Author’s response to reviews

Title: Facilitating Implementation of Research Evidence (FIRE): an international cluster randomised controlled trial to evaluate two models of facilitation informed by the Promoting Action in Research Implementation in Health Services (PARIHS) framework.

Authors:

Kate Seers (kate.seers@warwick.ac.uk)
Jo Rycroft-Malone (j.rycroft-malone@bangor.ac.uk)
Karen Cox (k.cox@fontys.nl)
Nicola Crichton (crichtnj@lsbu.ac.uk)
Rhiannon Tudor Edwards (r.t.edwards@bangor.ac.uk)
Ann Catrine Eldh (ann.catrine.eldh@liu.se)
Carole Estabrooks (carole.estabrooks@ualberta.ca)
Gill Harvey (gillian.harvey@adelaide.edu.au)
Claire Hawkes (c.a.hawkes@warwick.ac.uk)
Carys Jones (c.l.jones@bangor.ac.uk)
Alison Kitson (alison.kitson@flinders.edu.au)
Brendan McCormack (bmccormack@qmu.ac.uk)
Christel McCullan (c.mcmullan@bham.ac.uk)
Carole Mockford (carole.mockford@warwick.ac.uk)
Theo Niessen (t.niessen@fontys.nl)
Paul Slater (pf.slater@ulster.ac.uk)
Angie Titchen (a.titchen@gmail.com)
Teatske van der Zijpp (t.vanderzijpp@fontys.nl)
Lars Wallin (lwa@du.se)
**Author’s response to reviews:**

Reviewer 1

Reviewer Comment

a) If ACT was collected at baseline, I don't follow how it is listed as a secondary 'outcome'. I guess if only measured at baseline it should be reported much earlier in the results section.

Response: ACT was also collected at 12 and 24 months, however, this data is not available for all sites at 12 months and 24 months. In this paper it is the baseline data (the organisational context at the point of implementation) that seemed most relevant and we are really using ACT as an explanatory variable rather than as a secondary outcome variable in this paper. We would prefer to leave the results related to ACT at the end of the findings section.

Comment: b) Typically, ICCs are higher for process measures than for more distal outcomes. Process measures in primary care often have an ICC of 0.1 and outcome measures more like the stated 0.01. I appreciate the additions made in regard to ICCs - I might suggest that a revised sample size calculation could be helpful for readers.

Response: Retrospective sample size calculations are a statistically controversial issue and we do not think it will help the reader in this case. We have now included, at the suggestion of one of the other referees, the post estimation ICCs that follow from the fitting of the regression models.

Comment: c) If ACT was similar across sites, how do we understand differences observed across countries and sites? maybe ACT doesn't capture contextual factors relevant to improvement in these processes or relevant to responsiveness to the intervention?

Response: Thank you for this comment. We address this in the discussion section of our paper. The issue of whether ACT is sensitive to change we do not discuss as we have only used baseline data from ACT in this paper so do not examine change in ACT scores. The team of researchers that developed ACT are engaged in studies that are examining the sensitivity of ACT to change and this may be reported in a subsequent paper, but as stated in our response to your first point, we do not have data for all sites at 12 and 24 months.

Reviewer 3

Reviewer Comment

a) The primary outcomes are three measures of 'compliance with continence recommendations' (resident screened; assessed; treatment plan; specialist referral) (abstract and in the methods
p11). We are not clear how these are measured. Is this a percentage compliance at the level of the patient or the cluster? (nursing care site). If the latter, the regression models (tables 2 4 6 8) only include 24 units. We have assumed in the following comments that the analysis is based on patient-level outcomes. It would be helpful to expand the description of the outcomes in the methods section to explain how the measures have been constructed. Add a N= at the bottom of all the tables, to give the overall population, making clear whether these are sites or patients.

Response: Percentage compliance is at the resident level. A sentence has been added on page 9 in the outcome measures section. Supplementary file 1 provides details of the components of each of the recommendations. For each resident the percentage compliance with a recommendation is the total number of components of that recommendation that the documentation indicates have been complied with divided by the number of components and expressed as a percentage.

N=2313 has been added to the bottom of tables 1-7 to make clear that these tables present an analysis at the resident level.

Table 8 presents a summary of the staff responses to the ACT questionnaire, N=725 has been added to the title of this table so it is clear how many staff responses the table is based on overall.

Comment b) The analysis of the primary outcomes (three measures of compliance with continence recommendations) uses linear regression models. This is a cluster randomised trial with multiple levels (country, site, staff, patient). In a cluster randomised trial it is usual to perform an analysis which adjusts for clustering at the unit of randomisation, which in this case is nursing care site, and this is what is described. 'Regression' covers a range of different analyses, and in this case presumably some sort of multilevel model was used to adjust the standard errors to take site level clustering into account so it would be helpful to add in exactly which stata command was used to fit the regression models (e.g. mixed, xtmixed). It would be helpful to add a statement that the assumptions of linear regression have been examined and the data meets those assumptions,

Response: We agree with the referee that this study could be considered to have multiple levels. We have considered data at the resident level and clustering is at the site level. Country has been included as a covariate. With regard to the primary outcomes we have no information about staff so this cannot be considered as a level in the trial design.

In the previous version of the paper the models were linear regression models in which the standard errors were specified as robust (cluster) with the site as the cluster variable. This corrects the standard errors through the sandwich method (Huber-White method), inflating the se’s to correct for the clustering.
We have also fitted a multi-level mixed effect linear regression model (using the STATA15 mixed command, again with site as the level variable and SE set to robust(cluster)). The estimates are very similar to those from the simpler model we reported in the previous version, some of the se’s are increased a little and none of the conclusions from the models are changed. We have decided to update the results tables 2, 4, 6 to show the results from the multi-level mixed effect linear regression to be sure we have taken full account of the clustering. We have amended the explanation of the model fitting on page 11, to clarify what was done.

We have added a statement about the assumptions on page 13-14.

Comment

c) ICCs were calculated on the baseline data and we find this confusing. Once a multilevel model has been fitted to the data, it is usually possible to extract an overall ICC at the level of clustering which would be based on all data points (not just baseline) - see https://www.stata.com/features/overview/intraclass-correlations-for-multilevel-models/

Response:

The reason for calculating the ICCs at baseline is these are the values which, had they been available, we would have used in the sample size calculation.

We have, as you suggested, calculated the post-estimation ICC for each of the regression models. These post estimation ICCs are reported at the end of tables 2, 4 and 6 and in the text relating to the results for each of the three compliance variables (pages 14-16).

This analysis was not available in version 10 of STATA (the version we had been using for the analysis). We have therefore rerun all the analyses in STATA version 15 and throughout the text updated STATA10 to STATA15.

Comment: d) There were only 24 sites (clusters) and both 'country' and 'intervention' are fitted at a site level and have a total of 5 individual covariates between them - it is doubtful there is enough data here to provide stable estimates: We'd recommend that the authors try removing 'country' from the analysis as a sensitivity analysis to check that results are similar.

Response: We agree with the reviewers that the number of sites is small relative to the number of levels of the covariates. As the reviewers suggest have re-run the analysis removing country from the analysis as a sensitivity check. There is still no significant intervention effect for any of the three recommendations if country is removed. The results are similar, so the estimates appear to be reasonably stable. This has been noted in on page 13-14.

Comment e) Patients were recruited to this trial post randomisation - which means that recruiters already knew which arm of the trial a site was in before consenting patients. This lack
of allocation concealment can lead to differential recruitment in a cluster randomised trial - e.g. differences in numbers recruited or the type of patients recruited. We can see little information on which to judge whether or not this was a problem. We can't find a consort flow diagram and this would be helpful (and recommended by CONSORT) - how many patients were approached but did not consent, and how many consented but did not provide outcome data? Baseline compliance in the control group was much higher than in the intervention group (table 3) and this could be an indication of differential recruitment. Perhaps the authors can make some comment on this.

Response: On page 10 in the section Sample size and power calculation it says “Consent was sought at cluster and at individual level, the former before randomisation and the later after randomisation.” On page 11 in the section on allocation concealment and blinding it said “It was not possible to blind sites to intervention, although research fellows who collected data were initially blinded to intervention group.”

This sentence has been reworded (on page 11) to make it clear that where it was necessary to obtain consent from individual residents for outcome data collection this consent was obtained by the research fellow who was initially blind to the intervention allocated for the site.

To clarify for the reviewers, consent from residents was not necessary for access to records in Netherlands, Sweden or Ireland. All data from records was collected by the country research fellow who was blinded to the intervention group to which the care home had been allocated, so recruitment of resident records would not have been influenced by lack of allocation concealment. In UK, consent from residents (or their family) was necessary for access to the resident record. This consent was obtained by the research fellow who was unaware of which intervention the care home had been allocated. We noted that once the research fellows were in the long term care setting, the blinding could inadvertently be broken by the site, for example, mentioning the name of an external facilitator working with them, and thus revealing the allocation of that site. In all countries the consent of the resident or their family was necessary from the collection of EQ5D, this consent was obtained by the research fellow.

The study flow diagram is in supplementary file 4.

Comment f) How similar were patient demographics across groups? The paper would benefit from a table of baseline characteristics, by group. Provide mean(SD) for continuous variables, number(%) for categorical and median (IQR) for ordinal or non-normal, and include both the number included for each measure plus the overall total. It is not good practice to compare the groups using statistical tests (as described on page 11 and shown in table 1).

Response: The only resident demographic information that was collected was age and gender. We prefer not to introduce a further table to report this information by group, but we have included it within the text on page 12 in the description of resident sample.
Whilst we agree with the reviewers that using statistical tests to compare groups is not considered by everyone to be good practice, however it is a widely used practice. We only do this for demographic data and some secondary outcomes because we think some readers would expect (or prefer) to have evidence of statistical tests alongside statements of similarity or difference. With regard to Table 1 we have included 95% CI’s for those who prefer this approach to describing similarity or difference between groups.

Comment: g) EQ5D-VAS is reported (table 1) at 24 months only (because of least amount of missingness at that time point). It was collected at several time points, suggesting it might be an outcome measure, but it is not reported as such.

Response: The reviewers observation that EQ5D might have been a secondary outcome measure is correct. However, there were difficulties with the process of getting resident consent for collection of this data and process of getting this measure completed with the resident. Over time the issues related to consent and process were resolved sufficiently so that at the 24 month data collection point this measure was available for 77% of residents. Given the difficulties with collecting this data earlier in the trial we decide this measure could not be used as a secondary outcome. We did however, feel that the more complete 24 month data was helpful in providing a description of the health status of the residents.

Comment: h) Table 3 and 5 both suggest an effect in favour of the 2 intervention arms, and the effect is quite close to the 15 percentage point difference anticipated in the power calculation (Table 7 does not). This, combined with the very wide confidence intervals, might indicate that there is a hint of an effect here, but that there is insufficient power to detect a difference. We like the way the authors have used the findings from the process evaluation to explain the results: however, they cannot write off the idea that there may have been an effect, and the lack of power may be more of an issue than suggested in the conclusion.

Response: We agree with the reviewers’ observation that given the lack of power we cannot write off the idea that there may have been an effect from one or both of the intervention arms, and this is addressed in the limitations. We feel that the text on page 14 already identifies that Tables 3 and 5 show some improvement in the intervention arms, but given the large variability associated with the means reported and the lack of power we are reluctant to put any further emphasis on these results. As the reviewers note the confidence intervals are very wide.

Comment: i) Secondary outcomes at patient level seem to have been reported as a difference between baseline and 24 months, rather than compared by group. These outcomes should be reported as described in the methods, using Anova or chi squared tests, as appropriate, or even using regression methods

Response: All the secondary outcomes at patient level are binary outcomes we have therefore used chi-squared test to compare the groups at 24 months and this information has been added on page 16-17. As further exploration of the differences found between baseline and 24 months we
feel it is helpful to look at changes within the intervention groups, so the changes within intervention group previously reported remain in the text.

Comment: j) Given the lack of power in the study, we would very tentatively suggest that the authors give consideration to combining both arms and comparing facilitation against control, although this analysis was not anticipated and would have to be reported very carefully.

Response: Although the study was not designed with the intention of combining facilitation arms, the analysis has been rerun comparing facilitation (combining the groups Type A and type B) against control and there is no evidence that the facilitation is more effective than control.