noting in a comment, because the size of these ICCs explains why some apparently quite sizeable differences in care are not statistically significant. This is an issue that many readers will not be familiar with.

Response: See response to statistical reviewer below.

3. On the 4th line on page 6, there is “an effect size of approximately 0.25 in such measures.” Please put a unit on this number 0.25.

Response: An effect size of 0.25 corresponds to a difference in mean scores between groups equal to 0.25 multiplied by the standard deviation of the outcome measure of interest.

4. The validity of the cost data relies on the accuracy of patient reporting of care use. Please discuss this more as a serious limitation. Please provide some references that assess the accuracy of such self-reported utilization information, form other studies.

Response: We have added the following sentence to the methods: ‘Patients have been shown to be able to report these data reliably’ accompanied by the following reference: Thompson S and Wordsworth S on behalf of the UK Working Party on patient Costs

5. The narrative description of the intervention (page 8) needs to be clearer and more informative.

Response: We agree and have extended the description of the intervention as explained in our response to reviewer (1)’s comment number (3).

6. On page 12, second paragraph, there is a statement that “this (lipid) effect may be due to the increased delivery of dietary advice…” This actually seems quite unlikely in light of the fact that lipid drug use increased substantially (although not quite statistically significantly) in the intervention practices. I would delete this sentence, and suggest instead that lipid-lowering drug therapy is the core of lipid management in diabetes patients—to reinforce the importance of such care by physicians.

Response: Please see our response to reviewer (1)’s comment number 9. Because we have self-reported medication data which also suggests no impact on lipid lowering therapy, we do not think it appropriate to remove this sentence.

7. The sentence towards the bottom of page 12 that states “This is particularly important when, as in this case,…” Has words like “secondary care” and “responsibly for expenditure incurred in primary care” that are quite opaque to me, a reader in the U.S. I would drop the lingo and put in more descriptive wording.

Response: we have modified the description in this paragraph as follows: “This is particularly important when, as in this case, an innovation can reside in specialist services or hospital care where there is no responsibility for expenditure incurred in family or general practice.”

8. There are few references to prior diabetes research. These should be augmented. The publications of Steve Shortell, Victor Montori, Patrick O’Connor, and James Meigs come to mind. This is of course, subject to journal limits on number of refs, etc.

Response: While there is no journal limit on references, we think that by grounding our discussion in the most up to date international systematic review, we are linking the study’s findings systematically to previous international studies in this field.

Reviewer 3. Dr Martin Lee (statistician)
Major Compulsory Revisions

1) The sample size calculations were based on what appears to be an unadjusted comparison of the rates and means, yet the primary analysis is a regression adjusted comparison. What impact, positive or negative, could this have had on the study? In addition, the authors have assumed very large ICCs in these calculations. Could they provide estimates of the actual values seen? One of the other reviewers commented that this might have had an undue effect on the lack of significance, but this would be the case if the authors had underestimated the degree of clustering. Thus, it would be important to know if this, indeed, is an issue for the study.
Response: In the sample size calculation, in the analysis of patient outcomes such as the SF-36 we assumed an ICC of 0.07. In practice the observed ICCs for variables derived from the patient survey were smaller than this. The ICC for the SF-36 mental and physical health component scores were 0.02 and 0.03 respectively. For process and intermediate outcome variables the sample size calculation assumed an ICC of 0.14. In practice the observed ICCs were extremely variable and ranged from 0.01 (last recorded creatinine value) to 0.7 (was there a recording of an albumen creatinine ratio in the last 12 months). The ICCs quoted above were calculated with just a constant term in the regression model. When a register effect was fitted the ICCs were much reduced for many of the variables. For the variable described in the sample size calculation – was there an HbA1c recorded in the last 12 months – the ICC was 0.40 with no covariates in the model but fell to 0.30 with the inclusion of a register effect. The corresponding ICCs for the actual value of the last recorded HbA1c were 0.15 and 0.05 respectively. It is actually difficult to make general statements about the power to detect a particular effect. It is perhaps easier to consider the interval estimates for the effect of the intervention that are provided in the main tables of results. The effect of the intervention on the recording of HbA1c was not statistically significant but the 95% CI for the odds ratio was (0.81, 3.08). The odds of an HbA1c being recorded in an intervention practice might have been as much as three times the odds of an HbA1c being recorded in a control practice which would no doubt be regarded as clinically significant effect. For this variable the higher than expected clustering may have compromised our ability to detect a difference. However this is not the case for the actual clinical measures where the observed clustering was smaller. If we consider the difference between groups in the mean (last recorded) HbA1c the 95% confidence interval was from -0.18 to 0.10. We suspect that in this case we can conclude that there is unlikely to have been a significant clinical change in this variable and certainly not one that is consistent with that reported elsewhere. The discrepancy between the sample size calculation and specification of the analysis is addressed below.

2) There is a clear issue of patient non-response here, which is acknowledged in the Data Collection section and actually manifests itself quite significantly in the study with a 51% response rate. Was there any attempt to adjust the analyses for this or, at the very least, evaluate the data for any non-response bias?

Response: Unfortunately all we can do is acknowledge that this is potentially a problem for the variables derived from the patient survey (see response to reviewer 1’s comment 7). Within the structure of our trial there is very little that we can do to try and quantify this. A key issue in this respect was the requirement by the ethics/research governance organisations that we were unable to hold any patient identifiable data within our academic institution. We were supplied with a list of names and addresses of patients to whom we could send out a patient survey but were not allowed access to link that information with individual patient records on the registers. Thus we were unable to compare the clinical characteristics of respondents and non-respondents.

3) Some of the regression models are adjusted using "baseline" variables. I could find any indication as to what those variables were and whether they were identified in advance. If they were selected on the basis of statistical significance, then this becomes an "ex post facto" issue subject to Type 1 error.
4) In the same vein as above, the register effect was unexpected and added to the analysis after the fact when "large systematic differences" were found. It would, first of all, be important to state what this statement means and how it was identified. Clearly this was not part of the initial analytical plan. Therefore, it would be very important for the authors to reanalyze the data without this adjustment and, if the results are different, provide a caveat to the reader as to the legitimacy of the adjusted results.

5) Were the baseline comparisons between the control and intervention groups in Table 1 made with a cluster-adjusted analysis? If so, these should be unadjusted. In any event, it should be made clear to the reader. Further, the authors should state what statistical method(s) were used for these analyses.

6) For Table 2, it is very common practice today to present the unadjusted comparisons between the groups, so that the reader can see the effect of adjustment on the results. This is particularly important here given the uncertainty as to how the adjustment variables were selected.

7) Comment 6 is also relevant to Tables 3 and 4 with respect to the adjustment for register differences.

Response: Points 3 to 7 relate to adjustments made to the regression models that we report in the paper. We actually considered two sets of adjustments.

1. **Baseline values included as a covariate.** Here we failed to make it clear that for those variables derived from the register databases we had data on the process and outcomes of care both before and after the intervention and that, for each variable considered, the post intervention measure was specified as the dependent variable and the corresponding pre-intervention measure was specified as a covariate. So in the analysis of each of these variables there was only one baseline covariate—the value of the process or outcome measure prior to the intervention.

   The main effect of including baseline performance/outcome as a covariate was to reduce the width of the interval estimates of effect size slightly. In much of the work we have done in primary care past performance of practices has been a good predictor of current/future performance. When we include past performance/outcome as an explanatory variable we explain a significant amount of the variation in current performance/outcome with and the variance of the residual error terms is reduced. This is the primary justification for the collection of baseline performance and including it as a covariate; effectively we are assessing change in performance rather than absolute performance following the intervention.

   This adjustment was pre-specified although, as the reviewer notes, it was not incorporated into the sample size calculation. The main reason for not including it in the sample size was to reduce the number of assumptions made at that stage. It was believed that a calculation based on the analysis of absolute performance/outcome following the intervention would be conservative. It was felt that it was not possible to quantify with any certainty at that stage the gains that would be made from the inclusion of past performance/outcome as a covariate.

   As we only undertook a single postal survey of patients at the end of the intervention period, for those variables derived from patient questionnaires, we had no pre-intervention measure and no adjustment for differences at baseline was
possible.

2. **Inclusion of a register effect.** We allowed for a fixed difference in each measure of process and outcome between the two different registers. This adjustment was not pre-specified. This adjustment was possible for variables derived both from the register databases and the patient survey. In general the differences between registers were extremely small for the variables derived from the patient survey but very large for the variables arising from the register databases. The reviewer asks how it was that these differences came to light. In practice the differences were so large they were picked up during the data cleaning and validation phase of the study. It was felt that it would be inappropriate to ignore them during the main analysis.

In general, for the variables derived from the register databases, the main effect of including this effect in the regression model was to reduce the size of the interval estimates slightly.

In our “in house” final report we included both the adjusted and unadjusted comparisons. For the variables derived from the patient survey there were only two comparisons—with and without and adjustment for a difference between registers. For the variables derived from the register databases we presented four comparisons—unadjusted; adjusted for baseline performance/outcome; adjusted for a difference between registers; and then adjustments for both effects simultaneously. The advantage of having a single table with all the comparisons in is that it is possible to gauge the effect of each adjustment on both the point estimate of the effect of the intervention and on the precision of the estimate (by considering the width of the confidence intervals). The disadvantage is that the sheer volume of information in the table makes it less easy for the reader to get an overview of the effects of the intervention. So when writing this paper a decision was made to include one set of comparisons along with the caveat that the other analyses gave similar results.

The statistical reviewer has made it clear that he would prefer to see alternative estimates of effect size in the tables. Therefore we have retained the simplified versions of table 2 in the paper with links to electronic versions of the complete table 2. In addition we have included unadjusted estimates in Tables 3, 4 and 5.

Additional response to points as they relate to analysis of baseline characteristics of sample.

In the original report interval estimates of the difference between intervention and control practices were provided for each characteristic. In the paper we simply report that none of the differences was statistically significant. The reviewer asks whether the comparisons between control and intervention groups made with a cluster-adjusted analysis, suggests that the comparison should not be adjusted for clustering, and goes on to ask us to specify the analyses undertaken.

The analyses undertaken did adjust for clustering when clustering was observed. There was no adjustment for any difference between registers. For sex the estimated ICC was zero and the ratio of males to females in the intervention and control groups was compared using Fisher’s exact test; the odds of a subject in the control group being female divided by the odds of a subject in the intervention group being female was 1.08 (95% CI: 0.94, 1.23). For age the there was evidence of clustering (estimated ICC = 0.02). When we take into account clustering using a standard random effects model, the difference in mean age between groups was -1.05 years (95% CI: -2.12, 0.03) which was not significant at the 5% level (p = .06). Using an unadjusted independent sample t-test
the estimated difference in mean age was slightly smaller (-0.91 years) but was statistically significant at the 5% level. If we report the unadjusted analysis we would need to explain that the difference in mean age was statistically significant, give the 95% CI: -1.67, -0.15 years and suggest that this is probably not clinically significant. The remaining variables – which describe the baseline management of the sample were all analysed using exactly the same method as all the other therapy variables reported in the paper; we determined the total number of patients in each practice on a particular therapy and then compared intervention and control practices by fitting a negative binomial regression model. Because we have a single observation per practice this method does inherently adjust for the clustering of patients within practices.

8) A measure of variability (probably standard deviation) should be included in Table 3 as it is in Table 4.

Response: Table 3 is now Table 4 and SDs have been added.

I hope these responses are satisfactory but, if you have any further queries, please do not hesitate to contact me.

With best wishes.
Yours sincerely,

Dr Paula Whitty
Senior Lecturer in Medical Care Epidemiology
Institute of Health and Society
Newcastle University
Email: p.m.whitty@ncl.ac.uk