Reviewer's report

Title: Assessing fidelity of delivery of smoking cessation behavioural support in practice.

Version: 1 Date: 5 December 2012

Reviewer: Felix Naughton

Reviewer's report:

The authors present a highly useful and implementable method of measuring fidelity in smoking cessation support/counselling sessions and I applaud them for describing a clear and methodical process and for sharing high levels of detail in the accompanying tables.

Major compulsory revisions

None

Minor essential revisions

One aspect I struggled with is whether adherence to a service treatment manual is an ideal measure without knowing any history of how the manual was developed, though I realise the aim was not to look at clinical benefit. The authors indicate that these manuals “in theory incorporate national guidance and training standards” but there were no reported efforts to assess how well these particular treatment manuals met these standards. The point is this - if you have less BCTs specified in a manual then presumably, although this is dependent upon which BCTs were included, it would be easier to come out as having higher fidelity. For example, service 2 was reported as having lower treatment fidelity for the post-quit sessions, yet they delivered more BCTs as they specified more in their manual for this session, so while these sessions were less adherent to the manual they might arguably be more effective? I felt this needed to be discussed.

Discussion – presumably some BCTs take longer to deliver than others (e.g. set graded tasks vs provide reassurance), so given that time taken to deliver each BCT was not incorporated in the analyses, it is not entirely valid to conclude that as session duration was not associated with extent of fidelity insufficient time is unlikely to be a factor for failures to deliver prescribed content.

There was no acknowledgement in the discussion that recording advisors for a few sessions may have changed their behaviour in some way, so these sessions may reflect the best case scenario? Furthermore, as opportunistic sampling was used, this may have allowed for self-selection of which clients to record (could this explain the different numbers of recording between the two services?), which may have again affected the results. This should be discussed.

Discretionary revisions

In the abstract it is stated that there are no published studies looking at
assessing the fidelity of intervention implementation for smoking cessation and the current study has as aim 1 to evaluate a method of assessing fidelity using the BCT taxonomy. However, in the introduction, a study is cited which has used the BCT taxonomy to describe support sessions provided by stop smoking services using audio-recording (Lorencatto et al, 2011; Psych & Health) – so from a readers perspective it sounds like there is at least one study published looking at the fidelity of interventions delivered in the services? Perhaps there is a subtle difference between these studies that I have missed, but this stands as a bit of a contradiction from a reader’s perspective.

The last third of the first paragraph under procedure was a little difficult to follow, perhaps this could be made a bit clearer for the reader.

The way you numerically structured the results section really helps the reader, but it would be good if they directly corresponded to the objective numbers – I see that they do in that section 2 in the results corresponds with objective 2i, and 3 with 2ii etc. but at first I was a little confused.

In the results, the inter-rater reliability coding was 87.1% which seems satisfactory, although it is classed as ‘high’ (>75%) with a reference to a paper by Cohen (1968) on Kappa. Does this >75% figure correspond to percentage agreements or kappa agreements? The reference would suggest it is Kappa agreements, although I have not read the paper and am not an expert on inter-rater reliability, but I thought I would flag it up in case the citation was incorrect.

It would be valuable to have an indication of what the overlap between these two different services were in terms of BCT inclusion in their manuals. Table 2 is very helpful, although it doesn’t tell us anything about how the treatment manuals varied.

The supplementary document I had access to did not have a table 1a or 1b (which were referred to in the text), so I was unable to see what the differences were between the services, although the additional information in tables 1 and 2 in the supplementary document are very useful more generally.

Why was there such a difference between services in terms of the number of recordings? Might this have influenced the findings? How many participants did these recordings represent – were the same individuals followed up or were they always different clients (may explain why some BCTs were not used at subsequent visits if, in some cases, they had already been delivered).

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:

I declare that I have no competing interests