Underreporting of the active content of behavioural interventions: a systematic review and meta-analysis of randomised trials of smoking cessation interventions

Marijn de Bruin, Nicola Black, Neza Javornik, Wolfgang Viechtbauer, Maarten C. Eisma, Jamie Hartman-Boyce, A. Jess Williams, Robert West, Susan Michie & Marie Johnston

Routledge
Taylor & Francis Group

Underreporting of the active content of behavioural interventions: a systematic review and meta-analysis of randomised trials of smoking cessation interventions

Marijn de Bruin [a,b], Nicola Black[a,c], Neza Javornik[a], Wolfgang Viechtbauer [d], Maarten C. Eisma [e], Jamie Hartman-Boyce [f], A. Jess Williams[g], Robert West [h], Susan Michie [i] and Marie Johnston [a]

[a]Health Psychology Group, Institute of Applied Health SciencesUniversity of Aberdeen, Aberdeen, UK; [b]Radboud University Medical Center, Radboud Institute for Health Sciences, IQ Healthcare, Nijmegen, the Netherlands; [c]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; [d]Department of Psychiatry and NeuropsychologySchool for Mental Health and Neuroscience, Maastricht University, Maastricht, the Netherlands; [e]Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, the Netherlands; [f]Nuffield Department of Primary Care Health Sciences and National Institute of Health Research Oxford Biomedical Research Centre, University of Oxford, Oxford, UK; [g]Institute of Mental HealthSchool of Psychology, University of Birmingham, Birmingham, UK; [h]Department of Behavioural Science and Health, University College London, London, UK; [i]Centre for Behaviour Change, University College London, London, UK

ABSTRACT

Despite its importance, underreporting of the active content of experimental and comparator interventions in published literature has not been previously examined for behavioural trials. We assessed completeness and variability in reporting in 142 randomised controlled trials of behavioural interventions for smoking cessation published between 1/1996 and 11/2015. Two coders reliably identified the potential active components of experimental and comparator interventions (activities targeting behaviours key to smoking cessation and qualifying as behaviour change techniques, BCTs) in published, and in unpublished materials obtained from study authors directly. Unpublished materials were obtained for 129/204 (63%) experimental and 93/142 (65%) comparator groups. For those, only 35% (1200/3403) of experimental and 26% (491/1891) of comparator BCTs could be identified in published materials. Reporting quality (#published BCTs/ #total BCTs) varied considerably between trials and between groups within trials. Experimental (vs. comparator) interventions were better reported ($B(SE) = 0.34$ (0.11), $p < .001$). Unpublished materials were more often obtained for recent studies ($B(SE) = 0.093$ (0.03), $p = .003$) published in behavioural (vs. medical) journals ($B(SE) = 1.03$ (0.41), $p = .012$). This high variability in underreporting of active content compromises reader's ability to interpret the effects of individual trials, compare and explain intervention effects in evidence syntheses, and estimate the additional benefit of an experimental intervention in other settings.

Background

Identifying effective intervention strategies for modifying health behaviours and behaviours of healthcare professionals is critical to improving population health and reducing healthcare

expenditure (Jepson, Harris, Platt, & Tannahill, 2010; NICE, 2007; WHO, 2013). Accurate and comprehensive reporting of such interventions allows readers to understand exactly what has been developed and evaluated, what resources are required to deliver it, and what is needed for replication of the intervention and its implementation in other settings. Full intervention descriptions are also key to evidence syntheses and economic modelling of interventions to inform policy and practice.

Behavioural interventions are often experimentally evaluated against active comparator groups to investigate effectiveness; for example, behavioural support for smoking cessation, treatment-as-usual, or the same intervention delivered through different means (e.g., group vs. individual). These comparators may – like experimental interventions – include a range of active intervention ingredients, which is thought to influence success rates in the comparator arm and therefore trial effect sizes (de Bruin et al., 2010; de Bruin, Viechtbauer, Hospers, Schaalma, & Kok, 2009; Wagner & Kanouse, 2003). Hence, current guidance for reporting of trials and their interventions recommends that both experimental and comparator interventions are reported in similar detail (e.g., Albrecht, Archibald, Arseneau, & Scott, 2013; Hoffmann, Glasziou, et al., 2014; Montgomery et al., 2018).

Previous studies examining behavioural and complex interventions found incomplete reporting of intervention characteristics to be common (Candy, Vickerstaff, Jones, & King, 2018; Hoffmann, Erueti, & Glasziou, 2013; Lohse, Pathania, Wegman, Boyd, & Lang, 2018). Several studies suggest that comparator interventions may be more poorly reported than experimental interventions (Byrd-Bredbenner et al., 2017; de Bruin et al., 2009; de Bruin et al., 2010; Lohse et al., 2018). Comprehensive reporting of behavioural interventions includes, amongst others, how (e.g., in-person or via a mobile phone app), where (e.g., at home or in a community centre), by whom (e.g., a nurse or pharmacist), and how often (e.g., a single 10-minute consultation or ten 1-hour sessions over a 6-month period) the intervention was delivered (Hoffmann, Glasziou, et al., 2014; Kok et al., 2016; Montgomery et al., 2018). These characteristics are relatively easy to report, and for readers and systematic reviewers to identify as 'missing' when not reported. Except for Hoffman and colleagues (2013), these reporting studies used only the primary trial papers to examine the reporting of intervention characteristics and did not include study protocols, study websites, and supplemental materials.

The active content of behavioural interventions (*what* was done to induce change) is considerably more challenging to report, and to identify as missing when not reported, than the how, where, or whom of intervention delivery. Yet, understanding the active content of experimental and comparator interventions is important for interpreting effects of individual trials, for explaining heterogeneity in treatment effects in evidence syntheses, and for estimating the likely effectiveness of an intervention in another setting. Several reporting studies provide *indirect* evidence that active content of descriptions of interventions may be particularly poor, as they found that intervention materials were rarely provided in publications (e.g., Abell, Glasziou, & Hoffmann, 2015; Albarqouni, Glasziou, & Hoffmann, 2018; Bryant, Passey, Hall, & Sanson-Fisher, 2014; Hoffmann et al., 2013; Sakzewski, Reedman, & Hoffmann, 2016). Only one study provides *direct* evidence of underreporting of active content in behavioural trials: Lorencatto and colleagues (2012) compared published intervention descriptions with intervention descriptions in full intervention manuals obtained from study authors of smoking cessation trials. They found that on average 44.4% of potential active content that could be identified in manuals could also be identified in the trial paper. Limitations of this study were the low author response rate to requests for intervention manuals (28/152, 18.4%), that it only examined experimental interventions, and – like other reporting studies – used the primary trial paper only to identify published intervention content. Hence, despite a growing literature suggesting that underreporting of study characteristics in non-pharmacological or complex interventions is common, there is no robust evidence on the degree and variability in underreporting of active content in experimental and comparator interventions in behavioural trials.

Defining the potential active content of behavioural interventions

We consider the 'active content' of behavioural interventions to be comprised of those intervention activities (done to and/or by the intervention recipient) that target behaviours relevant to the outcome of interest, and that qualify as a behaviour change technique (BCT) according to the 93-item BCT Taxonomy v1 (BCTTv1) (Michie et al., 2013). For example, in order to motivate smokers to quit smoking a pharmacist may 'provide smokers with information about the health benefits of smoking cessation, namely … ..'. This intervention activity targets the behaviour 'quitting smoking' and qualifies as BCT '5.1 Providing information about health consequences'. Additionally, following an initial quit attempt, an intervention may include planning and problem solving exercises, which targets the behaviour 'abstinence' and qualify as the BCTs '1.1 Goal setting' and '1.2 Problem solving'. For the remainder of the paper we refer to these intervention activities as 'active content', but it should be noted that the evidence base regarding the effectiveness of BCTs and their application in specific contexts is still in development. Hence, this content is presumed to be active.

In simple interventions, as few as one or two BCTs may be identified, but usually behavioural interventions are more complex and can include 20 or more BCTs (Bartlett, Sheeran, & Hawley, 2014; Glidewell et al., 2018). Even routine treatment-as-usual conversations between health care professionals and patients can contain 15 or more BCTs (de Bruin et al., 2009; de Bruin, Dima, Texier, & van Ganse, 2018; Lorencatto et al., 2012; Oberjé, Dima, Pijnappel, Prins, & de Bruin, 2015). This makes accurate reporting of the active content of behavioural interventions particularly challenging. Additionally, it can be hard for readers to identify the extent to which the active content has been comprehensively reported in trial reports.

Study aim and research questions

The literature on intervention reporting raises some important concerns about the interpretation of trial reports. For example, do published materials reliably allow readers to assess an intervention's active content and understand what an experimental intervention was adding – in terms of active content – to the comparator group in a particular trial? Can systematic reviewers reliably use published intervention descriptions to examine whether differences in effect sizes between trials can be accounted for by differences in interventions' active content?

This study aimed to determine the degree to which published (i.e., either publicly accessible or available behind a journal paywall) descriptions of interventions in experimental and comparator groups accurately reflect the full active content of these interventions. This required comparing the active content identified in published materials with the active content identified in comprehensive descriptions of experimental and comparator interventions obtained from authors of a representative set of trials. The current study was conducted as part of a larger systematic review project of smoking cessation interventions, employing novel methods for synthesising effects of behavioural interventions ([NAME], PROSPERO [NUMBER]). The database includes a large number of trials published during a period of 20 years and a range of different types of experimental and comparator interventions.

This study was guided by three main research questions:

(1) What proportion of the active content of experimental and comparator interventions is described in published materials?
(2) Does underreporting of the active content of experimental and comparator interventions in published materials vary between trials?
(3) Does underreporting of the active content of interventions in published materials vary between experimental and comparator groups within the same trial?

Whereas research question 1 relates simply to the *amount* of underreporting of active content, research questions 2 and 3 concern whether or not the published information on the active

content of interventions can be reliably used to compare different trials, and different interventions (i.e., experimental vs comparator) within a trial. Additionally, this study examined which trial characteristics are associated with the quality of active content reporting, and with author responses to our requests for additional materials describing their experimental and comparator interventions.

Methods

We searched the Cochrane Tobacco Addiction Group Specialised Register for randomised controlled trials (RCTs) of behavioural interventions (with or without pharmacological support) to support smoking cessation (search date: first of November 2015). Trials were included if they had at least six months follow-up and biochemically verified smoking cessation. We excluded trials published before 1996, that were not reported in English or in peer-reviewed journals, or that recruited participants younger than 18 years. Detailed methods of the search strategy are published elsewhere (de Bruin et al., 2016).

Materials and procedures

We first obtained published materials from all of the 142 RCTs included in the [NAME] project, including any articles (e.g., trial, protocol, or intervention development papers), their supplementary files, or materials deposited elsewhere (e.g., repositories or study websites). Next, we attempted to contact study authors from all included trials as follows. First authors were contacted by email (including two reminders); if no response was received we called them by telephone; and if unsuccessful the second and last authors were contacted by email, followed by the middle authors. This process took about eight months, with one delayed response coming in at eleven months.

Authors were asked to send additional materials describing in detail the interventions delivered to the experimental and comparator groups in their trial. Additional materials could be anything from leaflets and website materials, to detailed intervention manuals. Two researchers (NJ and AJW) completed the online BCT training and practiced BCT coding on 25 trials not included in this systematic review, prior to initiating the formal BCT coding. Both researchers – blinded to the propose of this study – independently used the v1 taxonomy of BCTs (BCTTv1; Michie et al., 2013, 2015) to reliably identify intervention components that targeted any of four key behaviours of interest (i.e., the initial quit attempt, abstinence following a successful quit attempt, adherence to smoking cessation medication, and engagement with the intervention) and that qualified as BCTs (overall Prevalence and Bias Adjusted Kappa (PABAK) = .98, agreed absent = 96.7% (124.417), agreed present = 2.6% (3306), disagree = 0.8% (989)) (Black et al., 2018; de Bruin et al., 2016). The researchers first examined the published intervention materials and then identified the active content described in the additional materials that the authors had sent. The BCTTv1 taxonomy was used as this is the most widely-used instrument for describing and coding interventions in systematic reviews.

For comparator groups, especially those receiving treatment-as-usual, authors often do not have (detailed) intervention materials. In order to address this information gap, we adopted a previously validated method for collecting detailed descriptions of comparator interventions (de Bruin et al., 2009, 2010). For that, we first extracted from smoking cessation guidelines and related literature (Fiore et al., 2008; HSE Tobacco Control Framework Implementation group, 2013; McEwen, 2014; Michie et al., 2013; Michie, Hyder, Walia, & West, 2011; NHS Scotland, 2017), the potential active content (i.e., practical intervention activities that target smoking cessation and qualify as a BCT) of comprehensive treatment-as-usual smoking cessation programmes. We discussed this list with our advisory board comprised of smoking cessation professionals and policy makers, and generated a final list of 60 practical intervention activities that qualified as potential active content. These practical activities were included as items in a 'comparator intervention checklist', which was sent to authors, asking them to report which of these had been delivered to the majority of participants in the

comparator arm of their trial. Because all authors received the same checklist, a syntax was used to link author responses to BCTs and hence these were not part of the intercoder reliability calculations reported above. Also in this study, the checklist demonstrated high internal consistency reliability (Cronbach's alpha = .92) and predictive validity (see pre-print on https://osf.io/uvepa/).

In sum, the published materials as well as unpublished materials (manuals, websites, checklists) were all sources of information on what practical intervention activities had been delivered to experimental and comparator groups. From these materials we extracted the active content through identifying which of these activities targeted smoking cessation behaviours (quitting, abstaining, medication adherence, treatment engagement) and that qualified as a BCT using the BCTTv1. An elaboration on these methods is included in Appendix 1.

Finally, independent double coding (NJ, AJW, and NB) was conducted to evaluate the degree to which the published and unpublished materials combined, provided a clear and comprehensive description of the active content of each of the experimental and comparator interventions (reporting quality: 3 = well-described, 2 = moderately described, 1 = poorly described, 0 = not described; PABAK = .79). Disagreements were resolved through discussion.

The local ethics committee (Institute of Applied Health Science, University of Aberdeen) judged that ethics was not required for collecting data from study authors as this did not involve collecting personal information.

Analyses

In the absence of evidence on the relative effectiveness of the 93 BCTs in the BCTTv1 taxonomy in the context of smoking cessation interventions (i.e., some may be more effective than others), for simplicity we assumed that all intervention activities that qualify as active content are equally effective (de Bruin et al., 2009, 2010). Hence, the active content score of the experimental and comparator interventions in our analyses is reflected by the total number of BCTs that targeted any of the four behaviours considered to be of interest. When authors indicated in the manuscript or checklist that intervention groups received the experimental intervention as an addition to the comparator intervention (and not instead of), their active content score also included the comparator group checklist data (76/204 [37%] of intervention groups).

Descriptive statistics were used to examine the total number of BCTs identified in the published materials, and the additional number of BCTs identified in the unpublished materials obtained from study authors (RQ1). Graphs and correlational analyses were used to examine variability in underreporting of the number of BCTs between experimental and comparator groups from different trials, as well as between groups from the same trial (RQ2). These analyses were restricted to the experimental and comparator groups for which additional materials were obtained from study authors.

Binomial mixed-effects logistic regression was used to analyse factors that explained the quality of reporting of experimental and comparator interventions. The dependent variable was the logit transformed proportion of BCTs reported in published materials, out of the total number of BCTs identified. Predictors were (a) time (coded as 'publication year – 1996'), (b) journal impact factor (coded as 'log(impact factor)'), (c) behavioural versus medical journals (coded 1/0, respectively), (d) trial sample size (coded as 'log(trial sample size)'), (e) experimental versus comparator groups (coded 1/0, respectively), and (f) primary mode of delivery: person-delivered versus written interventions (i.e., digital or print interventions; coded 1/0, respectively). All predictors were entered into the model jointly. The model included random effects for trials (to account for the multilevel structure of the data, with groups nested within trials) and individual groups within trials (to account for potential overdispersion in the logit transformed proportions). For the dependent variable to be meaningful, the analysis was restricted to the groups coded as well-described.

Logistic regression analyses using generalised estimating equations (GEE) were conducted to identify whether groups for which additional unpublished materials could be retrieved were

different from those for which no unpublished materials were retrieved (i.e., missing data analyses), using publication year, journal impact factor, journal type, trial sample size, group and primary mode of intervention delivery as predictors (all variables were coded as described above). Trials were used as the clustering variable and an exchangeable correlation structure was assumed. Our a-priori published analysis plan, including hypotheses regarding the direction of effects of these predictors, can be found on the Open Science Framework (https://osf.io/23hfv/).

Results

From the 142 randomised controlled trials included in the review, additional unpublished materials (including checklists) were retrieved for 129/204 (63%) experimental and 93/142 (65%; 69% if 8 passive comparator groups are removed) comparator groups. For experimental interventions, authors sent a range of materials, such as the intervention manuals, trial protocols, and self-help materials. For the comparator groups, 80/93 (86%) completed checklists were obtained and for 53/93 (57%) existing documents describing the comparator intervention were sent, such as manuals and self-help materials.

Table 1 shows the quality judgements for active content reporting for the experimental and comparator interventions for which no additional materials were obtained versus those for which additional materials were obtained. Overall, the active content of 144/204 (71%) of the experimental and 110/142 (77%) of the comparator interventions were judged to be well-described. The quality of descriptions was rated significantly higher for groups for which additional materials were obtained than for groups for which no additional materials were obtained (Mann–Whitney $U =$ 6655, $p < .001$). Of the 29 experimental interventions rated as well-described based on the published materials alone, 23 had detailed descriptions in the primary trial paper and in six cases the intervention manuals were publicly available. Of the 21 well-described comparator groups based on the published materials alone, 11 were attention or wait list controls, 6 adequately described the content in the primary article, 2 reported using a publicly available standardised method (i.e., 5As method) (Fiore, 2000; Fiore et al., 2008), and 2 reported using a publicly available leaflet or website.

Proportion of active content identifiable in published materials

For the 222 experimental and comparator groups for which authors sent additional materials, 3403 BCTs were identified for the experimental interventions (*Mdn* = 25, *IQR* = 18, range 2–62 BCTs) and 1891 for the comparator interventions (*Mdn* = 19, *IQR* = 28, range 0–51 BCTs). Of these, only 35% (1200/3403) of the BCTs in the experimental interventions and 26% (491/1891) of those in the

Table 1. Coder judgements on how well the active content of smoking cessation interventions was described for treatment arms for which no additional intervention materials were obtained from trial authors, and those for which trial authors did sent additional intervention materials.

	Experimental		Comparator	
	No material retrieved	Material retrieved	No material retrieved	Material retrieved
Well-described	29 (39%)	115 (89%)	21 (43%)	89 (96%)
Moderately described	19 (25%)	8 (6%)	7 (15%)	0 (0%)
Poorly described	22 (29%)	6 (5%)	17 (36%)	0 (0%)
Not described	5 (7%)	0 (0%)	4 (8%)	4 (4%)[a]
Total	75 (100%)	129 (100%)	49 (100%)	93 (100%)

[a]For four control groups additional materials were received but these did not include any useable information.
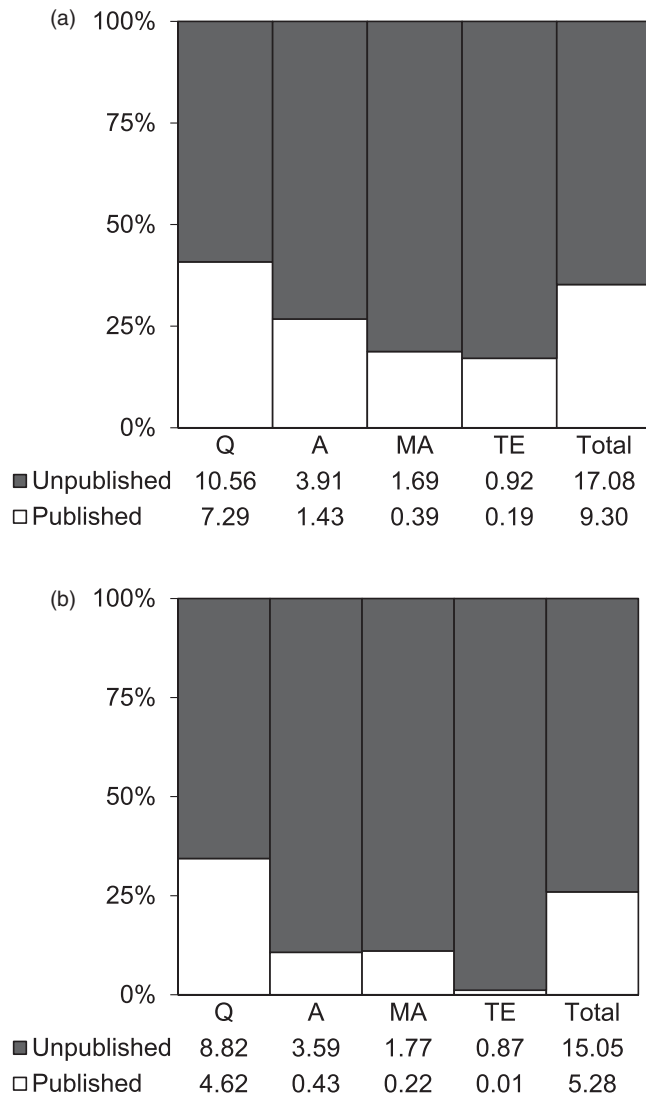
Figure 1. Percentage of total (bars) and mean (numbers) number of BCTs coded from published and the additional BCTs coded from unpublished available materials, by behavioural target (Q = quitting; A = abstinence; MA = medication adherence; TE = treatment engagement) and group (a. experimental; b. comparator).

comparator interventions could be identified in published materials. BCTs targeting initiating a quit attempt appeared to be best reported, followed by BCTs targeting smoking abstinence, medication adherence, and least well reported were those supporting intervention engagement (see Figure 1).

Table 2 shows the reporting quality at the level of individual BCTs (across the four behaviours), for the experimental and comparator groups separately; and across treatment arms. BCTs are ordered from best to worst reported, based on the total scores across treatment arms. The quality of reporting of different BCTs was very similar between experimental and comparator arms, with a correlation of $r = .64$ ($p < .01$) if all BCTs identified at least once in each treatment arm were considered; $r = .81$ ($p < .01$) if only BCTs were considered that were reported at least 5 times in each arm; and $r = .90$ ($p < .01$) if only BCTs were considered that were reported at least 10 times in each arm.

Table 2. Reporting quality of each behaviour change technique for experimental and comparator groups separately, and across treatment arms (total). Behaviour change techniques are sorted from highest to lowest reporting quality across treatment arms.

BCT	Experimental % Published	Comparator % Published	Total % Published
1.7 Review outcome goal(s)	3/3 (100.00%)	–	3/3 (100.00%)
7.4 Remove access to the reward	2/2 (100.00%)	–	2/2 (100.00%)
10.7 Self-incentive	2/2 (100.00%)	3/3 (100.00%)	5/5 (100.00%)
10.8 Incentive (outcome)	4/4 (100.00%)	–	4/4 (100.00%)
14.2 Punishment	1/1 (100.00%)	–	1/1 (100.00%)
14.3 Remove reward	6/6 (100.00%)	–	6/6 (100.00%)
14.5 Rewarding completion	2/2 (100.00%)	–	2/2 (100.00%)
14.6 Situation-specific reward	1/1 (100.00%)	–	1/1 (100.00%)
14.10 Remove punishment	1/1 (100.00%)	–	1/1 (100.00%)
16.2 Imaginary reward	1/1 (100.00%)	–	1/1 (100.00%)
10.10 Reward (outcome)	17/18 (94.44%)	0/1 (0.00%)	17/19 (89.47%)
16.3 Vicarious consequences	5/6 (83.33%)	–	5/6 (83.33%)
2.5 Monitoring of outcome(s) of behaviour without feedback	4/6 (66.67%)	4/4 (100.00%)	8/10 (80.00%)
2.7 Feedback on outcome(s) of behaviour	23/30 (76.67%)	7/8 (87.50%)	30/38 (78.95%)
9.1 Credible source	36/49 (73.47%)	18/23 (78.26%)	54/72 (75.00%)
10.3 Non-specific reward	9/13 (69.23%)	3/4 (75.00%)	12/17 (70.59%)
3.1 Social support (unspecified)	147/191 (76.96%)	71/129 (55.04%)	218/320 (68.13%)
14.9 Self-reward	5/6 (83.33%)	1/3 (33.33%)	6/9 (66.67%)
6.1 Demonstration of the behaviour	9/12 (75.00%)	2/5 (40.00%)	11/17 (64.71%)
2.4 Self-monitoring of outcome(s) of behaviour	8/11 (72.73%)	0/2 (0.00%)	8/13 (61.54%)
1.5 Review behaviour goal(s)	8/14 (57.14%)	4/6 (66.67%)	12/20 (60.00%)
7.3 Reduce prompts/cues	22/38 (57.89%)	8/16 (50.00%)	30/54 (55.56%)
11.1 Pharmacological support	107/175 (61.14%)	50/111 (45.05%)	157/286 (54.90%)
2.1 Monitoring of behaviour by others without feedback	9/13 (69.23%)	1/6 (16.67%)	10/19 (52.63%)
1.8 Behavioural contract	6/11 (54.55%)	2/5 (40.00%)	8/16 (50.00%)
1.3 Goal setting (outcome)	5/10 (50.00%)	–	5/10 (50.00%)
9.3 Comparative imagining of future outcomes	2/4 (50.00%)	–	2/4 (50.00%)
12.6 Body changes	1/2 (50.00%)	–	1/2 (50.00%)
8.3 Habit formation	3/7 (42.86%)	3/5 (60.00%)	6/12 (50.00%)
1.1 Goal setting (behaviour)	74/134 (55.22%)	33/88 (37.50%)	107/222 (48.20%)
3.2 Social support (practical)			

Table 2. Continued.

BCT	Experimental % Published	Comparator % Published	Total % Published
	9/20 (45.00%)	5/10 (50.00%)	14/30 (46.67%)
4.5 Tell to act	41/83 (49.40%)	29/71 (40.85%)	70/154 (45.45%)
14.1 Behaviour cost	4/7 (57.14%)	0/2 (0.00%)	4/9 (44.44%)
4.2 Information about antecedents	20/46 (43.48%)	9/20 (45.00%)	29/66 (43.94%)
6.2 Social comparison	24/57 (42.11%)	9/20 (45.00%)	33/77 (42.86%)
12.5 Adding objects to the environment	13/32 (40.63%)	4/8 (50.00%)	17/40 (42.50%)
3.3 Social support (emotional)	2/6 (33.33%)	2/4 (50.00%)	4/10 (40.00%)
6.3 Information about others' approval	5/16 (31.25%)	4/7 (57.14%)	9/23 (39.13%)
5.3 Information about social and environmental consequences	47/115 (40.87%)	21/61 (34.43%)	68/176 (38.64%)
1.2 Problem solving	66/152 (43.42%)	25/95 (26.32%)	91/247 (36.84%)
13.1 Identification of self as role model	10/34 (29.41%)	6/12 (50.00%)	16/46 (34.78%)
5.1 Information about health consequences	52/139 (37.41%)	22/77 (28.57%)	74/216 (34.26%)
11.4 Paradoxical instructions	2/4 (50.00%)	0/2 (0.00%)	2/6 (33.33%)
2.3 Self-monitoring of behaviour	19/65 (29.23%)	7/19 (36.84%)	26/84 (30.95%)
1.9 Commitment	21/66 (31.82%)	6/26 (23.08%)	27/92 (29.35%)
9.2 Pros and cons	29/79 (36.71%)	5/38 (13.16%)	34/117 (29.06%)
12.3 Avoidance/reducing exposure to cues for the behaviour	32/114 (28.07%)	17/55 (30.91%)	49/169 (28.99%)
4.1 Instruction on how to perform the behaviour	79/248 (31.85%)	30/143 (20.98%)	109/391 (27.88%)
2.2 Feedback on behaviour	15/49 (30.61%)	0/8 (0.00%)	15/57 (26.32%)
11.2 Reduce negative emotions	36/111 (32.43%)	7/59 (11.86%)	43/170 (25.29%)
8.6 Generalisation of target behaviour	10/42 (23.81%)	5/26 (19.23%)	15/68 (22.06%)
13.2 Framing/reframing	13/63 (20.63%)	8/38 (21.05%)	21 (20.79%)
8.1 Behavioural practice/rehearsal	11/42 (26.19%)	2/22 (9.09%)	13/64 (20.31%)
15.2 Mental rehearsal of successful performance	1/5 (20.00%)	–	1/5 (20.00%)
12.2 Restructuring the social environment	18/93 (19.35%)	9/47 (19.15%)	27/140 (19.29%)
13.5 Identity associated with changed behaviour	7/37 (18.92%)	4/24 (16.67%)	11/61 (18.03%)
1.4 Action planning	15/78 (19.23%)	8/62 (12.90%)	23/140 (16.43%)
1.6 Discrepancy between current behaviour and goal	6/37 (16.22%)	3/20 (15.00%)	9/57 (15.79%)
8.2 Behaviour substitution	13/80 (16.25%)	7/48 (14.58%)	20/128 (15.63%)
7.1 Prompts/cues	4/35 (11.43%)	3/15 (20.00%)	7/50 (14.00%)
5.5 Anticipated regret	2/8 (25.00%)	0/8 (0.00%)	2/16 (12.50%)

Table 2. Continued.

BCT	Experimental % Published	Comparator % Published	Total % Published
5.6 Information about emotional consequences	8/64 (12.50%)	4/40 (10.00%)	12/104 (11.54%)
12.4 Distraction	9/66 (13.64%)	3/43 (6.98%)	12/109 (11.01%)
10.1 Material incentive (behaviour)	1/6 (16.67%)	0/4 (0.00%)	1/10 (10.00%)
10.9 Self-reward	4/64 (6.25%)	7/46 (15.22%)	11/110 (10.00%)
10.4 Social reward	9/81 (11.11%)	3/40 (7.50%)	12/121 (9.92%)
12.1 Restructuring the physical environment	10/103 (9.71%)	5/50 (10.00%)	15/153 (9.80%)
15.3 Focus on past success	3/46 (6.52%)	2/38 (5.26%)	5/84 (5.95%)
15.4 Self-talk	3/31 (9.68%)	0/22 (0.00%)	3/53 (5.66%)
5.2 Salience of consequences	3/46 (6.52%)	0/31 (0.00%)	3/77 (3.90%)
15.1 Verbal persuasion about capability	1/53 (1.89%)	0/34 (0.00%)	1/87 (1.15%)
2.6 Biofeedback	0/30 (0.00%)	0/25 (0.00%)	0/55 (0.00%)
4.4 Behavioural experiments	0/1 (0.00%)	–	0/1 (0.00%)
5.4 Monitoring of emotional consequences	0/6 (0.00%)	–	0/6 (0.00%)
7.5 Remove aversive stimulus	0/1 (0.00%)	0/2 (0.00%)	0/3 (0.00%)
7.6 Satiation	0/1 (0.00%)	–	0/1 (0.00%)
7.8 Associative learning	0/1 (0.00%)	–	0/1 (0.00%)
8.4 Habit reversal	0/7 (0.00%)	0/1 (0.00%)	0/8 (0.00%)
8.7 Graded tasks/BS9 set graded tasks	0/2 (0.00%)	–	0/2 (0.00%)
10.2 Material reward (behaviour)	0/19 (0.00%)	0/17 (0.00%)	0/36 (0.00%)
11.3 Conserving mental resources	0/33 (0.00%)	0/31 (0.00%)	0/64 (0.00%)
14.4 Reward approximation	0/6 (0.00%)	0/1 (0.00%)	0/7 (0.00%)
4.3 Re-attribution	–	–	–
7.2 Cue signalling reward	–	–	–
7.7 Exposure	–	–	–
10.5 Social incentive	–	–	–
10.6 Non-specific incentive	–	–	–
10.11 Future punishment	–	–	–
13.3 Incompatible beliefs	–	–	–
13.4 Valued self-identify	–	–	–
14.7 Reward incompatible behaviour	–	–	–
14.8 Reward alternative behaviour	–	–	–
16.1 Imaginary punishment	–	–	–
Total	1200/3403 (35.26%)	491/1891 (25.97%)	1691/5294 (31.94%)

Variability in active content reporting between trials

To examine whether published intervention descriptions would still allow for the comparison of the interventions' active content despite the above-illustrated underreporting, we examined the degree to which reporting quality varied between trials.
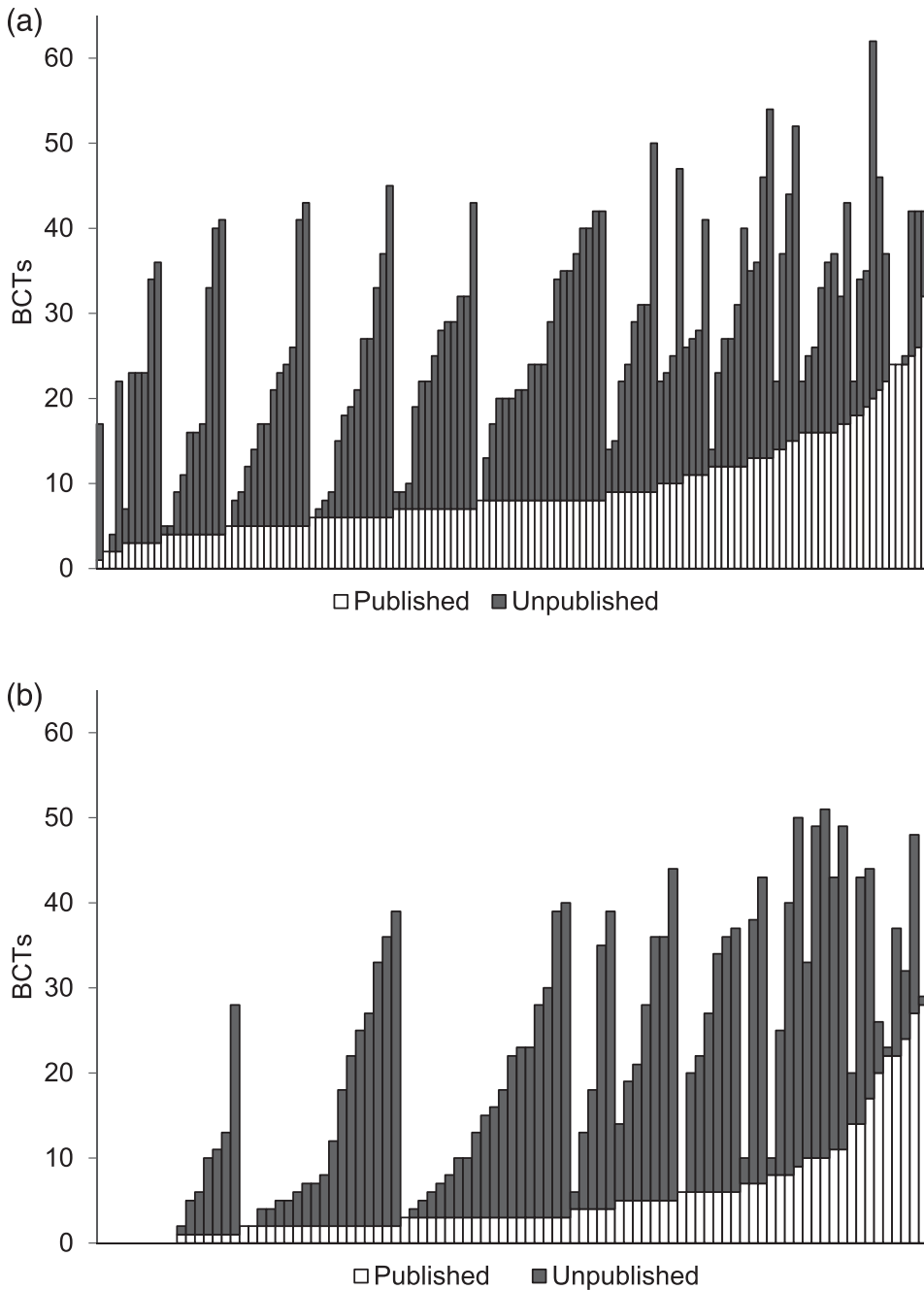
Figure 2. BCTs coded from published (white bars) and additional BCTs coded from unpublished (grey bars) materials for each experimental (a) and comparator (b) group. Groups are ordered by number of behaviour change techniques coded from published materials.

Figure 2(a and b) show the experimental and comparator groups respectively (x-axis), ordered by number of BCTs identified in published materials (white bars), after which the number of additional BCTs identified in the unpublished materials were added (grey bars). While the published number and the total number of BCTs are related to each other ($r = 0.45$ and $r = 0.53$ for experimental and comparator groups, respectively), one can immediately see that there is considerable variability in
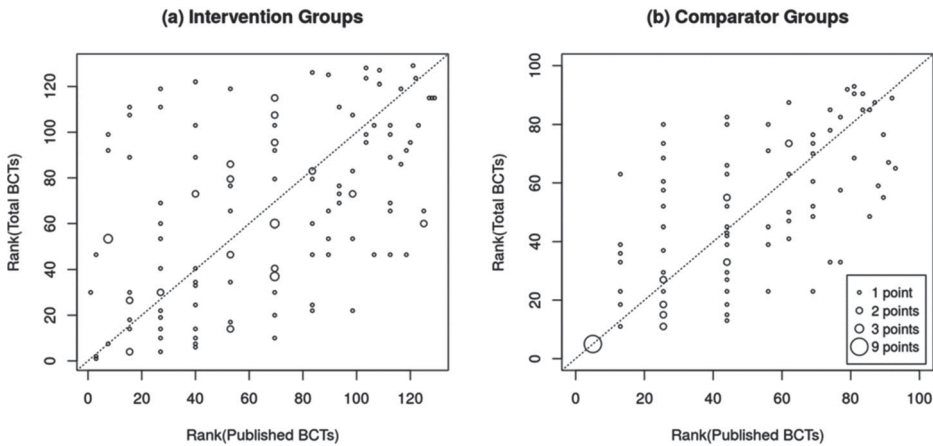
Figure 3. The ranking of (a) experimental and (b) comparator groups based on the number of BCTs identified in published versus all intervention materials obtained. The size of the circle reflects the number of experimental or comparator groups contributing to that data point.

reporting between studies. For example, experimental interventions that appear to deliver an identical number of BCTs based on published materials (e.g., five) turn out to vary widely when BCTs identified in unpublished materials are also considered (i.e., five to forty-three BCTs). Consequently, when ranking experimental and comparator groups based on their active content score (i.e., from receiving the least to receiving the most BCTs), the ranking of groups looks quite different when only published or also unpublished materials are considered (see Figure 3(a and b); $\tau_b = 0.35$ for experimental, $\tau_b = 0.55$ for comparators). Most groups were ranked substantially higher (above the diagonal line) or lower (below the diagonal line) when unpublished materials were also considered. Hence, there is considerable between-study variability in reporting quality, which affects how much active content experimental or comparator interventions from different trials appear to deliver relative to each other.

Variability in active content reporting within trials

Another important characteristic of accurate reporting is that readers and systematic reviewers should be able examine the additional value of an experimental intervention over its comparator. Reporting quality for unique active content (i.e., the BCTs identified only for the comparator or the experimental interventions within a trial) was nearly identical to the reporting quality for all BCTs identified: only 29.4% (83/282) of the unique comparator BCTs and 40.6% (729/1794) of unique experimental BCTs had been identified in published materials.

Figure 4a shows a scatterplot of the difference in the number of BCTs delivered to intervention and comparator groups from the same trial, based on published descriptions (x-axis) and all available materials (y-axis, $r = .57$). For 27/110 comparisons (25%; points on the diagonal line), the difference was identical. For 30/110 (27%) comparisons, the difference between the experimental and comparator intervention decreased (*Mdn* = 2, *IQR* = 9, range 1–27 BCTs) and for 53/110 (48%) comparisons, the difference increased when unpublished materials in addition to published materials were considered (*Mdn* = 4, *IQR* 8, range 1–22 BCTs). Accordingly, if you rank the trials based on the difference in the number of experimental and comparator BCTs (e.g., from interventions that add only 1 BCT to their comparator, to interventions that add 30 BCTs to their comparator), the order of the trials changed considerably when unpublished materials were also considered (see Figure 4b; $\tau_b = 0.52$). Hence, these analyses show that BCT reporting quality also varies between groups from within the same trial, which in turn affects how superior – in terms of active content – interventions from different trials appear to be in relation to their comparator.
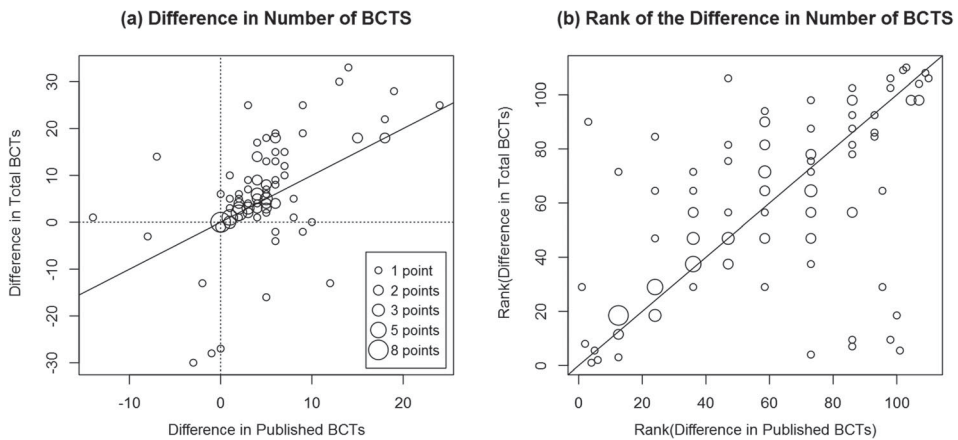
Figure 4. The difference in the number of BCTs between experimental groups and their comparator based on published versus all materials (a) and the ranking of these differences (b). The size of the circle reflects the number of experimental or comparator groups contributing to that data point.

Table 3. Results from the binomial mixed-effects logistic regression model predicting active content reporting quality.

Coefficient	Estimate	SE	z-val	p-val
Intercept	−0.922	1.019	−0.91	.365
Publication Year^S	0.023	0.023	0.99	.323
Journal Impact Factor$^\#$	−0.289	0.161	−1.79	.073
Journal Type (1 = behavioural vs. 0 = medical)	−0.341	0.264	−1.29	.197
Study Sample Size$^\#$	0.101	0.141	0.71	.475
Group (1 = Experimental vs. 0 = Comparator)	0.339	0.105	3.22	.001
Primary Delivery Mode (1 = interpersonal vs. 0 = written)	−0.133	0.348	−0.38	.702

Notes: ^{S} coded as 'publication year – 1996'; $^\#$ log transformed.

To examine whether these findings could in part be explained by method bias (i.e., checklists were more often used for comparators and protocols more often for experimental interventions, which could have produced differences in how comprehensive the information obtained was), we repeated these analyses only for the 96 trials with well-described experimental and comparator groups. However, results were virtually identical, suggesting that these primary results presented would be similar also if the methods of data collection would have been identical for all study arms.

Factors associated with reporting quality and the retrieval of additional materials

Based on the binomial mixed-effects logistic regression model, experimental interventions were better reported (i.e., number of BCTs published / total number of BCTs identified) in the published materials than comparator interventions (OR = 1.40, 95% CI [1.14–1.72], p = .001). Higher journal impact factor appeared to be negatively associated with reporting quality (OR = 0.75, 95% CI [0.55–1.03], p = .073). There was little evidence that other variables were associated with reporting quality (all p ≥ .20, see Table 3).

The results from the GEE logistic regression model (Table 4) indicated that the chances of obtaining additional materials from study authors were higher for more recently published trials (OR = 1.10, 95%CI [1.03–1.17], p = .003) and trials published in behavioural journals (OR = 2.80, 95%CI [1.25–6.26], p = .012). p-values for the other predictors were all ≥.25.

Discussion

In the current study, we managed to collect additional materials describing experimental *and* comparator interventions from a large sample of smoking cessation trials, and found that underreporting

Table 4. GEE logistic regression model[a] examining predictors of whether authors provided additional intervention materials upon request.

Coefficient	Estimate	SE	z-val	p-val
Intercept	−1.592	1.470	−1.08	.279
Publication Year[$]	0.093	0.032	2.94	.003
Journal Impact Factor[#]	0.208	0.245	0.85	.395
Journal Type (1 = behavioural vs. 0 = medical)	1.029	0.411	2.50	.012
Study Sample Size[#]	0.053	0.229	0.23	.818
Group (1 = intervention vs. 0 = control)	−0.152	0.134	−1.13	.257
Primary Delivery Mode (1 = interpersonal vs. 0 = written)	0.372	0.326	1.14	.254

Notes: [a]We originally had planned to report a mixed-effects logistic regression model with random trials effects for these analyses. Although that model produced very similar p-values as the GEE model (and identical conclusions), the mixed-effects model gave estimates that were unreasonably large and hard to interpret (since for many trials, additional materials were either provided for none or all groups, leading to a very large variance component); [$] coded as 'publication year − 1996'; [#] log transformed.

of the presumed active content of behavioural smoking cessation interventions is common and – importantly – varies considerably between different trials as well as between different groups from within the same trial (experimental interventions seem to be better reported than comparators). This is important as readers, systematic reviewers, practitioners, and policy makers require comparable descriptions of experimental and comparator interventions when interpreting (e.g., What content did the intervention add to the comparator?), comparing (e.g., What content makes intervention x more effective than intervention y?), or generalising (e.g., What content does intervention x add to the care provided in my clinic?) trial effects. Comprehensive intervention descriptions are also an important condition for the replication and implementation of interventions. When more than two-thirds of the presumed active intervention content is missing from a published literature, and the degree of omission varies so much between trials and between groups within trials – as observed here – it undermines confidence that the published literature can be used to confidently answer these questions.

On the up-side, and this also reflects one of the major strengths of this study, with a novel reporting tool (the checklist) and persistence of requests, we were able to contact almost all author teams of the included trials published over a period of 20 years. For two-thirds of the trials we received valuable additional material describing the experimental *and* comparator interventions. We were less likely to receive those from older studies and those published in medical journals, but as publication year and journal type were only very weakly associated with reporting quality, we can be reasonably confident that the results generalise to all the included trials.

The percentage of well-described interventions was an impressive 50% higher for the trials for which authors sent us additional materials (see Table 1). This difference was not based on just a comparison with intervention descriptions in the primary trial paper (which is what most reporting studies are using), but on comparisons with any published material we could find describing the intervention and trial (e.g., protocol papers, supplements, trial website, intervention development papers). Despite the high author response rate, the active content of 29% (60/204) of the experimental and 23% (32/142) of the comparator interventions (representing 37% [53/142] of trials), was scored as partly or poorly described. Note that most of these authors did respond to our emails but were not able to provide this information. Consequently, although they may report well on other trial and intervention characteristics, the absence of comprehensive descriptions of presumed active content complicates the interpretation, comparison, and generalisability of results in one-third of the included smoking cessation trials; and presents an important barrier to their replication and implementation.

Our study builds on and extends a growing evidence base on intervention reporting. There are a number of studies that show that reporting of intervention characteristics is suboptimal (Albarqouni et al., 2018; Candy et al., 2018; di Ruffano et al., 2017; Glasziou et al., 2014; Glasziou, Meats, Heneghan, & Shepperd, 2008; Hoffmann et al., 2013; Lohse et al., 2018; Mhizha-Murira, Drummond, Klein, & dasNair, 2018; Warner, Kelly, Reidlinger, Hoffmann, & Campbell, 2017; Lorencatto et al., 2012). The active content of interventions is evidently a crucial component of adequate trial reporting, yet

amongst all these studies we could find only one that directly examined active content reporting in behavioural trials (Lorencatto et al., 2012). This study revealed considerable underreporting of active content (55.6%) in the primary trial papers, for the 28/152 (18.4%) experimental interventions for which trial authors provided intervention manuals. Hence, this study provided an important signal that reporting of active content in behavioural interventions may be suboptimal, but suffered from methodological limitations that we were able to address in this study by collecting data on a larger and more representative set of experimental *and* comparator interventions, examine all published materials we could identify (not only the primary trial paper), explore variability in reporting quality between trials *and* between groups from the same trials, and run pre-registered regression models to explain reporting quality and author response rates.

The present study also has several potential limitations. First, we were unable to obtain unpublished information for 36% of the intervention and comparator groups, mainly from older studies and those published in medical journals. However, as we did not find that publication year and journal type were related to reporting quality, the high response rate combined with these meta-analyses provide reasonable confidence that the results obtained generalise to the full sample of smoking cessation trials. Secondly, we only examined smoking cessation interventions, which could limit the generalisability of our findings to other behavioural interventions. It is worth noting, however, that the smoking cessation literature captures a wide range of experimental and comparator interventions used in the behavioural intervention literature. Also, in discussions with systematic reviewers who have attempted to code BCTs in experimental and comparator interventions, an often-heard issue is the poor reporting of the interventions' active content. Hence, the present study has examined the scope, causes and possible implications of a well-recognised problem amongst systematic reviewers of behavioural interventions; and has provided methods to tackle this issue in future systematic reviews. Third, the active content for experimental interventions was primarily identified through protocols and manuals that authors sent us, whereas the active content of the comparator interventions was primarily identified from the checklist. This could have resulted in method bias (e.g., the checklist provided a better insight in the comparator intervention than the intervention protocols did for the experimental interventions), although we did not previously find evidence for this (de Bruin et al., 2010). Finally, a general limitation for these type of content analyses for behavioural interventions is that we do not know the relative effectiveness of each of the presumed active content components, and pragmatically chose to treat them all as equal.

The findings of this study have several implications. First, readers' access to comprehensive descriptions of the active content of experimental and comparator interventions has considerable room for improvement. Authors should ensure that detailed descriptions are published of the interventions delivered to all treatment arms in their trial, and upload this information on publicly accessible platforms such as the Open Science Framework. Our call for improved reporting of the potential active components of behavioural interventions extends beyond the mere absence/presence of BCTs, as it also seems important that authors report *how* they have applied the BCTs in their interventions (e.g., in terms of BCT tailoring and other possible parameters of effectiveness Kok et al., 2016; Peters, de Bruin, & Crutzen, 2015). Research funders, journal editors, reviewers, and readers could play an important role in stimulating this. Secondly, in an attempt to identify the most effective ingredients of behavioural interventions, many systematic reviews report meta-regression analyses in which they associate effect sizes with BCTs identified in published intervention descriptions. Even if intervention reporting were unexpectedly to be twice as good in other literatures, this underreporting would still raise concerns about potential biases, imprecision and confounding in these analyses. Hence, future systematic reviews should aim to ensure that comprehensive descriptions for experimental and comparator interventions are available prior to conducting such analyses, for example, by contacting study authors or restricting the analyses to well-described studies.

In conclusion, despite the availability of various reporting guidelines, published reports of the active content of smoking cessation interventions are often incomplete. This study demonstrated that active content reporting varies considerably between and within studies, which limits readers'

ability to interpret, compare, and generalise the effects reported by these trials. With considerable investment, however, a substantial amount of information can be retrieved from study authors up to 20 years back – although for approximately one-third of trials this was not possible. To tackle incomplete reporting – and its consequences for science, policy, and practice – a concerted effort from multiple stakeholders is required. We would also caution that readers – be it scientists, practitioners, or policy makers – keep in mind the issue of incomplete and variable reporting of the active content experimental and especially comparator interventions when interpreting, comparing, and generalising intervention effects reported in smoking cessation trials.

ORCID

Marijn de Bruin   http://orcid.org/0000-0003-3100-6803
Wolfgang Viechtbauer   http://orcid.org/0000-0003-3463-4063
Maarten C. Eisma   http://orcid.org/0000-0002-6109-2274
Jamie Hartman-Boyce   http://orcid.org/0000-0001-9898-3049
Robert West   http://orcid.org/0000-0001-6398-0921
Susan Michie   http://orcid.org/0000-0003-0063-6378
Marie Johnston   http://orcid.org/0000-0003-0124-4827

References

Abell, B., Glasziou, P., & Hoffmann, T. (2015). Reporting and replicating trials of exercise-based cardiac rehabilitation: Do we know what the researchers actually did? *Circulation: Cardiovascular Quality and Outcomes*. doi:10.1161/CIRCOUTCOMES.114.001381

Albarqouni, L., Glasziou, P., & Hoffmann, T. (2018). Completeness of the reporting of evidence-based practice educational interventions: A review. *Medical Education*, *52*(2), 161–170. doi:10.1111/medu.13410

Albrecht, L., Archibald, M., Arseneau, D., & Scott, S. D. (2013). Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations. *Implementation Science*, *8*(1), 52. doi:10.1186/1748-5908-8-52

Bartlett, Y. K., Sheeran, P., & Hawley, M. S. (2014). Effective behaviour change techniques in smoking cessation interventions for people with chronic obstructive pulmonary disease: A meta-analysis. *British Journal of Health Psychology*, *19*(1), 181–203. doi:10.1111/bjhp.12071

Black, N., Williams, A. J., Javornik, N., Scott, C., Johnston, M., Eisma, M. C., … de Bruin, M. (2018). Enhancing behavior change technique coding methods: Identifying behavioral targets and delivery styles in smoking cessation trials. *Annals of Behavioral Medicine*. doi:10.1093/abm/kay068

Bryant, J., Passey, M. E., Hall, A. E., & Sanson-Fisher, R. W. (2014). A systematic review of the quality of reporting in published smoking cessation trials for pregnant women: An explanation for the evidence-practice gap? *Implementation Science*, *9*(1), 94. doi:10.1186/s13012-014-0094-z

Byrd-Bredbenner, C., Wu, F., Spaccarotella, K., Quick, V., Martin-Biggers, J., & Zhang, Y. (2017). Systematic review of control groups in nutrition education intervention research. *International Journal of Behavioral Nutrition and Physical Activity*, *14*(1), 91. doi:10.1186/s12966-017-0546-3

Candy, B., Vickerstaff, V., Jones, L., & King, M. (2018). Description of complex interventions: Analysis of changes in reporting in randomised trials since 2002. *Trials*, *19*(1), 110. doi:10.1186/s13063-018-2503-0

de Bruin, M., Dima, A. L., Texier, N., & van Ganse, E. (2018). Explaining the amount and consistency of medical care and self-management support in asthma: A survey of primary care providers in France and the United Kingdom. *The Journal of Allergy and Clinical Immunology: In Practice*. doi:10.1016/j.jaip.2018.04.039

de Bruin, M., Viechtbauer, W., Eisma, M. C., Hartmann-Boyce, J., West, R., Bull, E., … Johnston, M. (2016). Identifying effective behavioural components of intervention and comparison group support provided in SMOKing cEssation (IC-SMOKE) interventions: A systematic review protocol. *Systematic Reviews*, *5*(1), 77.

de Bruin, M., Viechtbauer, W., Hospers, H. J., Schaalma, H. P., & Kok, G. (2009). Standard care quality determines treatment outcomes in control groups of HAART-adherence intervention studies: Implications for the interpretation and comparison of intervention effects. *Health Psychology*, *28*(6), 668–674. doi:10.1037/a0015989

de Bruin, M., Viechtbauer, W., Schaalma, H. P., Kok, G., Abraham, C., & Hospers, H. J. (2010). Standard care impact on effects of highly active antiretroviral therapy adherence interventions: A meta-analysis of randomized controlled trials. *Archives of Internal Medicine*, *170*(3), 240–250. doi:10.1001/archinternmed.2009.536

di Ruffano, L. F., Dinnes, J., Taylor-Phillips, S., Davenport, C., Hyde, C., & Deeks, J. J. (2017). Research waste in diagnostic trials: A methods review evaluating the reporting of test-treatment interventions. *BMC Medical Research Methodology*, *17*(1), 32. doi:10.1186/s12874-016-0286-0

Eldredge, L. K. B., Markham, C. M., Ruiter, R. A., Fernández, M. E., Kok, G., & Parcel, G. S. (2016). *Planning health promotion programs: An intervention mapping approach*. John Wiley & Sons.

Fiore, M. C. (2000). A clinical practice guideline for treating tobacco use and dependence: A US public health service report. *JAMA: Journal of the American Medical Association*. doi:10.1001/jama.283.24.3244

Fiore, M. C., Jaén, C. R., Baker, T. B., Bailey, W. C., Benowitz, N. L., & Wewers, M. E. (2008). *Treating tobacco use and dependence: 2008 update*. Clinical Practice Guideline. DIANE publishing.

Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., … Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, *383*(9913), 267–276. doi:10.1016/S0140-6736(13)62228-X

Glasziou, P., Meats, E., Heneghan, C., & Shepperd, S. (2008). What is missing from descriptions of treatment in trials and reviews? *BMJ*, *336*(7659), 1472–1474. doi:10.1136/bmj.39590.732037.47

Glidewell, L., Willis, T. A., Petty, D., Lawton, R., McEachan, R. R., Ingleson, E., … Hartley, S. (2018). To what extent can behaviour change techniques be identified within an adaptable implementation package for primary care? A prospective directed content analysis. *Implementation Science*, *13*(1), 32. doi:10.1186/s13012-017-0704-7

Hoffmann, T. C., Erueti, C., & Glasziou, P. P. (2013). Poor description of non-pharmacological interventions: Analysis of consecutive sample of randomised trials. *BMJ*, *347*, f3755. doi:10.1136/bmj.f3755

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., … Lamb, S. E. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, *348*, g1687. doi:10.1136/bmj.g1687

HSE Tobacco Control Framework Implementation group. (2013). National standard for tobacco cessation support programme. Retrieved from http://www.hse.ie/cessation

Jepson, R. G., Harris, F. M., Platt, S., & Tannahill, C. (2010). The effectiveness of interventions to change six health behaviours: A review of reviews. *BMC Public Health*, *10*(1), 538. doi:10.1186/1471-2458-10-538

Kok, G., Gottlieb, N. H., Peters, G. J. Y., Mullen, P. D., Parcel, G. S., Ruiter, R. A., … Bartholomew, L. K. (2016). A taxonomy of behaviour change methods: An intervention mapping approach. *Health Psychology Review*, *10*(3), 297–312. doi:10.1080/17437199.2015.1077155

Lohse, K. R., Pathania, A., Wegman, R., Boyd, L. A., & Lang, C. E. (2018). On the reporting of experimental and control therapies in stroke rehabilitation trials: A systematic review. *Archives of Physical Medicine and Rehabilitation*. doi:10.1016/j.apmr.2017.12.024

Lorencatto, F., West, R., Stavri, Z., & Michie, S. (2012). How well is intervention content described in published reports of smoking cessation interventions? *Nicotine & Tobacco Research*, *15*(7), 1273–1282. doi:10.1093/ntr/nts266.

McEwen, A. (2014). Standard treatment programme: A guide to behavioural support for smoking cessation. 2nd ed. *National Centre for Smoking Cessation and Training*. Retrieved from http://www.ncsct.co.uk/usr/pdf/NCSCT-standard_treatment_programme.pdf

Mhizha-Murira, J. R., Drummond, A., Klein, O. A., & dasNair, R. (2018). Reporting interventions in trials evaluating cognitive rehabilitation in people with multiple sclerosis: A systematic review. *Clinical Rehabilitation*, *32*(2), 243–254. doi:10.1177/0269215517722583

Michie, S., Hyder, N., Walia, A., & West, R. (2011). Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. *Addictive Behaviors*, *36*(4), 315–319. doi:10.1016/j.addbeh.2010.11.016

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., … Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, *46*(1), 81–95. doi:10.1007/s12160-013-9486-6

Michie, S., Wood, C. E., Johnston, M., Abraham, C., Francis, J., & Hardeman, W. (2015). Behaviour change techniques: The development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). *Health Technology Assessment*, *19*(99), 1–188. doi:10.3310/hta19990

Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., & Moher, D. (2018). Reporting randomised trials of social and psychological interventions: The CONSORT-SPI 2018 extension. *Trials*, *19*(1), 407. doi:10.1186/s13063-018-2733-1

NHS Scotland. (2017). A guide to smoking cessation in Scotland, 2010. Edinburgh, United Kingdom. Retrieved from http://www.healthscotland.scot/media/1096/a-guide-to-smoking-cessation-in-scotland-2017.pdf

NICE. (2007). Behaviour change: General approaches. Retrieved from https://www.nice.org.uk/guidance/ph6/chapter/1-public-health-need-and-practice

Oberjé, E. J., Dima, A. L., Pijnappel, F. J., Prins, J. M., & de Bruin, M. (2015). Assessing treatment-as-usual provided to control groups in adherence trials: Exploring the use of an open-ended questionnaire for identifying behaviour change techniques. *Psychology & Health*, *30*(8), 897–910. doi:10.1080/08870446.2014.1001392

Peters, G. J. Y., de Bruin, M., & Crutzen, R. (2015). Everything should be as simple as possible, but no simpler: Towards a protocol for accumulating evidence regarding the active content of health behaviour change interventions. *Health Psychology Review*, *9*(1), 1–14.

Sakzewski, L., Reedman, S., & Hoffmann, T. (2016). Do we really know what they were testing? Incomplete reporting of interventions in randomised trials of upper limb therapies in unilateral cerebral palsy. *Research in Developmental Disabilities*, *59*, 417–427. doi:10.1016/j.ridd.2016.09.018

Wagner, G. J., & Kanouse, D. E. (2003). Assessing usual care in clinical trials of adherence interventions for highly active antiretroviral therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *33*(2), 276–277.

Warner, M. M., Kelly, J. T., Reidlinger, D. P., Hoffmann, T. C., & Campbell, K. L. (2017). Reporting of telehealth-delivered dietary intervention trials in chronic disease: Systematic review. *Journal of Medical Internet Research*, *19*(12), e410. doi:10.2196/jmir.8193

World Health Organization. (2013). Global action plan for the prevention and control of noncommunicable diseases 2013–2020.

Appendix 1. Elaboration on the methods for assessing potential active content and the data collection process (a reply to a Reviewer, added as Appendix upon reviewer request)

In our methods, we make a distinction between practical intervention activities (which you would expect to read in an intervention protocol or a smoking cessation guideline for practitioners) and behaviour change techniques (which are given in a taxonomy and describe techniques that may modify behaviours or its determinants at a more abstract level). This is similar to for example the concepts of *practical applications* and *behaviour change methods* in Intervention Mapping (Eldredge et al., 2016. Planning Health Promotion Programmes: an Intervention Mapping Approach. Jossey-Bass).

When we collect published materials and unpublished materials from study authors, we typically obtain documents (e.g., articles, protocols) describing practical intervention activities. We then identify the potential active content of these interventions by examining which of these activities target any of the four behaviours (quit, abstain, adherence, engagement) and that qualify as a behaviour change technique. The approach is identical for the checklist: The checklist contains a list of practical intervention activities that we identified from multiple smoking cessation guidelines. The guideline activities included in the checklist qualified as representing active content (i.e., qualify as behaviour change technique/s and target one of the four behaviours of interest). Authors are asked to report which of these practical activities have been delivered to their comparator groups, after which we link their responses to the BCTs in the v1 taxonomy (this can be done with a syntax instead of with coders). So, both the intervention materials and the completed checklists contain the same level of information (practical intervention activities). We use this information to identify the potential active content of the smoking cessation support provided by identifying the activities that qualify as behaviour change techniques targeting one of the four behaviours. Hence, there is no checklist versus BCTTv1, nor are there any BCTs in the checklist, and the BCTs included in this study were those already defined in the BCTTv1 taxonomy.

The reasons that the checklist items describe the practical intervention activities instead of the more abstract behaviour change technique definitions from the BCTTv1, are (a) this would require individual authors to make the translation from practical intervention activities (i.e., what happened in the study) to BCTs – potentially introducing noise/error, and (b) because we think recall is better when we ask for concrete, observable activities. For the same reason of recall, we ask authors to report the activities that were delivered to the majority of their comparator group participants (i.e., the regularly recurring activities). We also asked authors to only complete the checklist if they felt confident about their responses. In a previous study (de Bruin et al., 2009, 2010) we found that approximately 65% of the authors felt confident in their ability to complete this checklist. In this study that was slightly lower (56%; note that we had

established communication with approximately 95% of the study teams). Hence, a bit less than half of the teams reported not feeling they had the information available to give meaningful responses.

We developed the checklist for assessing what the comparator groups were receiving. The reason for that is that there is a major gap in the quality and quantity of materials that authors have available on what their experimental group received versus their comparator group. Hence, solely relying on what the authors could send us in terms of intervention materials (i.e., no method bias), would result in a major discrepancy in the amount and quality of information available for experimental versus comparator interventions. The purpose of this checklist is to try to close that information gap. In this way, we see the use of the checklist as an important strength of our methodology.

For developing the items for that checklist, we require access to the materials describing the interventions (i.e., to identify the potential active practical activities to quit smoking). In this case, we used publicly available smoking cessation guidelines from various countries (e.g., UK, USA). We extracted from these guidelines the practical intervention activities that qualified as BCTs targeting any of the four behaviours. We then placed these 'potentially active' practical intervention activities in the checklist. We discussed this checklist with smoking cessation experts (professionals, researchers, and clients) to ensure that it was clear, comprehensive, and accurate. It is important to note that this checklist is not all-encompassing: it should include a good range of the most commonly provided activities in smoking cessation programmes in treatment-as-usual contexts. But it is unlikely to contain many of the potential active ingredients of the experimental smoking cessation interventions, as this would require us to first obtain all the intervention materials; identifying in these materials the potential active ingredients; including those potentially active elements as items in the checklist, and then sending that checklist back to trial authors to ask them what their intervention group received. Besides that this would create considerable delay in the project, we could also anticipate low response rates and annoyed authors as they had already provided us with all the materials necessary to infer what was happening in their interventions.

In our previous checklist study (de Bruin et al., 2009, 2010) we observed a high internal consistency reliability of the checklist and good predictive validity for behaviour and clinical outcomes. Such high reliability is unlikely if author's responses were sketchy. Also, in our current smoking cessation study, the internal consistency reliability of the checklist is high (Cronbach's alpha .92) and (sneak preview) in a manuscript under review, we again see good predictive validity of the checklist responses for smoking cessation rates in the comparator groups and in effect sizes. In other words, the checklist has been carefully and systematically developed and examined both previously and in this study, and the analyses all support that these checklist data are reliable and valid.