Author's response to reviews

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

Authors:

    Fabiana Lorencatto (fabi.lorencatto@ncsct.co.uk)
    Robert West (robertwest100@gmail.com)
    Charlotte Christopherson (c.christopherson@ucl.ac.uk)
    Susan Michie (s.michie@ucl.ac.uk)

Version: 2 Date: 21 February 2013

Author's response to reviews: see over
To the editor,

On behalf of my co-authors, I would like to thank you and the reviewers for the helpful suggestions and comments provided on our manuscript for the above study. Please find below a list of all reviewers’ comments in bold. In italics, we have listed our responses and underlined any changes made to the manuscript (which are highlighted in yellow on the main manuscript).

I. COMMENTS FROM REVIEWER 1:

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:

1) 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.

Response from authors: We agree that it is important to consider what practitioners are being asked to adhere to, by considering the evidence-base of manuals and the extent to which they are fit for purpose. Although this was not part of our study, we have added the following content to the discussion:

‘This study raises the issue of the extent to which treatment manuals are fit for purpose. The evidence base for the BCTs in the services’ manuals was not assessed, nor was the extent to which manuals are clearly written and conform to training standards and national guidelines.'
This is not only necessary for interpreting results of fidelity assessments but also for comparing the quality of services provided, since both the planned content and the extent to which content is delivered are essential aspects of assessing the quality and hence likely impact of a service. For instance, the post-quit sessions delivered in Service 2 had an average lower percentage fidelity (56%) than those delivered in Service 1 (69%). However, the post-quit manual from Service 2 contained more BCTs (17) than that from Service 1 (10). The mean number of BCTs delivered per post-quit session in Service 2 was higher than that in Service 1 (~10 vs ~7 BCTs respectively). Therefore although fidelity appears to be poorer in Service 2, the post-quit sessions may in fact have been more effective in helping clients successfully quit as a higher number of techniques were delivered. This raises the question as to whether 100% fidelity is necessary to produce desired treatment outcomes[16]. Combining an analysis of the extent to which manuals are based on good evidence with an assessment of fidelity will give a more comprehensive assessment of delivery and stronger evidence of intervention quality than considering either evidence or fidelity on its own.

2) 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.

Response from authors: We agree that time taken to deliver each BCT is an important factor to consider when interpreting the association between session duration and extent of fidelity. We have amended the relevant paragraph in the discussion (p.15) to reflect this, so that it now reads:

‘Session duration was not significantly associated with extent of fidelity. Insufficient time to deliver manual-specified content is therefore unlikely to be an important contributing factor for failures to deliver prescribed content in this area. However, time taken to deliver each BCT was not accounted for in analyses. It is possible that some complex BCTs, such as ‘barrier identification and problem solving,’ take longer to deliver than BCTs such as ‘provide reassurance.’ Such variation across BCTs may have in part influenced the relationship between overall observed fidelity and session duration.’

3) 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.
Response from authors: We have added the following sentences to the ‘limitations’ paragraph in the discussion to acknowledge the potential influence of demand characteristics on practitioner’s delivery:

‘Nonetheless, practitioners were aware that their sessions were being audio-recorded and may thus have been susceptible to demand characteristics and attempted to improve their practice under observation. Therefore, these sessions may not be representative of typical practice. However, these sessions are likely to represent a ‘best case scenario,’ and therefore over-estimate rather than under-estimate fidelity of delivery.’

Concerning the potential for practitioners to self-select clients to record, this is unlikely given the sampling procedures which have been clarified in response to Reviewer 1’s 10th comment. Nonetheless, this has been acknowledged in the following sentences that have been added to the procedure section of the Methods:

‘Audio-recordings of consecutive behavioural support sessions delivered to consenting clients as part of routine clinical practice were obtained during a two month data collection period. This minimised the opportunity for practitioners to select which clients to record. The resulting sample comprised 30 recordings from Service 1 and 13 recordings from Service 2.’

Discretionary revisions

4) 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.

Response from authors: The study by Lorencatto et al. (now In Press with Journal of Consulting & Clinical Psychology; reference updated accordingly throughout the manuscript) aimed to establish whether the taxonomy was a reliable method for identifying BCTs delivered during behavioural support sessions. It did not investigate treatment fidelity, the subject of the current study. This difference between the two studies has been clarified in the 5th paragraph of the introduction in the following sentences:

‘This taxonomy has been reliably applied in a previous study as a coding framework for identifying and categorising component BCTs present in English Stop-Smoking Service treatment manuals [4, 6,25] and transcripts of audio-recorded behavioural support sessions delivered by these services [26]. However, the taxonomy has not yet been used to compare the
content of treatment manuals with the transcripts of corresponding behavioural support sessions to assess fidelity.

This study aimed to evaluate the taxonomy as a method for investigating variations in the fidelity of delivery.

5) 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.

Response from authors: We have rephrased the relevant section of the first paragraph under procedure (p.9) for clarification purposes, so that it now reads:

‘This adapted taxonomy is an updated version of the original taxonomy of 43 BCTs. Adaptations included merging typically co-occurring BCTs and refining existing BCT labels and definitions [26]. The resulting content of the taxonomies is therefore largely comparable and comprise the same BCTs.’

6) 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.

Response from authors: We have restructured the results (p.11-13) and rephrased the headings for each sub-section within the results so that they clearly correspond with the research objectives outlined in the last paragraph of the introduction.

7) 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.

Response from authors: We have now removed the Cohen (1968) reference from the results on p.11 and reference list as it was incorrect.

8) 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.

Response from authors: We have created a new Supplementary Table (Supplementary Table 2) which lists, side-by-side the BCTs present in each section (i.e. pre-quit, quit day, post-quit) of the manual from each service. We have emphasized in bold the BCTs present in both of the manuals. This table is referred to in the sub-section of the results ‘variation in fidelity by session type,’ in the following sentence:

‘A full list of BCTs identified within each manual is available in Supplementary Table 2.’
9) 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.

Response from authors: We have corrected this error in the text, removing references to supplementary table 1a or 1b. The relevant supplementary table, which presents results across both services, has now been re-numbered as supplementary table 3 as a result of the insertion of additional supplementary tables 1 and 2 in response to Reviewer 1’s 8th comment and Reviewer 2’s 10th comment. This has been corrected in the relevant text of the results section so that it now reads:

‘The proportion of sessions individual manual-specified BCTs were delivered in with fidelity according to session type across both services is available in Supplementary Table 3.’

10) 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).

Response from authors: Audio-recordings of sessions were obtained for a fixed time period (i.e. two months). Practitioners were able to audio-record sessions delivered to consenting clients only. There were differences within each service in the number of clients who provided consent during the data collection time period; this resulted in the uneven sample size. To clarify this in the manuscript, we have added the following sentences:

‘Audio-recordings of consecutive behavioural support sessions delivered to consenting clients as part of routine clinical practice were obtained during a two month data collection period. This minimised the opportunity for practitioners to select which clients to record. The resulting sample comprised 30 recordings from Service 1 and 13 recordings from Service 2.’

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.

II. COMMENTS FROM REVIEWER 2:

The authors of this manuscript have attempted to tackle an important issue that is often ignored and is, most likely, an important variable in treatment outcome. Additionally, they have tackled it in the smoking population who are well known to have many health sequellae. The following are points that I believe should be addressed or clarified before this manuscript is accepted:
1. The data is original and the question posed is important. I recommend the authors reword objective 2 to increase clarity. I also recommend that adding an additional reviewer would strengthen the findings.

Response from reviewers: To increase clarity, we have split objective 2 into two smaller objectives and rephrased these. The list of objectives on p. 7 in the introduction now reads:

1. ‘To evaluate a method of assessing fidelity of behavioural support for smoking cessation using a taxonomy of behaviour change techniques.

2. To assess using this method the fidelity of delivery of behavioural support in two English Stop-Smoking Services

3. To examine variation in fidelity according to: i) session type (i.e. pre-quit, quit-day, post-quit), ii) session duration, iii) stop smoking practitioner, and iv) the specific BCT.

4. To assess the extent of use of BCTs not included in the particular treatment manual in operation.

2. I am not sure the interpretation of findings is warranted. One of the major weaknesses of the manuscript is the lack of information as to what warrants acceptable fidelity for this treatment. They identify 43 separate behaviors that are part of the treatment taxonomy and both treatment sites had an average of 2/3 completion. For many treatments this may be good fidelity. A justification for their conclusion needs to be strengthened. Are all sessions required to have 100% of the BCTs?

Response from authors: We agree that it is important to consider whether 100% fidelity is required. We have in part addressed this in our response to Reviewer 1’s first comment [see above]. We have added the following paragraphs to the discussion to address these comments:

‘This study raises the issue of the extent to which treatment manuals are fit for purpose. The evidence base for the BCTs in the services’ manuals was not assessed, nor was the extent to which manuals are clearly written and conform to training standards and national guidelines. This is not only necessary for interpreting results of fidelity assessments but also for comparing the quality of services provided, since both the planned content and the extent to which content is
delivered are essential aspects of assessing the quality and hence likely impact of a service. For instance, the post-quit sessions delivered in Service 2 had an average lower percentage fidelity (56%) than those delivered in Service 1 (69%). However, the post-quit manual from Service 2 contained more BCTs (17) than that from Service 1 (10). The mean number of BCTs delivered per post-quit session in Service 2 was higher than that in Service 1 (~10 vs ~7 BCTs respectively). Therefore although fidelity appears to be poorer in Service 2, the post-quit sessions may in fact have potentially been more effective in helping clients successfully quit as a higher number of techniques were delivered. This raises the question as to whether 100% fidelity is necessary to produce desired treatment outcomes[16]. Combining an analysis of the extent to which manuals are based on good evidence with an assessment of fidelity will give a more comprehensive assessment of delivery and stronger evidence of intervention quality than considering either evidence or fidelity on its own.

The question as to whether 100% fidelity of intervention delivery is a desirable aim is under debate [16]. Strict adherence to treatment manuals may be detrimental to therapeutic interactions, as not all content specified in manuals will be relevant to all the individual needs and concerns of intervention recipients [35, 36]. The delivery of additional, non-manual specified BCTs may be one means by which practitioners are tailoring the content of support provided to client needs and increasing flexibility in their practice. Furthermore, the manuals from both services contained a high number of BCTs, which may not always be feasible or appropriate to deliver in practice. However, manuals are essential to maintaining a degree of consistency and standards in service provision. Some argue in favour of a middle ground in which core, prescribed intervention components are delivered with a degree of additional flexibility and tailoring in how content is provided. Such an approach does not compromise fundamental treatment integrity, and offers a potentially more feasible, realistic and beneficial model of treatment delivery [16, 36].

3. Methods are relatively strong but there are some amendments that may further strengthen the report:

a) Why 2 reviewers? More info on who they were, how they were related to the project? Also clarify did the same reviewers who identified the taxonomy from the manuals also review the transcripts? This is unclear.

*Response from authors: We have added more information regarding the two reviewers (i.e. coders) to the procedure sub-section of the methods. It now reads as follows:

‘Coding was conducted by two research psychologists (researcher initials: FL, CC) with previous training and experience in coding using the taxonomy. Both of the researchers independently coded all study materials (i.e. both manuals, 34 transcripts).’*

b) A better rationale for use of % agreement is required as Kappa tends to be the standard. Would avoid the term inter-rater reliability and use percent agreement, since reliability typically implies a more stringent test.
Response from authors: There are several practical and theoretical reasons why calculating Kappa is not suitable for analyzing the reliability of coding these transcripts. First, the items being coded are not mutually exclusive, as more than one BCT may be present within a single excerpt/sentence. Secondly, BCTs are typically present more than once within a transcript; with coders potentially agreeing that a BCT is present in one instance within a transcript but not another. It is therefore not possible to conduct a simple 2x2 coding classification exercise and calculate kappa based on whether coders agree/disagree as to whether a BCT is globally ‘present/absent’ within a transcript. Results generated from such an analysis would not be representative. Instead, we have chosen to examine the proportion of all BCTs identified within a transcript that were identified with agreement by both coders (i.e. % positive agreement), as this overcomes the issue that one excerpt can contain multiple categories, and BCTs can feature multiple times within a transcript. We have clarified this in the analysis section of the manuscript as follows:

‘Inter-rater coding reliability was assessed by examining the proportion of all BCTs identified within a transcript that were identified by both coders (i.e. % positive agreement). Percentage agreement was used rather than Cohen’s Kappa for numerous reasons. First, the items being coded (i.e. sentences within transcripts) were not mutually exclusive, as multiple BCTs may be present within a single sentence. Secondly, BCTs may occur multiple times within a single transcript, with coders potentially agreeing in one instance within the transcript that the BCT is present, but not in another. This does not allow a global present/absent rating for the entire transcript for each BCT. Furthermore, given the high number of 43 BCTs, the probability of selecting a particular code by chance is low. Since Kappa corrects for chance agreement amongst multiple coders, use of Kappa is likely to underestimate reliability [29].’

c)- not sure if non-BCTs is as interesting at this point, given the need for some of these process oriented actions to make a relationship with the client.

Response from authors: In this study we are not coding ‘non-BCTs,’ but rather, BCTs that are delivered but are not included in the service treatment manual. The importance of this is that there may be effective BCTs being used in practice, that for some reason, are not included in the manual. The delivery of these may be beneficial to treatment outcomes, and has implications for improvement of manuals. This has been acknowledged in the discussion in the following paragraph:

‘In the present study, of the additionally delivered BCTs, some featured consistently in all sessions despite not being prescribed in either service’s manual (i.e. provide feedback on performance), and others such as ‘boost motivation and self-efficacy’ have been shown to be effective [25]. It is possible that practitioners recognise the value of these BCTs, or that they are easier to deliver routinely or intuitively. If research evidence suggests such BCTs contribute to treatment success, they should be considered for inclusion in treatment manuals.’

d) more info on how sample was obtained is needed and over what period of time
Reviewer’s response: Our amendments to the manuscript in order to clarify sampling procedures have already been outlined in response to comment number 10 by Reviewer 1. These changes are visible in the methods section in the following sentences:

‘Audio-recordings of consecutive behavioural support sessions delivered to consenting clients as part of routine clinical practice were obtained during a two month data collection period. This minimised the opportunity for practitioners to select which clients to record. The resulting sample comprised 30 recordings from Service 1 and 13 recordings from Service 2.’

c) unclear why there cannot be some indication as to whether fidelity impacts outcome? Even if it is not the most powered analyses, I am not sure why there can’t be some indication of the relationship in this sample

Response from authors: In order to examine associations between fidelity and quit outcomes, we would need to have a complete set of recordings of all four behavioural support sessions (pre-quit, quit-day, post-quit sessions) received by each smoker, as each session will contribute in part to the resulting quit outcome. As recordings were collected opportunistically within a limited time-frame, it was not possible to collect the complete set of sessions for all smokers. We also do not have data on the subsequent quit outcome for these individual smokers, or information on their smoker and socio-demographic characteristics, which would need to be controlled for in analyses. It would therefore not be feasible to examine associations between fidelity and quit outcomes based on this data set. We agree that the impact of fidelity is a key question and this is part of our future programme of work.

4. The following are issues of writing/organization that I believe may strengthen this manuscript:

a) Since many non-British persons may read this study, it may be essential to provide greater description on this program, who conducts the sessions, how long are they, etc. Also, please define the three categories pre-quit, etc as they are confusing

Response from authors: Each of the 152 NHS Stop Smoking Services vary in terms of the practitioners they employ, and the format, content and duration of typical sessions provided in their service. We have provided a general overview of the typical support provided by these services in the ‘sample’ section of the methods by including the following sentences:

‘… typically offer medication and four weekly behavioural support sessions. Behavioural support is typically provided by trained, specialist advisors, often of multidisciplinary backgrounds (i.e. nurses, midwives, GPs, pharmacists). The first session is typically a ‘pre-quit session’, which aims to enhance a smoker’s motivation and self-confidence to quit, set clear goals, discuss medication options, and address general preparations for quitting. The second session is the ‘quit-day’ session, which focuses on general strategies for avoiding smoking cues and overcoming barriers to cessation, as well as maintaining motivation and self-efficacy. The final two sessions are post-quit sessions, which concentrate on equipping the client with strategies for avoiding smoking in the long term by facilitating relapse prevention and coping, alongside promoting an ex-smoker identity.’
b) briefly describe the evidence base for this program, how do we know these 43 behaviors produces results?

Response from authors: In the introduction, we have now cited two studies that have assessed the evidence base for BCTs included within the taxonomy:

‘Fourteen BCTs from the taxonomy have been supported by RCT evidence, and 16 have been shown to be significantly associated with improved four-week CO-validated quit outcomes [41, 25].’

5. Use table to provide descriptive details for the persons providing and receiving the intervention including relevant details about their discipline, years education, etc. and for the smokers, relevant details about years smoking, etc.

Response from authors: These data were not collected in order to minimize practitioner burden and interference with actual practice by introducing additional questionnaires required to be completed with smokers during sessions.

6. P5 is confusing. I was not aware that fidelity did not appraise quality of intervention

Response from authors: According to conceptual models and definitions of treatment fidelity (see Borrelli, 2011), the dimension of fidelity of delivery specifically concerns whether or not intervention components are actually delivered as pre-specified or intended. Quality of delivery is a closely associated concept, but is separate in that it concerns how well components are delivered; components in a manual can be present/delivered in practice, and in that sense fulfill fidelity of delivery, but not necessarily be delivered well, with high quality. This is outlined in the 4th paragraph of the introduction (bottom of page 5). We have rephrased this for clarification purposes, so that it now reads as:

‘Fidelity of intervention delivery refers to the extent to which interventions are delivered as intended, with adherence to specifications in intervention manuals [16,17]. It specifically concerns whether core, prescribed intervention components are delivered, rather than the associated but separate question of how components are delivered, for example, in terms of quality or tailoring of delivery.’

7. Again, it has to be clarified as to what the gold standard is for this particular intervention...is 100% required to produce results?

Response from authors: We have now addressed this issue in the discussion in response to comment 1 by reviewer 1 and comments 2 and 4b by reviewer 2 [see above].

8. info on training on the manual is necessary. Are those that provide the intervention put through some sort of standardized training? Also, as manuals are written in house..how do we know they are written in a clear, manner?
Response from authors: We do not have data on the training that practitioners received regarding manual use. Regarding whether manuals are written clearly, in the discussion we have acknowledged the need to examine this when assessing whether manuals are fit for purpose:

‘This study raises the issue of the extent to which treatment manuals are fit for purpose. The evidence base for the BCTs in the services’ manuals was not assessed, nor was the extent to which manuals are clearly written and conform to training standards and national guidelines. This is not only necessary for interpreting results of fidelity assessments but also for comparing the quality of services provided, since both the planned content and the extent to which content is delivered are essential aspects of assessing the quality and hence likely impact of a service.’

9. Rather than using the term average, be specific, is it the mean?

Response from authors: We have changed the term average to mean throughout the relevant content of the methods and results section.

10. Table 1 is not interesting and appears to be a data table rather than a synthesis of results. Table 2 (which is supplementary) is much more interesting and informative...why not use that?

Response from authors: We have now created a new Table 1 which summarizes the results from the existing Table 1. The revised Table 1 now presents the mean session characteristics and number (percentage) of manual-specified BCTs delivered according to session type, by service. The existing Table 1 that lists results by individual transcript has now been made a supplementary table (Supplementary Table 1). The results have been rephrased accordingly throughout. In particular, results regarding fidelity by session type have been re-written as follows so that there is not substantial overlap between the content of the revised Table 1 and results reported in text:

‘3. Variation in fidelity of delivery

(i) According to session type

The number of BCTs identified in the pre-quit, quit day and post-quit section of each service’s treatment manual is provided in Table 1. A full list of BCTs identified within each section of the manual is available in Supplementary Table 2. The mean number (%) of manual-specified BCTs delivered in each session (i.e. % fidelity) is presented according to session type, by service, in Table 1. This, alongside general session characteristics, is available for each of the 34 individual transcripts in Supplementary Table 1.

Across both sets of transcripts, the mean proportion of manual-specified BCTs delivered per session was 66% (SD 14; range: 38-83%) for pre-quit sessions, 72% (SD 15.01; range: 50-85%) for quit-day sessions, and 62% (SD 16.4; range: 35-90%) for post-quit sessions (Table 1; Supplementary Table 1).

In Service 1, fidelity was on average highest for post-quit sessions, with a mean of 69% of manual-specified BCTs delivered per post-quit session, and lowest for pre-quit sessions (mean
58%) (Table 1). In Service 2, fidelity was on average highest in quit-day sessions (mean 81%) and lowest in post-quit sessions (56%) (Table 1).

11. A general discussion of fidelity may not be as useful compared to the issue of the fidelity cut off for this treatment. I am not sure you can use these general guidelines for this particular treatment. Perhaps 2/3 fidelity is enough to produce a change?

Response from authors: We have now addressed this issue in the discussion in response to comment 1 by reviewer 1 and comments 2 and 4b by reviewer 2 [see above].

12. Further discussion on the limits to generalizability may be useful and a list of lessons learned" from this study for the reader.

Response from authors: The limitations in generalizing and extrapolating these findings beyond the current sample have already been outlined on p.17 of the discussion, in the following sentences:

‘Limitations of the current study firstly include the sample size of only two services, which means that these findings may not reflect all sessions delivered by practitioners, other services or behavioural support provided in contexts other than the English Stop-Smoking Services.’

13. You hint at issue of treatment individualization. Please expand this discussion in relation to the issue of fidelity and non-BCT

Response from authors: In the 7th paragraph of the discussion, we have included the following sentences to reflect the issue of additional, non-manual-specified BCTs contributing to treatment individualization:

‘The delivery of additional, non-manual specified BCTs may be one means by which practitioners are tailoring the content of support provided to client needs and increasing flexibility in their practice.’

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Needs some language corrections before being published
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: I have no competing interests

REVIEWER 3

Reviewer's report:

This manuscript offers a strong contribution to the literature on fidelity
assessment, providing an informative initial step in evaluating the fidelity of smoking-cessation programs. As the authors note, subsequent application of this approach will be necessary to derive generalizable inferences about the state of the field more broadly, but even a convenience sample of n=2 is sufficient to indicate that such a step will be quite worthwhile.

In discussion, the authors appropriately note important issues not always carefully considered. The provision of techniques that are not specified in a manual might typically be considered as noise; but these could range from detrimental, to neutral though representing an opportunity cost in provider effort, to beneficial though of yet unknown benefit (my phrasing, not necessarily theirs). The authors rightly suggest that these need to be measured and accounted for in research analyses. They also suggest that more careful attention should be paid to the relevance and benefit of specific intervention components to individuals of varying characteristics, allowing more nuanced specification of model components.

Discretionary revisions

One change in the manuscript would help many readers. The initial section of the results provides quantitative results in narrative form, providing summations of data referred to in Table 1 by Service and phase. An alternative would be to provide some of this information in tabular form, likely following rather than built into Table 1. This section could be more concise, and readers could get a better view of cross-service and cross-phase differences. This recommendation is offered as a “discretionary revision” rather than “major compulsory revision” because it is not essentially substantive, but this reader does think it would strengthen the manuscript.

Response from authors: We have now created a new summarized Table 1 in response to comment 10 by Reviewer 2. The content of the in-text results describing variation in fidelity according to session type has also been revised accordingly (see response to Reviewer 2, comment 10).

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.