

# Evaluation of reporting quality of the 2010 and 2012 National Surgical Congress oral presentations by CONSORT, STROBE and Timmer criteria

Mustafa Hasbahçeci<sup>1</sup>, Fatih Başak<sup>2</sup>, Ömer Uysal<sup>3</sup>

## ABSTRACT

**Objective:** This study aimed to evaluate the abstracts of oral presentations that were accepted to the National Surgical Congress by CONSORT, STROBE and Timmer criteria and to recommend development of a national abstract assessment system.

**Material and Methods:** Presentation scores were calculated for oral presentations that have been accepted to the 2010 and 2012 National Surgical Congresses and have been included in the digital congress abstract booklets by two independent reviewers who were blinded to information regarding both the author and the institution. The CONSORT and Timmer criteria were used for randomized controlled trials, and for observational studies the STROBE and Timmer criteria were used. The presentation score that was obtained by three different evaluation systems was accepted as the main variable. The score changes according to the two congresses, the influence of the reviewers on the presentation scores, and compatibility between the two reviewers were evaluated. Comparisons regarding study types and total presentation number were made by using the chi-square test, the compatibility between the total score of the presentations were made by the Mann-Whitney U test and the compatibility between the reviewers were evaluated by the Wilcoxon signed ranks test.

**Results:** There was no difference between the two Congresses in terms of study type distribution and total number of accepted presentations ( $p=0.844$ ). The total scores of randomized controlled trials and observational studies from the 2010 and 2012 National Surgical Congresses that were evaluated by two independent reviewers with different assessment tools did not show any significant difference ( $p>0.05$ ). A significant difference was observed between the reviewers in their evaluation by CONSORT, STROBE and Timmer criteria ( $p<0.05$ ).

**Conclusion:** Implementation of standard criteria for the evaluation of abstracts that are sent to congresses is important in terms of presentation reporting quality. The existing criteria should be revised according to national factors, in order to reduce the significant differences between reviewers. It is believed that discussions on a new evaluation system will be beneficial in terms of the development of a national assessment system.

**Key Words:** Congress, abstract, oral presentation, reporting quality

## INTRODUCTION

Congresses where scientific studies are shared as oral or poster presentations are important occasions. The majority of presentations are included only as abstracts in congress proceedings. Considering that less than 50% of congress presentations are published in the literature, it is obvious that the majority of presentations are not converted into publications (1, 2). Similarly, it was found that in Turkey the rate of publications in international journals out of the general surgery congress presentations between 1996-2004 was 5.7% (3). Therefore, congress abstracts are the only source available for a large portion of scientific studies (4, 5).

A scientific study abstract should be sufficient and qualified enough to allow screening for the subject in order to fulfill the reader's interests and needs. Some abstracts that are presented at congresses regarded important in their respective fields are taken into account in determining clinical practice (6). The limitation of abstracts with word quantity, reporting of studies in the summary form, and the publication process result in serious problems in terms of reporting quality (4, 5). Although abstracts are peer reviewed, only considering the subject of the congress or presentation, the classification of presentations as oral or poster presentations, and insufficient details concerning the methodology of the study within the abstract result in biased assessments (4, 6). It is believed that an established and well-written summary will provide sufficient information regarding the validity and feasibility of study findings (5). Considering all these factors, contents and reporting quality of congress abstracts are very important.

In recent years, within the framework of implementation of evidence-based medicine, various criteria were introduced on the contents and reporting quality of congress abstracts. The CONSORT (Consolidated Standards of Reporting Trials) evaluation system, which was first proposed in 1996 in relation to the publication process of randomized controlled clinical trials, has also been used for the evaluation of congress abstracts since 2008 (5-7). Similarly, the STROBE (Strengthening the Reporting of observational Studies) evaluation system was developed for case-control and observational studies and its fourth edi-

<sup>1</sup>Department of General Surgery, Bezmialem Vakif University Faculty of Medicine, Istanbul, Turkey

<sup>2</sup>Clinic of General Surgery, Ümraniye Training and Research Hospital, Istanbul, Turkey

<sup>3</sup>Department of Biostatistics and Medicine Informatics, Bezmialem Vakif University Faculty of Medicine, Istanbul, Turkey

### Address for Correspondence Mustafa Hasbahçeci

Department of General Surgery, Bezmialem Vakif University Faculty of Medicine, Istanbul, Turkey  
Phone: +90 533 356 01 32  
e-mail: hasbahceci@yahoo.com

Received: 19.03.2014

Accepted: 03.04.2014

©Copyright 2014  
by Turkish Surgical Association

Available online at  
[www.ulusalcerahidergisi.org](http://www.ulusalcerahidergisi.org)

tion was published in 2007 to be used for the assessment of congress abstracts (8, 9). The CONSORT criteria and STROBE criteria that were developed for the evaluation of congress presentation abstracts consists 17 and 11 parameters, respectively, and the evaluation is carried out with a similar scoring by a checklist (10).

Timmer et al. (6) have introduced an assessment tool for congress abstracts in 2003. This system, which was developed because CONSORT and STROBE evaluation systems could not be used in both observational and randomized controlled studies, contains 19 parameters (4). The system that was developed by Timmer has a significant feature of being applicable to any type of study including meta-analysis, randomized controlled trials and observational studies, case series and experimental studies (6).

In recent years, the concept of selected papers or best papers has been introduced in congresses held in Turkey. What type of evaluation system will be used during this evaluation is usually not disclosed. Therefore, what criteria should be taken into account in the evaluation process of congress presentation reporting quality and compliance of these systems with the conditions of Turkey has not been yet investigated.

This study aimed to evaluate the abstracts of oral presentations that were accepted to the 17<sup>th</sup> and 18<sup>th</sup> National Surgical Congress, which is a national meeting in the field of general surgery, by CONSORT, STROBE and Timmer criteria, assess the changes in presentation reporting quality between the two congresses and to recommend development of a national congress presentation abstract assessment system.

## MATERIAL AND METHODS

All oral presentations that have been accepted to the 17<sup>th</sup> National Surgical Congress (UCK-2010) and the 18<sup>th</sup> National Surgical Congress (UCK-2012) and have been included in the digital congress abstract booklets were searched in the electronic environment. Presentations were classified as "randomized controlled", "observational" and "experimental" studies according to study type. Presentations that cannot be included in one of these categories, such as cost analysis studies or surveys were classified as "other".

Prospective studies that stated random allocation of participants to either treatment or control groups were identified as randomized controlled trials. Prospective descriptive (cohort), retrospective case-control and cross-sectional studies, descriptive case series and case reports were defined as observational studies, and all the studies carried out in the laboratory on any animal form or human tissues and cells were evaluated as experimental studies.

The distribution of oral presentations sent to the congresses was given in Table 1. The sample group for observational studies was selected from 168 and 201 observational studies in both congresses, in order to predict the ability to detect a 10-15% difference with 90% accuracy by using computer-assisted random numbers, and 70 studies were selected from each congress. The sample group for randomized controlled trials was created by including all 14 randomized trials presented in both congresses. Experimental studies and studies classified

as other were excluded from the analysis. In order to blind the reviewers on information regarding both the author and the institution, someone other than the reviewers copied the presentations in a way not to include author or institution names.

Two reviewers (MH, FB) evaluated the reports in the sample group, independently, and according to the type of study. For scoring, the CONSORT (Attached file 1) and Timmer criteria (Attached file 2) were used for randomized controlled trials, and for observational studies the STROBE (Attached file 3) and Timmer criteria were used.

The original English text of each evaluation system was translated into Turkish, and was revised by each reviewer. The Turkish texts that were agreed on were used in the scoring process. Scores were recorded by the reviewers to computer-based datasheets that were prepared according to the used systems. Every single parameter that had an equal weight on the total score was scored as either 0 or 1 depending on whether the presentation possessed that characteristic in the CONSORT and STROBE criteria, and the resultant total score was recorded as the presentation score (CONSORT score range: 0-17, STROBE score range: 0-11).

Study type was excluded from scoring when Timmer score was used for evaluation and a binary scoring system (0: none, 1: yes) was used instead of the suggested triple scoring system (0: none, 1: partially valid, 2: completely valid) (6). Four parameters from a total of 19 parameters that are included in the Timmer score are not applicable to observational studies, therefore, the relevant four parameters were not used in the evaluation of observational studies. Finally, the Timmer score range was 0-19 for randomized controlled trials, while it was 0-15 for observational studies. Table 2 outlines the use of each assessment tool according to study type.

## Statistical Analysis

Comparisons regarding study types and total presentation number were made by using the chi-square test. The compatibility between the total score of the presentations were made by the Mann-Whitney-U test based on the scores given by the reviewers. The compatibility between the reviewers was evaluated by the Wilcoxon signed ranks test by consideration of congress total score individually or together.

## RESULTS

There was no difference in terms of distribution of study types and total number of accepted papers between UCK-2010 and

Table 1. Distribution of study type among 17<sup>th</sup> and 18<sup>th</sup> National Surgical Congress oral presentations

Study type	UCK-2010*	UCK-2012*
Randomized controlled	8 (3.8)	6 (2.5)
Observational	168 (80)	201 (82.4)
Experimental	25 (11.9)	27 (11)
Other	9 (4.3)	10 (4.1)
Total	210	244
*: n (%)		

UCK-2012 Congresses ( $p=0.844$ ). The most common type of study was observational studies (80% in UCK-2010 and 82.4% in UCK-2012). The total presentation scores given by the two reviewers according to study type were given in Table 3 (Figure 1). The highest score for randomized trials using the CONSORT evaluation system was 13 (maximum score 17), while the highest by Timmer tool was 12 (maximum score 19). The highest score for observational studies using the STROBE system was 9 (maximum score 11), while this score was 11 by the Timmer scale (maximum score 15).

There was no statistically significant difference in terms of total scores obtained from two independent reviewers with each assessment system for observational studies and randomized controlled trials in both Congresses (Figure 2) (Table 4).

There were no significant differences between the reviewers in terms of CONSORT scores used for UCK-2012 Congress randomized controlled trials and Timmer scores used for UCK-2010 observational studies. However, significant differences were detected between the reviewers in other evaluation systems (Table 5).

To eliminate the effect of Congress, when overall assessment by each reviewer was compared, significant difference was observed for each assessment system (Table 6).

Table 2. Distribution of evaluation tools according to study type

Evaluation type	Study type	Number of Parameters	Score value range
CONSORT	Randomized controlled	17	0-17
STROBE	Observational	11	0-11
Timmer	Randomized controlled	19	0-19
Timmer	Observational	15	0-15

## DISCUSSION

This study, which was designed to evaluate reporting quality of oral presentations in the 17<sup>th</sup> and 18<sup>th</sup> National Surgical Congresses, showed that there was no difference between the two congresses in terms of reporting quality, whereas evaluation performed by using standardized criteria revealed significant differences between the reviewers.

The rate of randomized controlled trials in UCK-2010 and UCK-2012 Congresses (2.5%-3.8%), which were similar in terms of distribution of study types and total number of presentations, were parallel to rates reported in the literature 2-7% (6, 11, 12). Nevertheless, due to the low rate of randomized controlled trials within congress presentations, a detailed assessment of the CONSORT system that was designed for randomized studies could not be made.

It is well known that using standard criteria to assess the quality of Congress presentation abstracts is important in improving reporting quality (1, 5, 6, 9, 13). The definition of evaluation systems that will be implemented in a Congress and notification of the authors in advance will ensure improvement in reporting and appropriateness of the study to scientific criteria. Since the parameters in the system are related to reporting as well as methodology, the authors will need to make the necessary arrangements in the study design. Although a one to one relationship could not be detected between quality of reporting and quality of the study, it is generally accepted that studies with low quality of reporting are particularly troublesome in terms of methodology (6). Improving reporting quality of abstracts and publications will also positively affect quality of the content. However, in the relevant study the rate of publications out of congress presentations were not examined, and no conclusions could be drawn regarding the possible hypotheses that presentations with high scores had higher likelihood of becoming a publication or that they would have a higher quality of content. In addition, an evaluation for any congress can only be used to compare presentations in that particular congress.

Table 3. Total score of UCK-2010 and UCK-2012 according to type of evaluation, study and reviewers

Congress	Type of evaluation	Type of study	N	Maximum score	Reviewer 1 total score*	Reviewer 2 total score*
UCK 2010	CONSORT	Randomized controlled	8	17	7.50±2.27 (8)	10±2.14 (10)
	STROBE	Observational	70	11	4.46±1.48 (4)	6.11±1.23 (6)
	Timmer	Randomized controlled	8	19	5.0±3.26 (5)	8.88±3.09 (10)
	Timmer	Observational	70	15	5.91±2.4 (6)	6.19±1.6 (6)
UCK 2012	CONSORT	Randomized controlled	6	17	7.83±1.60 (9)	9.50±1.64 (9)
	STROBE	Observational	70	11	4.33±1.44 (4)	5.87±1.27 (6)
	Timmer	Randomized controlled	6	19	4.67±1.97 (5)	8.50±1.64 (9)
	Timmer	Observational	70	15	5.23±2.2 (5)	6.09±2.2 (6)

\*: mean±standard deviation (median)

Table 4. Comparison of UCK-2010 and UCK-2012 congresses according to type of evaluation, study and reviewer

Study type	Evaluation type/reviewer	p
Randomized controlled	CONSORT/1	1.0
	CONSORT/2	0.431
	Timmer/1	0.948
	Timmer/2	0.558
Observational	STROBE/1	0.549
	STROBE/2	0.355
	Timmer/1	0.068
	Timmer/2	0.678

Table 5. Comparison of reviewers according to congresses type of evaluation, and study

Study type	Evaluation type/congress	p
Randomized controlled	CONSORT/UCK-2010	0.018
	CONSORT/UCK-2012	0.078
	Timmer/UCK-2010	0.027
	Timmer/UCK-2012	0.045
Observational	STROBE/UCK-2010	<0.001
	STROBE/UCK-2012	<0.001
	Timmer/UCK-2010	0.284
	Timmer/UCK-2012	0.001

Table 6. Comparison of reviewers according to type of evaluation

Type of evaluation	Reviewer 1 total score*	Reviewer 2 total score*	p
STROBE	4.39±1.46 (4)	6.0±1.26 (6)	<0.001
Timmer/observational	5.57±2.33 (5)	6.14±1.91 (6)	0.001
Timmer/randomized controlled	4.86±2.69 (5)	8.71±2.50 (9)	0.003
CONSORT	7.64±1.95 (9)	9.79±1.89 (10)	0.004

\*: mean±standard deviation (median)

That is why, it is impossible to make an overall assessment of what congress total scores mean.

It has been suggested that scales or checklists developed for evaluation of congress presentation abstracts mostly focus on reporting quality, while the system developed by Timmer evaluates quality of research method in addition to reporting quality (4). Although Timmer assessment system includes more criteria for methodology than the STROBE and CONSORT systems that are used in this study, there are repetitive criteria especially related to statistical methods (Timmer, 13-16 criteria, Additional file 2). This partially results in practical difficulties and leads to biased reviews.

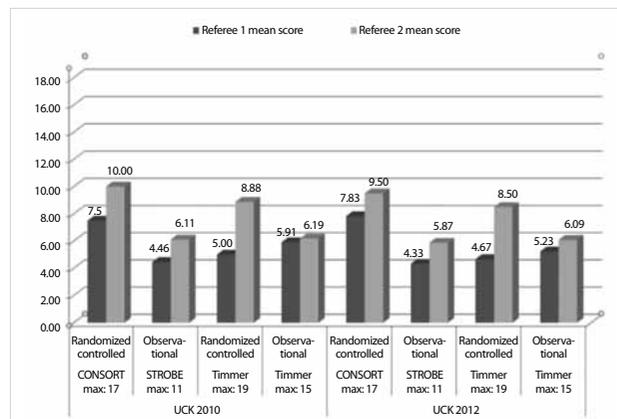


Figure 1. UCK-2010 and UCK-2012 Congress presentation score type of evaluation, study and reviewers

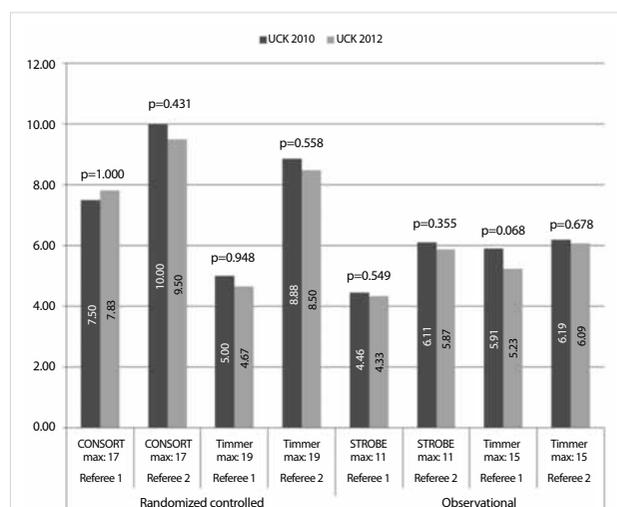


Figure 2. Comparison of UCK-2010 and UCK-2012 Congress presentation scores according to type of evaluation, study and reviewers

Evaluation of studies with subjective criteria that can be biased such as compliance with ethical principles, scientific validity or authenticity can lead to major problems (4). Evaluation of these criteria separately, as proposed by Timmer, and not considering these during standard evaluation process seems to be a more appropriate approach (4, 6).

If the reviewers do not know or do not consider the criteria in standardized systems the study may be evaluated insufficiently independent of its quality.

Concerning the identification of address and e-mail account, which are considered as important factors to contact the author even in case of potential institution changes, within the STROBE and CONSORT evaluation systems, and criteria such as specifying any funding source in the CONSORT system, in both Congresses none of the presentations obtained scores due to these. Additionally, because specification of the study type in the title was considered as a separate score in both systems, it was observed that these issues played a role in the relatively low scores obtained within the scope of this study. It is also considered that utilization of scoring systems for the evaluation of clinical trials can lead to biased conclusions (14). In some publications reviewer dependent criteria such as the

Table 7. Proposition for new congress presentation abstract evaluation system

Number	Group	Parameter
1	Title	Identification of study design in the title with common terms (randomized controlled, cohort, case-control, cross-sectional, case series, case report)
2	Author	Identification of contact information of corresponding author including e-mail address and exclusion of the Institution/Hospital name within the abstract
3	Aim	Identification of problem/aim/hypothesis
4	Method	Relevancy of the study design with the problem/aim/hypothesis of the study
5		Appropriate inclusion criteria, identification of selection, resource and methods
6		Identification of control group-if applicable
7		Identification of randomized allocation and management-if applicable
8		Identification of blinding the investigators, participants and reviewers, if applicable
9		Planned interventions for treatment groups
10		Specification of follow-up and result periods
11		Identification of variables and main outcome
12		Statement of participant number at the beginning and end of the study and reasons for exclusion
13	Statistics	Statistical methods, p value and confidence interval identification
14	Results	Reporting results in detail and according to aim/hypothesis
15	Conclusion	Interpretation of results in relevance to the aim/hypothesis
16		Relevancy of conclusions with results

importance of the subject, originality, overall quality of the study, possibility of raising discussion were also used as part of abstract evaluation (3, 15). In the evaluation of these types of criteria, personal interest of reviewers plays a greater role (3). Therefore, it is believed that using the total score rather than individual components that make up the score and having criteria that are objective and have less variability depending on the reviewer would be more beneficial in the evaluation of different designs or various subjects (4). It is believed that since the National Surgical Congress is a main specialty congress with general contents, using total score consisting of multiple criteria during the evaluation of reports is a more appropriate approach.

Different results were obtained from the two reviewers after evaluation of presentations accepted to the congress by three different systems. The finding that these differences were detected in studies from both Congresses and in both observational and randomized controlled trials suggests that the single variable was the reviewer. A similar study conducted by Montgomery et al. (15) stated that during evaluation of abstracts, reviewers were in greater agreement in criteria that were related to design and methodology rather than subjective components. In case of participation of multiple reviewers in the evaluation process, taking the average of these scores is common practice (4, 6, 16). In this study, comparison of the quality of presentations was not determined as an end-point, and the average of two different scores were not taken into account.

Inclusion of subjective criteria as part of evaluation is stated as the most important reason for mismatches between the reviewers (2, 15). It is believed that subjective criteria that are included in the three different systems used in the study such as clearly, sufficiently, well-defined and general interpretation

play an important role in the differences between reviewers. It has also been observed that some clarification issues occurred during translation of the texts and terms into Turkish, which are originally in English. One-to-one translation of some statements may cause understanding problems, and additions for clarity may lead to diversion from the original text. Therefore, the current international presentation evaluation systems should be arranged according to national regulations. It is believed that a new assessment system to be used in the evaluation of presentation abstracts in National congresses is necessary. It is considered that taking into account the parameters of existing evaluation criteria, a new evaluation system consisting of 16 parameters to be used for this purpose is appropriate (Table 7). It is predicted that this more simple and applicable scoring system with binary scoring (yes: 1, no:0) can be applied initially both to randomized controlled and observational studies.

#### CONCLUSION

In national based general surgery congresses, the relation between the best presentations announced and the publication rates of the presentations in the congresses should be illuminated. It will be possible to construct a widely accepted evaluation system with the help of these studies and other contributions.

**Ethics Committee Approval:** In this study, evaluation of reporting quality of oral presentations at 2010 and 2012 National Surgical Congresses was performed using CONSORT, STROBE and Timmer criteria. Therefore, Ethics Committee Approval was not taken. In scope of the study, there was no intervention on patients.

**Informed Consent:** Due to the lack of patient participation, patient consent was not taken.

**Peer-review:** Externally peer-reviewed.

**Author Contributions:** Concept - M.H., F.B., Ö.U.; Design - M.H., F.B., Ö.U.; Supervision - M.H., F.B., Ö.U.; Data Collection and/or Processing - M.H., F.B.; Analysis and/or Interpretation - M.H., F.B.; Literature Review - M.H., F.B.; Writer - M.H., F.B., Ö.U.; Critical Review - M.H., F.B., Ö.U.

**Conflict of Interest:** No conflict of interest was declared by the authors.

**Financial Disclosure:** The authors declared that this study has received no financial support.

## REFERENCES

1. Yoon U, Knobloch K. Assessment of reporting quality of conference abstracts in sports injury prevention according to CONSORT and STROBE criteria and their subsequent publication rate as full papers. *BMC Med Res Methodol* 2012; 12: 47. [\[CrossRef\]](#)
2. Bydder S, Marion K, Taylor M, Semmens J. Assessment of abstracts submitted to the annual scientific meeting of the Royal Australian and New Zealand College of Radiologists. *Australas Radiol* 2006; 50: 355-359. [\[CrossRef\]](#)
3. Kabay B, Teke Z, Erbiş H, Koçbil G, Tekin K, Erdem E. The rate of international publications out of National Surgical Congress presentations. *Ulusal Cer Derg* 2005; 21; 130-134.
4. Timmer A, Sutherland LR, Hilsden RJ. Development and evaluation of a quality score for abstracts. *BMC Med Res Methodol* 2003; 3: 2. [\[CrossRef\]](#)
5. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008; 371: 281-283. [\[CrossRef\]](#)
6. Knobloch K, Yoon U, Rennekampff HO, Vogt PM. Quality of reporting according to the CONSORT, STROBE and Timmer instrument at the American Burn Association (ABA) annual meetings 2000 and 2008. *BMC Med Res Methodol* 2011; 11: 161. [\[CrossRef\]](#)
7. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276: 637-639. [\[CrossRef\]](#)
8. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenberg JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370: 1453-1457. [\[CrossRef\]](#)
9. Sorensen AA, Wojahn RD, Manske MC, Calfee RP. Using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement to assess reporting of observational trials in hand surgery. *J Hand Surg Am* 2013; 38: 1584-1589. [\[CrossRef\]](#)
10. STROBE Statement-Items to be included when reporting observational studies in a conference abstract. Available from URL: [http://www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE\\_checklist\\_conference\\_abstract\\_DRAFT.pdf](http://www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_conference_abstract_DRAFT.pdf).
11. Hill CL, Buchbinder R, Osborne R. Quality of reporting of randomized clinical trials in abstracts of the 2005 annual meeting of the American College of Rheumatology. *J Rheumatol* 2007; 34: 2476-2480.
12. Becker A, Blümle A, Antes G, Bannasch H, Torio-Padron N, Stark GB, et al. Controlled trials in aesthetic plastic surgery: a 16-year analysis. *Aesthetic Plast Surg* 2008; 32: 359-362. [\[CrossRef\]](#)
13. Vandenberg JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? *J Clin Epidemiol* 2009; 62: 594-596. [\[CrossRef\]](#)
14. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282: 1054-1060. [\[CrossRef\]](#)
15. Montgomery AA, Graham A, Evans PH, Fahey T. Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Serv Res* 2002; 2: 8. [\[CrossRef\]](#)
16. Sehra J, Wright NS, Polychronopoulou A, Cobourne MT, Pandis N. Reporting quality of abstracts of randomized controlled trials published in dental specialty journals. *J Evid Based Dent Pract* 2013; 13: 1-8. [\[CrossRef\]](#)

### Addendum: 1. CONSORT abstract evaluation system

Number	Parameter	Description
1	Title	Identification of the study as randomized
2	Authors	Contact details for corresponding author
3	Study design	Identification of the study design
4	Participants	Eligibility criteria for participants and the environment in which the data is collected
5	Interventions	Planned interventions for each group
6	Aim	Special purpose or hypothesis
7	Results	Clearly defined outcomes for this article
8	Randomization	In what way the participants are allocated to interventions
9	Blinding	Whether participants, care providers and reviewers were blinded to allocation into groups
10	Randomization number	The number of randomized participants in each group
11	Inclusion	Study status
12	Analysis number	The number of participants analyzed in each group
13	Results	For the primary outcome results, data for each group, its estimated power and accuracy
14	Harm	Important adverse events or side effects
15	Results	General review of the findings
16	Study registration	Registration number and name of study record
17	Funds	Source of funding

**Addendum: 2. Timmer abstract evaluation tool**

<b>Number</b>	<b>Parameters</b>
1	Is the question / aim adequately defined?
2	Is the study design clear and suitable for answering the question?
3	Are participant characteristics identified sufficiently?
4	Are the participants eligible for study questions?
5	Has control been used and is it suitable? (Check no if there is no control).
6	Is participant selection method defined and is it suitable?
7	If randomization is used within treatment groups, is it clearly identified? (Check not applicable if not used).
8	If investigators were blinded for interventions, is this situation clearly identified? (Check not applicable if not used).
9	If participants were blinded for interventions, is this situation clearly identified? (Check not applicable if not used).
10	Is the result evaluation well-defined and is it bias proof? Are assessment tools reported?
11	Are the reasons for confounding results specified?
12	Is the sample size adequate?
13	Are post hoc power calculations or confidence intervals reported for statistically insignificant results?
14	Are statistical analysis suitable?
15	Are statistical tests identified?
16	Are the exact p-values and confidence intervals defined?
17	Are non- participation numbers and reasons for non-participation reported?
18	Are the findings reported in sufficient detail?
19	Do results support the conclusions?

**Addendum: 3. STROBE abstract evaluation tool**

Number	Parameter	Description
1	Title	Specifying study design within the title with common terms (eg cohort, case-control, cross-sectional)
2	Authors	Contact details for corresponding author
3	Study design	Defining the study design (eg cohort, case-control, cross-sectional)
4	Aims	Specific objectives or hypotheses.
5	Adjustment	Specifying the adjustments, the follow-up period or the time results occurred (for example, the prevalence in the age of 18, 1998-2007).
6	Participants	<p><i>Cohort study</i>- identification of the most important eligibility criteria, and the most important sources and methods of participant selection. A brief description of follow-up methods.</p> <p>Case-control study- identification of eligibility criteria, and sources and methods of case and control selection</p> <p><i>Cross-sectional study</i>- identification of the eligibility criteria, and selection methods and main resource of participants</p> <p><i>Cohort studies</i>- for paired studies, identifying number and matching of exposed and non-exposed participants</p> <p>Case-control study- for paired studies, identifying matching criteria and the number of controls per case</p>
7	Variables	Clear definition of primary outcome variable for this article
8	Statistical methods	Explanation of statistical methods, including those used to control confounding results.
9	Participants	Are the number of participants at the beginning and end of the study reported?
10	Main outcomes	<p>Reporting relations of links. If relevant, consider translating estimates of relative risk into absolute risk for a meaningful period.</p> <p>Identify appropriate measures of uncertainty and variability (eg odds ratio with confidence intervals)</p>
11	Results	General interpretation of study results