The impact of STRICTA and CONSORT on reporting of randomised control trials of acupuncture: a systematic methodological evaluation

Simen Svenkerud, Hugh MacPherson

Department of Health Sciences, Faculty of Science, University of York, York, UK

Correspondence to
Simen Svenkerud, Department of Health Scienses, University of York, York YO10 5DD, UK; ss1996@york.ac.uk

ABSTRACT

Background Clear and unambiguous reporting is essential for researchers and clinicians to be able to assess the quality of research. To enhance the quality of reporting, consensus-based reporting guidelines are commonly used.

Objectives To update and extend previous research by evaluating the more recent impact of STRICTA (STandards for Reporting Interventions in Controlled Trials of Acupuncture) and CONSORT (CONsolidated Standards Of Reporting Trials) guidelines on the quality of reporting of acupuncture trials.

Methods By random sampling, approximately 45 trials from each of five 2-year time periods between 1994 and 2015 were included in the study. Using scoring sheets based on the STRICTA and CONSORT checklist items (range 0 to 7 and 0 to 5, respectively), the distribution of items reported over time was investigated, with changes shown using scatterplots. The primary analysis used a before-and-after t-test to compare time periods. A meta-analysis investigated whether or not trials published in journals that endorsed STRICTA were associated with better reporting.

Results The study included 207 trials. Improved reporting of items over time was observed, as represented by changes in the scatterplot slope and intercept. The mean STRICTA score increased from 4.27 in the 1994–1995 period to 5.53 in 2014–2015, an 18% improvement. The mean CONSORT score rose from 1.01 in the 1994–1995 period to 3.32 in 2014–2015, an increment of 46%. There was proportionately lower reporting for items related to practitioner background (STRICTA) and for randomisation implementation and allocation concealment (CONSORT). Trials published in journals that endorsed STRICTA had statistically significantly superior reporting of both STRICTA and CONSORT items overall.

Conclusion This study has provided evidence of an improvement in reporting of STRICTA and CONSORT items over the time period from 1994 to 2015. Journals that endorse STRICTA have a better record in terms of reporting quality. Some evidence suggests that the publication of STRICTA has had a positive impact on reporting quality.

INTRODUCTION

When Begg *et al* published the first edition of the 'CONsolidated Standards Of Reporting Trials' (CONSORT) guidelines, their aim was to improve reporting quality in randomised control trials by providing a checklist for authors to follow when writing up their articles for publication.[1] CONSORT has since been updated in 2001 and 2010.[2][3] Today CONSORT is widely accepted as standard for reporting of clinical trials. A series of studies have been conducted to assess the impact of CONSORT in general medical fields, which show that CONSORT has positively impacted reporting quality and research quality.[4–8]

The 'STandards for Reporting Interventions in Controlled Trials of Acupuncture' (STRICTA) guidelines were first published in 2001[9] and a revised version was published in 2010.[10] The aim of STRICTA is to improve reporting of the interventions within clinical trials of acupuncture, thereby extending the CONSORT guidelines. An early review that evaluated the impact of STRICTA on acupuncture trial reporting concluded that there had been a statistically significant increase in the reporting of CONSORT items between 1996 and 2005, but over the same period there was no statistical evidence for an improvement in STRICTA reporting.[11]

The aim of this specific study was to update and expand upon previous research by Prady *et al*[11] by investigating the impact of STRICTA and CONSORT on the quality of reporting of acupuncture research. Our objective was to compile a larger dataset over a longer time period and conduct a higher level statistical analysis in order to improve and extend previous research.
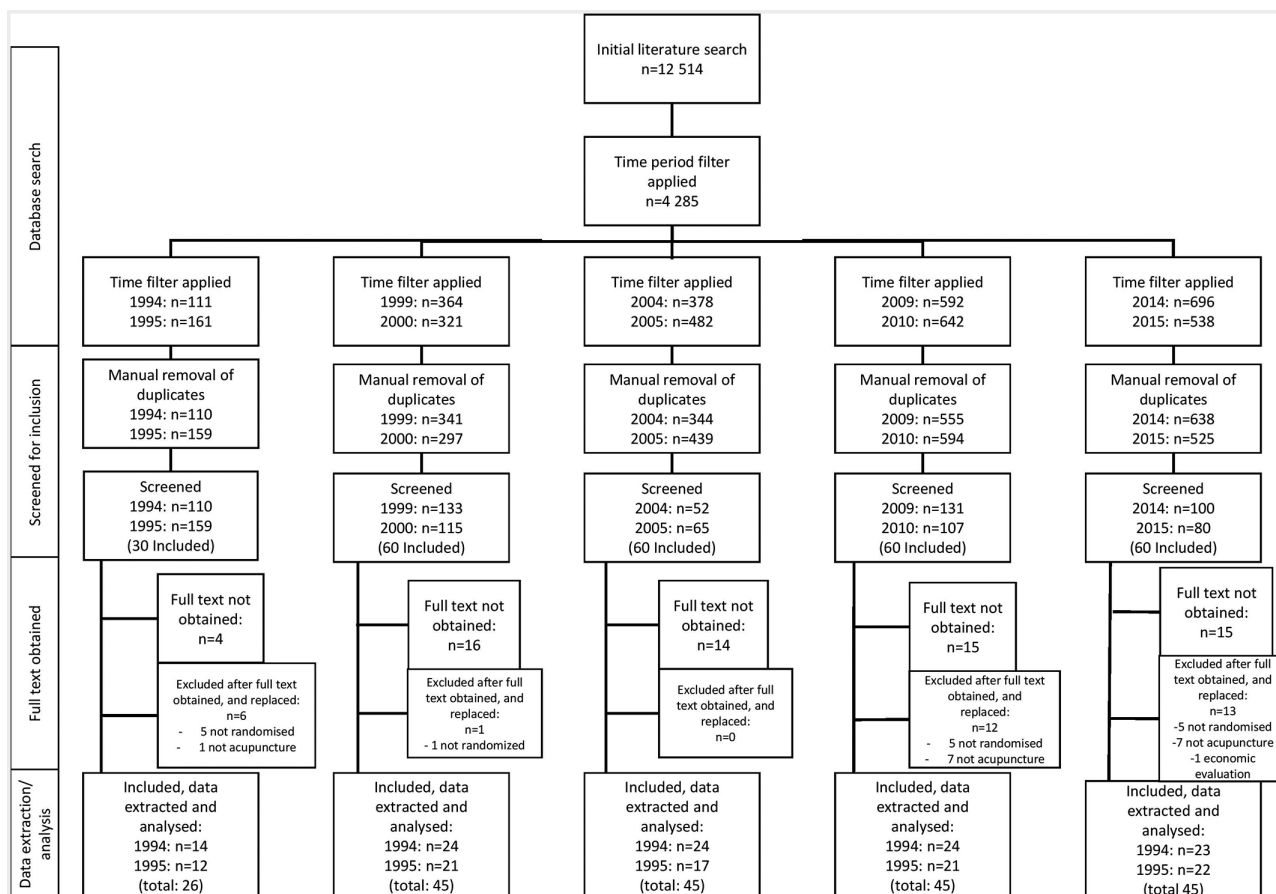
Figure 1 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing the flow of articles through the study.

METHODS

The methodological basis of the study was a systematic review, however the area of interest was the methodology of the published articles rather than a specific condition. The gold standard for a systematic review was adjusted accordingly to allow for this focus. This study can be seen as an expansion of previous research by Prady et al.[11]

Time periods

As a baseline, we chose to collect data over the 2-year time period 1994–1995 as this was before the publication of CONSORT and STRICTA. We also collected data from a further four distinct time periods, namely 1999–2000, 2004–2005, 2009–2010, and 2014–2015. These periods can be interpreted in the context of the publication dates of both STRICTA (in 2001 and 2010) and CONSORT (in 1996, 2001 and 2010). For comparability between this study and the one by Prady et al,[11] we used the same inclusion/exclusion criteria.

Screening process and study selection

Our initial search found 12 514 potential acupuncture trials (figure 1). After manually removing duplicates, and applying time filters, we then used random selection (with the aim of achieving a pre-defined sample of 45 randomised control trials in each time period) and screening for inclusion and exclusion criteria. In

order to prevent a high attrition rate resulting from the screening process, as was observed by Prady et al,[11] it was decided that we would continue randomly selecting and screening articles until the full sample size had been achieved. Additionally, in the event that trials had to be excluded after the full text had been obtained, these were replaced using the same random allocation procedure as applied during initial screening. This procedure ensured a fully random final sample.

Sample size calculations

We assumed that the initial differences in STRICTA scores between time periods and related standard deviations (SDs) found by Prady et al[11] (13.5 and 22.4, respectively) represented a reasonable estimate for all time periods. For the group comparisons, at a significance level of 0.05 with 80% statistical power, we estimated that we would require a sample size per time period of n=45 papers, that is, n=225 papers overall. Due to limited resources available to the project, and the high level of comparability, we requested and received permission to include data from Prady et al[11] in our analysis.

Primary outcomes

Outcome measures for this study were the item scores using the STRICTA and CONSORT rating

Table 1 STRICTA item distribution

STRICTA items	1994–1995 (%)	1999–2000 (%)	2004–2005 (%)	2009–2010 (%)	2014–2015 (%)
1.1) Is there a statement on the style of acupuncture?	53.9	68.9	69.6	75.6	**86.7**
1.2) Is there a rationale for what they did? (ie, selection of points)?	53.9	57.8	**78.3**	62.2	77.8
1.3) Are sources quoted to justify the rationale for the selection?	38.5	42.2	50.0	42.2	**57.8**
2.1) Is there a description of the location of points?	84.6	84.4	91.3	93.3	**95.6**
2.2) Is it explicit that the points are unilateral or bilateral?	50.0	48.9	**73.9**	64.4	66.7
2.3) Are the numbers of needles inserted/points used reported?	53.9	53.3	71.7	80.0	**82.2**
2.4) Are the depths of insertion reported?	50.0	35.6	41.3	**60.0**	53.3
2.5) If the study protocol requires that a response be elicited, are the responses described?	73.1 Not specified 26.9 Specified	57.8 Not specified 42.2 Specified	43.5 Not specified 56.5 Specified	33.3 Not specified **66.7** Specified	44.4 Not specified 55.6 Specified
2.6) Is the needle stimulation described?	50.0	71.1	58.7	**82.2**	73.3
2.7) Is needle retention time recorded?	50.0	48.9	47.8	**82.2**	68.9
2.8) Is the needle type reported?	11.5	28.9	45.7	**73.3**	68.9
3.1) Is number of treatments reported?	69.2	66.7	56.5	86.7	**97.8**
3.2) Is the frequency of treatment reported?	69.2	71.1	67.4	84.4	**95.6**
4) Are co-interventions reported?	15.4 Yes 84.6 No	26.7 Yes 73.3 No	10.9 Yes 89.9 No	53.3 Yes 46.7 No	55.6 Yes 44.4 No
5.1) Is the duration of relevant training reported?	0	2.2	10.9	24.4	**35.6**
5.2) Is length of clinical experience reported?	3.9	4.4	15.2	22.2	**31.1**
5.3) Is any expertise of the practitioner in the specific condition reported?	11.5	0	10.9	31.1	**33.3**
6.1) Is the intended effect of the control interventions and its appropriateness to the research question described?	**80.8**	66.7	76.1	68.9	80.0
6.2) Are the explanations that were given to participants of treatment and control interventions provided?	3.9	11.1	15.2	35.6	**57.8**
6.3) Type of control used	53.8 Sham/needle control 46.2 Other control	53.3 Sham/needle control 46.7 Other control	47.8 Sham/needle control 52.2 Other control	35.6 Sham/needle control 64.4 Other control	44.5 Sham/needle control 55.6 Other control
For sham or needle control					
6.4) Is there a description of the location of points used?	50.0	80.0	75.0	86.2	**92.0**
6.5) Is it explicit that the points are unilateral or bilateral?	50.0	56.0	62.0	58.6	**76.0**
6.6) Are the numbers of needles inserted/placed reported?	42.9	64.0	58.3	79.3	**88.0**
6.7) Are the depths of insertion reported?	57.1	68.0	62.5	**69.0**	60.0
6.8) If the study protocol requires that a response be elicited, are the responses described?	85.7 Not specified 14.3 Specified	80.0 Not specified 20.0 Specified	75.0 Not specified 25.0 Specified	41.4 Not specified 58.6 Specified	48.0 Not specified 52.0 Specified
6.9) Is the needle stimulation described?	78.6	68.0	70.8	**82.8**	60.0
6.10) Is needle retention time recorded?	64.3	56.0	45.8	75.9	**76.0**
6.11) Is the needle type reported?	28.6	40.0	45.8	**82.8**	72.0
6.12) Is number of treatments reported?	78.6	68.0	62.5	82.8	**96.0**
6.13) Is the frequency of treatment reported?	78.6	76.0	66.7	79.3	**92.0**
For other controls					
6.14) Is there a precise description of the intervention, including treatment regimen?	58.3	40.0	59.1	81.3	**95.0**
6.15) Are the sources that justify the choice of control quoted?	50.0	42.2	43.5	51.1	**66.7**

scales based on the checklists used by Prady et al.[11] The primary outcomes for the study were the STRICTA and CONSORT checklist item scores over time for acupuncture trials as an indicator of the impact of publication and revision of the STRICTA and CONSORT guidelines. Each item on the STRICTA and CONSORT checklists was re-framed as a question or series of sub-questions (see tables 1 and 2 for exact

Table 2 CONSORT item distribution

CONSORT items	1994–1995 (%)	1999–2000 (%)	2004–2005 (%)	2009–2010 (%)	2014–2015 (%)
1) Is there a description of the method used to generate the random allocation sequence?	7.7	22.2	50.0	62.2	**77.8**
2.1) Is there a description of the method used to implement the random allocation sequence?	0	26.8	41.3	42.2	**53.3**
2.2) Is there clarification about whether the sequence was concealed until interventions were assigned?	3.9	20.0	43.5	48.9	**51.1**
3.1) Is there a statement about blinding of participants or open trial?	38.5	40.0	50.0	57.8	**75.6**
3.2) Is there a statement about blinding of practitioner or open trial?	38.5	35.6	50.0	53.3	**77.8**
3.3) Is there a statement about blinding outcome assessor?	26.9	31.1	41.3	53.3	**75.6**
4) Are baseline demographic and clinical characteristics of each group presented?	53.6	68.9	**89.1**	84.4	88.9
5) Is there a statement of whether or not the analysis was by 'intention-to-treat'?	3.9	20.0	32.6	24.4	**37.8**

formulations). Each item was given equal weight, such that the item's question(s) contributed a score between 0 and 1 that, when totalled, was within the range 0 to 7 and 0 to 5 points for STRICTA and CONSORT, respectively.

Statistical analysis

We used the statistical analysis software package SPSS for Windows (2013; IBM-Corp, Armonk, NY, USA) to identify a random selection of articles. Using Stata 14 (2015; StataCorp, College Station, TX, USA), we performed t-tests to compare total STRICTA scores and combined STRICTA+CONSORT scores between time periods. The data for the CONSORT scores were found not to meet the requirements for a parametric analysis and thus were analysed using the Wilcoxon rank sum test instead. We examined for a statistically significant change between adjacent time periods, as well as between the period before first publication and the most recent time period.

Secondary analyses included the distribution of individual STRICTA and CONSORT items (using percentages) over time, a simple regression of item scores against time, and a meta-analysis of the effect of endorsing STRICTA and/or CONSORT on adherence to item reporting.

RESULTS
Study characteristics

After removal of duplicates, 4285 articles matched the search criteria. Through random sampling, 1052 articles were screened for inclusion, and 207 articles were included in the final analysis. See figure 1 for our PRISMA flowchart, and table 3 for relevant study characteristics. Note should be made that for 1994–1995 we only achieved a sample size of 26 papers. The

reason for this was that our search only identified 26 studies eligible for inclusion within this time period.

Overall STRICTA and CONSORT total scores over time

We observed a rise in the mean total STRICTA score (see table 2) from 4.27 in the 1994–1995 period to 5.53 in 2014–2015, which represents a 1.25 point improvement (18%). The mean CONSORT score rose from 1.01 in the 1994–1995 period to 3.32 in 2014–2015, an increment of 2.31 points (46%). The overall score when STRICTA and CONSORT were combined as a single value rose from 5.29 in the 1994–1995 period to 8.85 in 2014–2015, an improvement of 3.56 points (30%).

When these data were presented graphically, we saw an overall trend towards increased quality of reporting following the publication of STRICTA (figure 2A). The red line, as a simple regression line, shows the trend in the data across all time periods, ignoring the gaps between them. The slope was relatively unchanged over the time period before the publication of STRICTA in 2001. By the 2004–2005 time period there was evidence of increased reporting quality according to the STRICTA scores, a trend that continued through to 2014–2015. A trend towards improvement was observed in both the CONSORT scores (figure 2B) and the combined STRICTA and CONSORT scores (figure 2C).

Reporting of individual STRICTA and CONSORT items over time

Looking at the overall trends for STRICTA, we saw that, for all items except 2, 5, and 18, the highest reporting scores were in the more recent periods of either 2009–2010 or 2014–2015 (table 2). In 2014–2015 the best-reported items, adhered to in over 90% of all included articles, were frequency and number of

Table 3 Characteristics of included studies	1994–1995	1999–2000	2004–2005	2009–2010	2014–2015	Total
Number of trials included	26	45	46	45	45	20
Mean STRICTA score (min–max)	4.27 (2.21–6.24)	4.29 (2.14–6.33)	4.53 (2.715–7)	5.06 (2.48–7)	5.53 (3.48–7)	4.78 (2.14–7)
SD of STRICTA score	0.85	0.88	0.90	0.96	0.90	1.02
Mean CONSORT score (min–max)	1.01 (0–3)	1.70 (0–5)	2.61 (0–5)	2.72 (0–5)	3.32 (1–5)	2.39 (0–5)
SD of CONSORT score	0.84	1.47	1.39	1.51	1.41	1.57
Mean STRICTA+CONSORT score (min–max)	5.29 (3.21–7.6)	5.99 (2.14–11)	7.14 (4.17–11.5)	7.78 (3.48–12)	8.85 (4.48–12)	7.17 (2.14–12)
SD of STRICTA+CONSORT score	1.21	1.96	1.81	2.05	1.99	2.22
Trial setting: inpatient/outpatient (%)	2/24 (7.7/92.3)	7/38 (15.6/84.4)	11/35 (23.9/76.1)	10/35 (22.2/77.8)	11/34 (24.4/75.6)	41/166 (19.8/80.2)
Trial population age: paediatric/adult (%)	3/23 (11.5/88.5)	1/44 (2.2/97.8)	0/46 (0/100)	7/38 (15.6/84.4)	0/45 (0/100)	11/196 (5.3/94.7)
Endorsing STRICTA: yes/no (%)	4/22 (15.4/84.6)	2/43 (4.4/95.6)	3/43 (6.5/93.5)	24/21 (53.3/46.7)	21/24 (46.7/53.3)	54/153 (26.1/73.9)
Endorsing CONSORT: yes/no (%)	10/16 (38.5–61.5)	22/23 (48.9/51.1)	24/22 (52.2/47.8)	30/15 (66.7/33.3)	29/16 (64.4/35.6)	115/92 (55.6/44.4)
Average number of authors (min–max)	3.56 (1–8)	4.31 (1–9)	5.41 (1–13)	6.11 (2–15)	6.57 (1–17)	5.37 (1–17)
Average number of participants randomised (min–max)	80.80 (8–568)	70.06 (12–238)	100.13 (10–386)	84.91 (10–638)	95.47 (20–517)	86.80 (8–683)

treatments as well as the description of point locations (items 4, 12, 13, 21, 29 and 30). The area with the weakest reporting was practitioner information (items 15–17).

There was a relatively steady increase in the reporting of most CONSORT items throughout most of the time periods (table 3). Although the majority of CONSORT items were reported in over 75% of included studies in 2014–2015, the poorest reporting related to information on allocation concealment and implementation—which was only reported in 51.1% (concealment; item 2.2) and 53.3% (implementation; item 2.1)—and related to 'intention-to-treat' (37.78%; item 5).

Before-and-after analysis of STRICTA and CONSORT

We observed statistically significant results for the change in STRICTA scores (table 4) for all time periods compared, with the exception of 1994–1995 versus 1999–2000 (P=0.91) and 1999–2000 versus 2004–2005 (P=0.2). Looking at the overall change from the earliest period (1994–1995) to the most recent time period (2014–2015) we found an overall statistically significant (P<0.001) improvement of 1.25 points (95% CI 1.72 to 0.84).

For CONSORT, Wilcoxon rank sum tests only provided P values (table 5). Statistically significant results were found for 1999–2000 versus 2004–2005 (P<0.001), while the remaining adjacent time periods showed no statistically significant differences. A statistically significant (P<0.001) improvement was found when comparing baseline (1994–1995) against the data from 2004 to 2005 and 2014–2015. For combined STRICTA and CONSORT scores (table 6), we found an overall change of 3.55 (95% CI 4.32 to 2.78).

Meta-analysis of the effect of STRICTA endorsement

When comparing the relative risk of reporting quality in journals that endorsed the use of STRICTA with those that did not (figure 3), we found that overall there was more complete reporting in journals endorsing STRICTA. The overall relative risk of reporting completeness was 1.27 (95% CI 1.18 to 1.36). Looking at the individual items, the point estimates for a few items showed that they were better reported in non-endorsing journals (items 1.2, 1.3, 2.1, 6.3). By contrast, several items were significantly better reported in journals endorsing STRICTA (items 2.8, 3.1, 4, 6.12, 6.13, and 6.15).

As observed for the STRICTA items, figure 4 shows a similar trend, in that CONSORT items appeared to be more completely reported in journals that endorsed the use of STRICTA compared with non-endorsing journals. The overall estimate of reporting completeness was 1.36 (CI 1.25 to 1.55). CIs of items 1, 3.2, 3.3 showed that they were significantly better reported in STRICTA-endorsing journals.
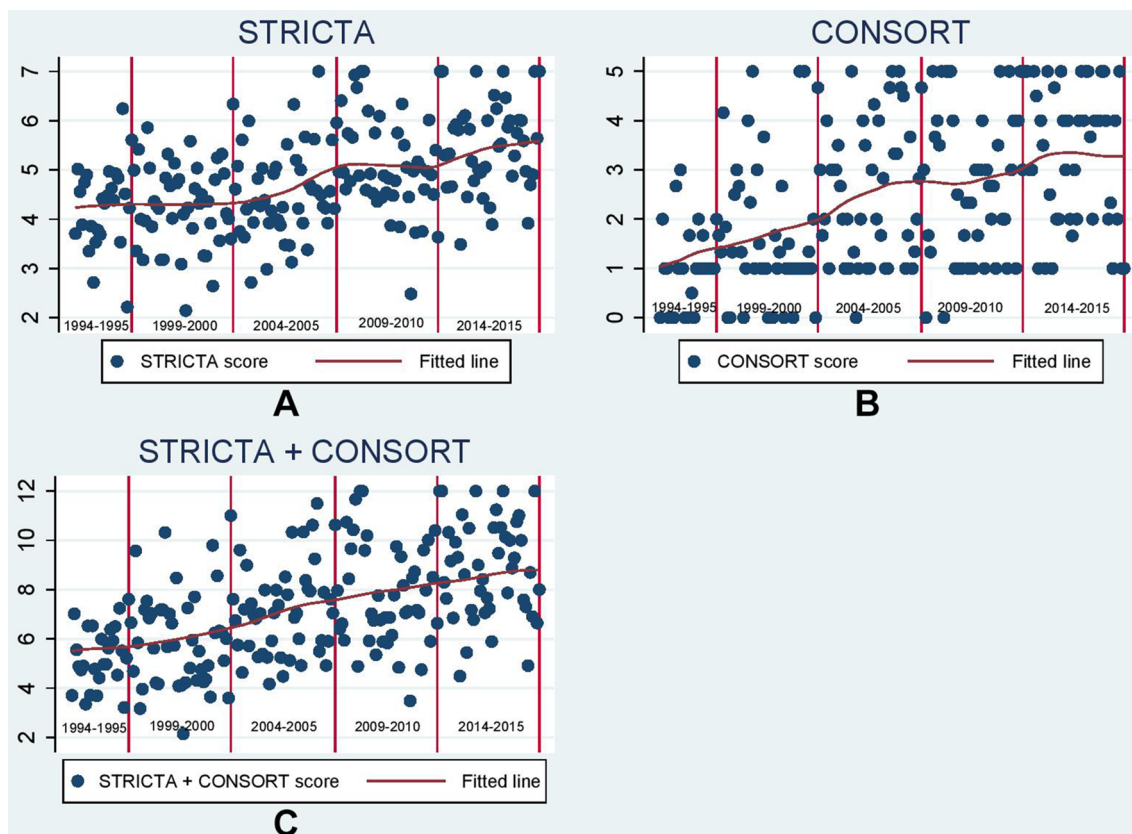
Figure 2 Distribution of item reporting scores over time. All plots show an increase in reporting quality over time. (A) STRICTA item reporting score. (B) CONSORT item reporting score. (C) combined STRICTA+CONSORT item reporting score. Solid red line=fitted least squares for aid of visualisation of data.

DISCUSSION
Principal findings
We found statistical evidence for an improvement of reporting of both STRICTA and CONSORT checklist items over time. The meta-analysis provided evidence that endorsement of STRICTA by journals is associated with improved reporting not only of STRICTA items but also of CONSORT items. Our evidence suggests that STRICTA and its subsequent update has had an impact on reporting quality.

With respect to individual item reporting, we found that the highest reported STRICTA items were frequency and duration of treatment, as well as point descriptions, which were all better reported in later years. The weakest reported items were related to practitioner information, which displayed little change from baseline. The lowest reported CONSORT items were related to randomisation implementation, allocation concealment and intention-to-treat.

Our findings in context
We note that our study provides fresh evidence that updates the earlier study by Prady et al,[11] who found little evidence to support the claim that the reporting of STRICTA items had improved by 2005, some 4 years after STRICTA was first published. A study published in 2011 investigated the reporting quality of trials using a tool they developed on the basis of STRICTA and CONSORT, named the Oregon-STRICTA-CONSORT-Index (OSCI).[12] Possibly due to the larger sample size and extended time period, they observed a statistically significant rise in the STRICTA score from 1997 to 2007.[12] Our study,

Table 4 Before-and-after t-test results for STRICTA scores

STRICTA time period	t-value	df	P value	Mean difference	SE	95% CI
1994–1995 vs 1999–2000	0.10	69	0.91	0.02	0.21	0.45 to −0.40
1999–2000 vs 2004–2005	1.29	89	0.20	0.24	0.18	0.61 to −0.13
2004–2005 vs 2009–2010	2.71	89	<0.001	0.53	0.19	0.92 to 0.14
2009–2010 vs 2014–2015	2.33	87.64	0.02	0.46	0.19	0.85 to 0.06
1999–2000 vs 2014–2015	6.55	88	<0.001	1.23	0.18	1.60 to 0.86
1994–1995 vs 2014–2015	5.79	68	<0.001	1.25	0.22	1.72 to 0.84

Table 5 Wilcoxon rank sum test results for CONSORT scores

CONSORT time period	z value	P value
1994–1995 vs 1999–2000	1.89	0.05
1999–2000 vs 2004–2005	3.06	<0.001
2004–2005 vs 2009–2010	0.36	0.71
2009–2010 vs 2014–2015	1.86	0.06
1994–1995 vs 2004–2005	4.65	<0.001
1999–2000 vs 2009–2010	3.13	<0.001
1994–1995 vs 2014–2015	5.70	<0.001

which is an update and extension of Prady et al[11] that includes subsequent time periods to 2015, larger sample size and additional analyses, supports the case that it might take between 5 and 10 years for reporting guidelines to become more widely adopted within the research community.

Studies on the impact of STRICTA and CONSORT on acupuncture trials published in the Chinese and Korean languages[13 14] have shown that there has been an increase in reporting quality over time. A review of the impact of STRICTA found an increase in the citations of STRICTA over time, with publication date being a significant predictor of STRICTA citation.[15] Kim et al[16] investigated the level of reporting of selected STRICTA items in systematic Cochrane reviews compared with reporting within the constituent randomised controlled trials (RCTs). They found that the level of reporting was statistically significantly lower in the systematic reviews, and argued for better reporting quality within such reviews.

Strength and limitations

The primary strength of our study lies in the extended time span as well as a large sample size. This allowed us to investigate and observe a broader picture compared with previous studies. Furthermore, we used assessment tools that have previously been tested, although not statistically validated. Limitations of the study include the fact that the STRICTA and CONSORT checklists were never intended to be used as rating scales of research reporting quality.[17] The original intent was to provide guidance on what items should be reported when writing up RCTs for publication. It is further debatable as to the value of the weighting chosen in this study, which was

based on that used by Prady et al,[11] whereby individual checklist items were all equally weighted. We suggest that some items are clearly more important than others. Likewise Hammerschlag et al[12] also used an equal weighting system.

A further weakness includes the fact that data extraction was performed by only one investigator, due to limited resources. Language restrictions were applied, such that all included studies were published in English. There is, however, little reason to believe that this would have introduced excessive systematic bias, as similar reviews performed on non-English studies[13 14 18] are consistent with our own results. This study did not assess the impact of unknown potential confounding factors on the question of causality, which could have led to quantification of the extent by which STRICTA has influenced the improvement in the quality of reporting.

Implications for research and practice

While we endorse the practice of regular revisions of reporting guidelines in order that they are adapted to the changing environment of acupuncture research, Liu et al[19] have suggested that it might be time to consider alternative routes to improving reporting quality. They reported limited adherence to reporting of STRICTA items among journals that have endorsed STRICTA. One possible way forward would be to better support journals and editors in order to enhance reporting of items, with editors engaging more specifically with authors on their reporting of STRICTA items at the submission stage.

In the section on strengths and limitations above, we discussed the issue that STRICTA has not been validated as a quality scoring tool, regardless of the fact that several studies have used it in that way. The development of an appropriately weighted and properly validated scoring tool to assess the quality of reporting of acupuncture research is needed.

During the conduct of this study, we found that the use of before-and-after tests to assess changes in levels of reporting presents a series of challenges. These challenges include lack of control over confounding factors, natural changes in reporting trends, and causality of impact. From this, we recommend future research to adopt an interrupted

Table 6 Before and after t-test results for STRICTA and CONSORT

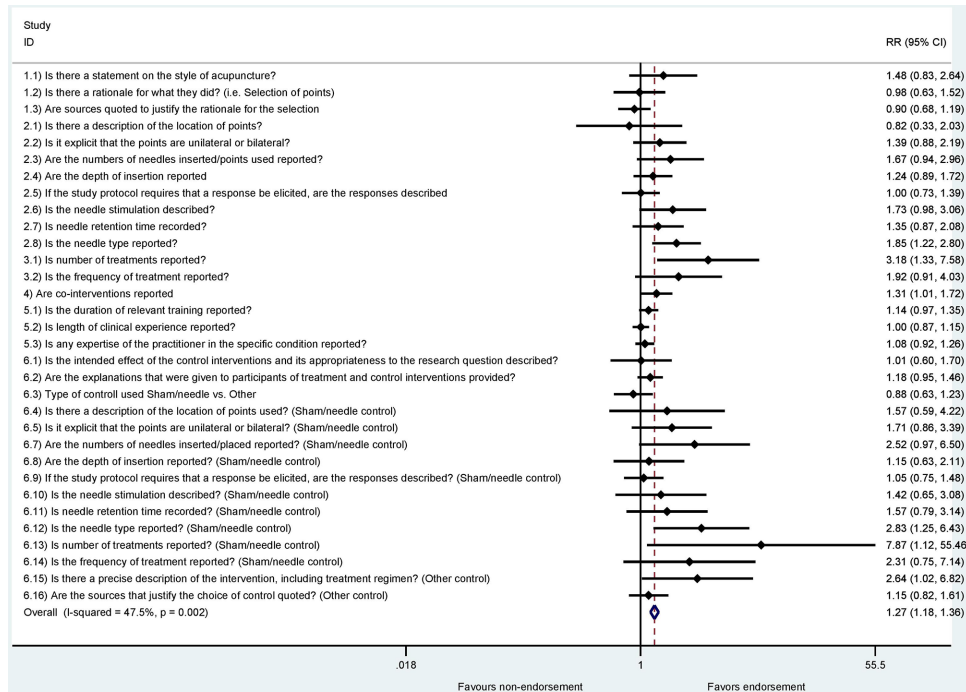| Overall time period	t-value	df	P value	Mean difference	SE	95% CI
1994–1995 vs 1999–2000	1.80	66.74	0.07	0.69	0.38	1.45 to −0.07
1999–2000 vs 2004–2005	2.91	89	<0.001	1.15	0.39	1.94 to 0.36
2004–2005 vs 2009–2010	1.58	89	0.11	0.64	0.40	1.44 to −0.16
2009–2010 vs 2014–2015	2.50	88	0.01	1.06	0.42	1.91 to 0.22
1999–2000 vs 2014–2015	6.85	88	<0.001	2.86	0.41	3.69 to 2.03
1994–1995 vs 2014–2015	9.16	67.04	<0.001	3.55	0.38	4.32 to 2.78

Figure 3 Meta-analysis of the effect of endorsement of STRICTA. The pooled estimate of likelihood of an item being reported clearly favours the paper being published in a STRICTA endorsing journal. This indicates that papers published in STRICTA endorsing journals are more likely to have higher reporting quality than papers reported in non-endorsing journals.

time series design (ITSA) to address these challenges. Computation using ITSA requires continuous data, thereby avoiding any null periods, as was the case in the study reported here. This approach would require a larger sample size of trials to be evaluated.

CONCLUSION
We found a statistically significant improvement in reporting quality of both STRICTA and CONSORT

checklist items from 1994 to 2015. Journal endorsement of STRICTA is associated with better reporting. By combining the data across analyses, we provided some evidence to suggest that the publication of STRICTA has had a positive impact on reporting quality. We are cautious regarding this causal relationship, due to the possible influence of confounding factors. Future research on the impact of reporting guidelines would benefit from using an ITSA in order
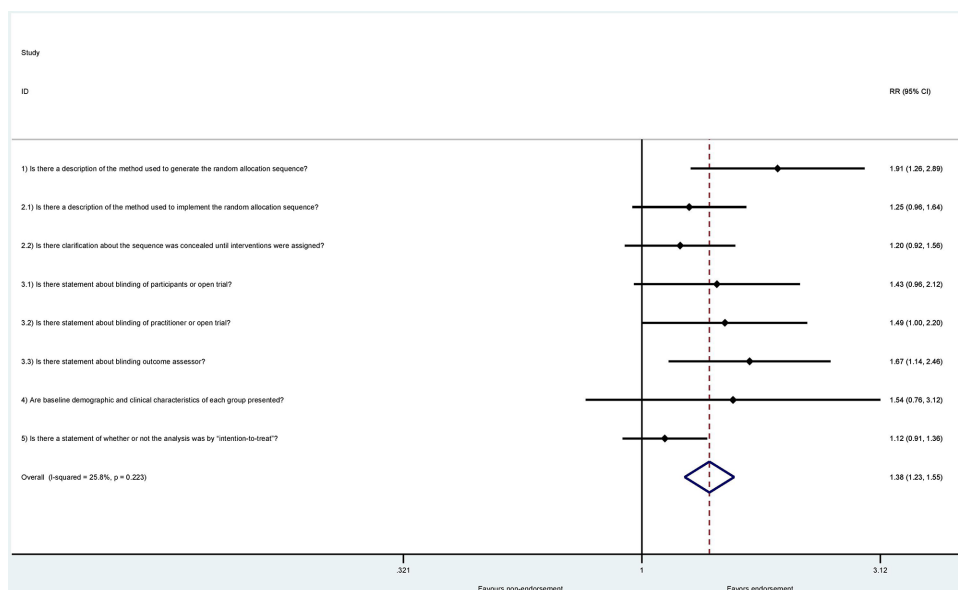

Figure 4 Meta-analysis on effect of endorsement of STRICTA on CONSORT reporting. The pooled likelihood of a CONSORT item being reported favours STRICTA endorsing journals; although not as clear as for STRICTA items it remains a statistically significant difference.

to be able to explore causality. We endorse the practice of regular revisions of reporting guidelines in order to adapt to the ever changing environment of acupuncture research.

REFERENCES

1 Begg C, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637–9.

2 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med Res Methodol* 2001;1.

3 Schulz KF, Altman DG, Moher D, *et al*. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.

4 Devereaux PJ, Manns BJ, Ghali WA, *et al*. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Control Clin Trials* 2002;23:380–12,.

5 Han C, Kwak KP, Marks DM, *et al*. The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. *Contemp Clin Trials* 2009;30:116–22.

6 Mills EJ, Wu P, Gagnier J, *et al*. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. *Contemp Clin Trials* 2005;26:480–7.

7 Mills E, Wu P, Gagnier J, *et al*. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *J Clin Epidemiol* 2005;58:662–7.

8 Turner L, Shamseer L, Altman DG, *et al*. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012;1:60.

9 MacPherson H, White A, Cummings M, *et al*. Standards for reporting interventions in controlled trials of acupuncture: the STRICTA recommendations. *Complement Ther Med* 2001;9:246–9.

10 MacPherson H, Altman DG, Hammerschlag R, *et al*. Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. *PLoS Med* 2010;7:e1000261.

11 Prady SL, Richmond SJ, Morton VM, *et al*. A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS One* 2008;3:e1577.

12 Hammerschlag R, Milley R, Colbert A, *et al*. Randomized controlled trials of acupuncture (1997-2007): an assessment of reporting quality with a CONSORT- and STRICTA-based instrument. *Evid Based Complement Alternat Med* 2011;2011:1–25.

13 Kim KH, Kang JW, Lee MS, *et al*. Assessment of the quality of reporting in randomised controlled trials of acupuncture in the Korean literature using the CONSORT statement and STRICTA guidelines. *BMJ Open* 2014;4:e005068.

14 Ma B, Chen ZM, Xu JK, *et al*. Do the CONSORT and STRICTA checklists improve the reporting quality of acupuncture and moxibustion randomized controlled trials published in Chinese journals? A systematic review and analysis of trends. *PLoS One* 2016;11:e0147244.

15 Liu L, Skinner M, McDonough SM, *et al*. STRICTA: is it time to do more? *BMC Complement Altern Med* 2015;15:190.

16 Kim KH, Kang JW, Lee MS, *et al*. Assessment of the quality of reporting for treatment components in Cochrane reviews of acupuncture. *BMJ Open* 2014;4:e004136.

17 MacPherson H, Jobst KA. Improving the reporting of interventions in clinical trials of acupuncture: the updated and revised STRICTA. *J Altern Complement Med* 2010;16:929–30.

18 Lu LM, He J, Zeng JC, *et al*. Impact evaluation of CONSORT and STRICTA guidelines on reporting quality for randomized controlled trials of acupuncture conducted in China. *Chin J Integr Med* 2017;23:10–17.

19 Liu L, Skinner M, McDonough SM, *et al*. STRICTA: is it time to do more? *BMC Complement Altern Med* 2015;15.