Author’s response to reviews

Title: Decreasing the number of arthroscopies in knee osteoarthritis – the Personalised Knee Improvement Programme (P-KIP). A service evaluation of a de-implementation strategy

Authors:

Timothy Barlow (timbarlow1@hotmail.com)
Timothy Rhodes-Jones (Tim.Rhodes-Jones@uhcw.nhs.uk)
Sue Ballinger (sue.ballinger@uhcw.nhs.uk)
Andrew Metcalfe (A.Metcalfe@warwick.ac.uk)
David Wright (David.wright@uhcw.nhs.uk)
Peter Thompson (Peter.Thompson@uhcw.nhs.uk)

Version: 1 Date: 02 Dec 2019

Author’s response to reviews:

Dear Editor,

Thank you for the reviewers' comments and the opportunity to respond. Please find a detailed response to each comment below. Additionally, we have supplied tracked and untracked copies of the manuscript.

Yours faithfully,

Tim Barlow

REVIEWER ONE

1. Abstract/Conclusions: 'Results suggest that P-KIP reduces the number of arthroscopies performed, while improving patients' knee and general health outcomes. P-KIP has the potential to deliver large efficiency savings and relieve pressure on operative lists, however replication in other sites is required.' --&gt; the authors cannot state that patients' knee and general health outcomes improve as result of P-KIP, because they did not compare the results with a control group.
Thanks. We have changed this sentence to the sentence below, which reflects a factual statement. We expand on this in the discussion section (i.e. the improvement is in line with previous published reports of surgical and conservative care).

“Results suggest that P-KIP reduces the number of arthroscopies performed, and patients who took part in P-KIP had an improvement in their knee and general health outcomes.” Abstract, Line 78

2. Row 116-131. The authors describe that they assessed the barriers among surgeons and patients. However, the barriers mentioned in the manuscript are all described from the perspective of the surgeons. What barriers did the patients experience?

This is a really interesting area and surprised us when we initially investigated it. Briefly, patients generally expected (and to a large degree wanted) a paternalistic interaction where “something” was prescribed. There was not a great preference placed on what that “something” was (although simply “more physiotherapy” was not seen as satisfying that “something” for a proportion of people interviewed). We have included the reference that goes into this in detail, and we have altered the manuscript as below:

- Perceived pressure from patients to do something (with arthroscopy perceived as the management that was desired).

- Patients’ desire for something to be done, but with generally no fixed view on what that “something” should entail (although simply “more physiotherapy” was not desirable from a proportion of people.

Background, Lines 119-123.

3. Row 131: time pressure --&gt; who feels time pressure? And from what?

Altered to read:

“• Perceived time pressure from surgeons in busy clinics” . Background, Line 130.

4. Row 141-3: 'There is some evidence that such targeted interventions have greater effects,16 and a growing body of evidence on de-implementation strategies (reducing low value care) has emerged in recent years.17-19 ' . --&gt; these and the following sentences do not provide a logical structure for the argumentation in this part of the manuscript. For example, why 'AND' in de above sentence?

We have re-written this entire section to allow for easier understanding. Please see response below (point 5).

5. Row 145-7: I do not understand the following sentence: "Although there is a lack of robust evidence for the methods or rationale behind identifying the substituting
behaviour, we used the barriers identified in the TDF above to develop PKIP”. It seems not logical to use barriers though there is a lack of robust evidence for methods or rationale begin identifying substituting behavior. Methods for substituting behavior cannot be translated to barriers. One of the barriers may be the lack of alternatives but they way the authors worded this is not clear.

We have changed the manuscript to separate out these two ideas as below:

“The Personalised Knee Improvement Programme (P-KIP) was designed to address these barriers. The referral process was an online referral embedded in established clinical systems that took less than one minute to complete, and was accompanied by high quality, professionally printed pamphlets and patient atlases (to facilitate the feeling that this was not something that patients had been through before). There is some evidence that such targeted interventions have greater effects.16

Additionally, P-KIP could be used as an alternative (or a substitute) for arthroscopy. A specific feeling among surgeons was a lack of alternative treatments, along with patients wanting “something” done (although not just “more physiotherapy”). Therefore, we designed P-KIP to offer a substitute pathway that surgeons and patients felt was an acceptable alternative. A growing body of evidence on de-implementation strategies (reducing low value care) has emerged in recent years,17-19 with one strategy (based on psychological models of cognition) suggesting that substitution may facilitate the de-implementation of low value care.20”Background, Lines 132 to 165.

6. Result section table 2: why do the authors only give the characteristics of the patients included in P-KIP, it would be valuable to have also the characteristics of the patients that are included before the implementation of P-KIP - Jan 2014 till July 2015 (to check for comparability of the patients before and after the implementation). This would strengthen the manuscript because the manuscript has still a major limitation (which is now discussed in the discussion section, which is good) because the authors still use te number of arthroscopies as main outcome measure, and this number can also change as result of the number of patients seen and de case mix of patients seen.

Thank you. We have included a table on the baseline characteristics (age and gender) of patients receiving an arthroscopy before and after the introduction of PKIP. Results, Lines 381-387.

We have discussed in the discussion the limitations of our approach.

7. Row 369-371: I was a bit surprised that the cost savings are described in the result section. This is not mentioned in the method section. And how many patients can be included in the P-KIP program for £90,000?

Thank you. We have included the cost and throughput of the programme in the methods section (lines 248-249), the methods for estimating savings in the methods section (lines 283-285) and the results of this analysis in the results section. Further discussion of this (along with the limitations of this approach) are in the discussion section.
8. Row 407 - Why are this early results? You did not describe this in the aim or method section

This was left over from our original draft and was a typo. We have corrected it.

9. Row 537-538 - 'This project joins the body of evidence supporting the efficacy of interventions targeted to specific barriers, specifically of substitution strategies in de-implementation of low value care, and also provides an example of how interrupted time series analysis can be used in assessing the efficacy of interventions.' Too long sentence. I advise the authors the split it up. Furthermore I do not understand the part 'the body of evidence supporting the efficacy of interventions targeted to specific barriers, specifically of substitution strategies in de-implementation of low value care..'. Providing a substitute is a way to handle a barrier (for example to give the surgeons an option to do something for the patients because they want to help the patients). But the sentence says something else. So the language must be improved.

Thank you. This has been re-written:

"This project joins the body of evidence supporting the efficacy of interventions targeted to specific barriers, and the potential value of substitution strategies where alternative options are in short supply. It also provides an example of how interrupted time series analysis can be used in assessing the efficacy of interventions." Discussion, lines 510-518

10. Row 533-536 - 'The result of this service evaluation suggests that the conservative care pathway P-KIP decreases the number of arthroscopies performed for knee osteoarthritis and leads to improvements in knee specific and general health outcomes.' the authors cannot state that patients' knee and general health outcomes improve as result of P-KIP, because they did not compare the results with a control group.

Thank you. This is a similar mistake as was made in the abstract. We have made similar changes:

"The result of this service evaluation suggests that the conservative care pathway P-KIP decreases the number of arthroscopies performed for knee osteoarthritis. Patients also experienced improvements in knee specific and general health outcomes." Results, Lines 508-511.

11. Language Editing

Thank you. We have made all the suggested changes.

REVIEWER TWO

1. The analysis was based on the number of arthroscopies instead of the percentage of patients that had arthroscopies. This could not rule out the possibility that the total numbers of patients
might be changing during the studied period, and play as a confounding factor that lead to the changes in number of arthroscopies.

Thank you. Yes, this is a limitation of the study design. However, obtaining a useful denominator is challenging (i.e. generating a rate). New patient referrals, new patient diagnosis of arthritis, total clinic appointment etc could all be used. However, each has their drawback. For example, it is not clear the number of patients that would be eligible for PKIP from new patient diagnosis (some would have mild disease and be discharged, some severe and be offered arthroplasty) – case mix as well as absolute numbers could change. Identifying such patients is also a challenge -- data on the number of new patients with a diagnosis of knee OA is not routinely captured and patients come from variety of sources.

These drawbacks were the reason we not done this. Given that the incidence of OA is increasing, we have taken the (highly pragmatic) view that the number of patients seen that are eligible are unlikely to be going down significantly over the periods of the study (1.5 year lead in, over 2 year follow up). It is possible the baseline incidence is changing or the case mix is changing (masking or exaggerating the effects). We have discussed this limitation in the discussion section.

2. Line 258. "Normality was assessed informally using histograms". Histogram is a crude way of assessing normality. There are more advanced methods to perform such analysis, for example, QQ plot, or statistical tests like Kolmogorov-Smirnov test or Shapiro-Wilk test

Yes, it is crude. However, we performed both t-test and Wilcoxon tests on the data. Both were highly significant. We presented the data on the Wilcoxon test based on the histogram.

We did not feel that performing tests such as the Kolmogorov-Smirnov test would add to this. We have highlighted this in the manuscript:

“Comparison of outcome scores was conducted with the Wilcoxon Singed Rank Test for non-parametric data. No correction for multiple tests was conducted. Paired T-tests were also conducted for non-parametric data as this test is robust against deviations from normality, and no formal tests for normality were undertaken. All analysis was performed in SPSS version 22.30” Lines 351 to 355

3. Autoregressive integrated moving average (ARIMA) is a flexible and complex modeling regime that combine the 3 features of auto-regression, integration, and moving averages, and has a set of 3 parameters to tune upon to result in vastly different models. This study did not include sufficient modeling details to see what parameters were used, how these parameters had been chosen, and what were estimations of coefficients in the final models.

Thank you. We have included a completely re-written background section and updated the results section as below, including reference to the full modeling procedure.
“Time series modelling was carried out in SPSS using the non-seasonal autoregressive expert modeller. The best fit pre and post-intervention lines were estimated using linear regression, and autocorrelation was adjusted for by using the maximum likelihood methods, with first-order autocorrelation tested for using the Ljung-box statistic. We first compared the slope change pre and post- intervention, and secondly we estimated the level change. To allow an estimate of the effect of the “bedding in” period on arthroscopy rates we performed two step change analysis: the one month effect which took account of all data from the time of intervention on; and the six-month effect, which excludes data for six months after the intervention (the “bedding in” period we expected). This was performed by extrapolating the preintervention regression line to the post intervention regression line. The difference between these points gave a point estimate for the change in level. Full details of the model building procedure are available through the Cochrane Effective practice and Organisation of Care (EPOC) resource.32” Background Lines 373 to 387.

“Interrupted time series analysis was performed on the data from January 2014 to October 2016. The model (ARIMA (1,0,0); stationary R2 =0.8; Ljung-box Q statistic= 17.2, p=0.44) demonstrated the baseline (before intervention) arthroscopy rate was stable (coefficient of slope = 0.072; p=0.688; 95% CI = -0.284 - 0.428). The slope after intervention changes significantly (coefficient = -1.231; p <0.001; 95% CI = -0.647 – 1.779), indicating the number of arthroscopies is decreasing significantly after intervention.

To further investigate this effect, the level change was determined at the one month (i.e. the effect including all data after the intervention) and six month intervals (i.e. not including the first six months after the programme to remove the “bedding in” period). The analysis revealed that the number of arthroscopies decreased by 9.2 (95% CI = 3.7-14.7; p<0.001) each month from intervention. Six-month effects (removing the “bedding in period”) revealed a decrease in arthroscopy rate of 15.4 (95% CI = 9.4-21.4; p<0.001) per month: this equates to 184 arthroscopies a year (95% CI ranges between 108 and 252). The change in estimates suggests that, as expected, there was a bedding in period during PKIP’s introduction.” Results lines 425 to 442.

4. Line 287-289 "Paired T-tests were also conducted for non-parametric data as this test is robust against deviations from normality." T-test is based on the assumption of normal distributions, and generally considered a parametric test method. There are many non-parametric tests for such circumstances, like Mann-Whitney U test and Wilcoxon signed-rank test.

We used the Wilcoxon test to analyse these data, based on histograms suggesting non-parametric distribution.

We also did a T-test (on the same data) as we used histograms to assess for normality (mentioned in point 2), and wanted to reassure the reader that we had performed the analysis thoughtfully.
5. Line 319-326. Why the estimations are inconsistent between tests based on 1-month intervals and 6-month intervals? How to interpret the difference? Also, "95% CI=5.469" and "95%=6.000" are not shown as "intervals", and might be hard to understand.

Thank you. We fully expected this difference and as such have failed to explain it in the text. We added an explanation for this and changed the CI to “intervals”.