










Criteria for selection and classification of studies in medical events

René Aloisio da Costa Vieira^{1*} , Regis Resende Paulinelli² , Fábio Francisco Oliveira Rodrigues³ ,
Marise Amaral Rebouças Moreira² , Ricardo Caponero⁴ , Eduardo Carvalho Pessoa¹ ,
Rosemar Macedo Sousa Rahal² , Gil Facina⁵ , Ruffo de Freitas Junior² 

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the impact of study methodology and evaluation type on the selection of studies during the presentation of scientific events.

METHODS: A prospective, observational, transversal approach was applied to a cohort of studies that were submitted for presentation at the 2021 Brazilian Breast Cancer Symposium. Three forms of criteria (CR) were presented. CR1 was based on six criteria (method, ethics, design, originality, promotion, and social contribution); CR2 graded the studies from 0 to 10 for each study, and CR3 was based on five criteria (presentation, method, originality, scientific knowledge, and social contribution). To evaluate the item correlation, Cronbach's alpha and factorial analysis were performed. For the evaluation of differences between the tests, we used the Kruskal-Wallis and *post-hoc* Dunn tests. To determine the differences in the study classifications, we used the Friedman test and Nemenyi's all-pairs comparisons.

RESULTS: A total of 122 studies were evaluated. There was also a good correlation with the items concerning criterion 1 ($\alpha=0.730$) and 3 ($\alpha=0.937$). Evaluating CR1 methodology, study design and social contribution ($p=0.741$) represents the main factor and CR3 methodology, and the scientific contribution ($p=0.994$) represents the main factor. The Kruskal-Wallis test showed differences in the results ($p<0.001$) for all the criteria that were used [CR1-CR2 ($p<0.001$), CR1-CR3 ($p<0.001$), and CR2-CR3 ($p=0.004$)]. The Friedman test showed differences in the ranking of the studies ($p<0.001$) for all studies ($p<0.01$).

CONCLUSION: Methodologies that use multiple criteria show good correlation and should be taken into account when ranking the best studies.

KEYWORDS: Planning techniques. Congress. Meeting abstract. Societies, scientific. Methods.

INTRODUCTION

Medical scientific events (MSEs) are spaces for the recycling of scientific knowledge where updates are presented on changing trends in basic science, diagnosis, or treatment. In addition, they allow the strengthening of medical societies and the presentation of new inputs and novel technologies¹.

The refinement observed in the selection criteria of articles is not extended to scientific events. The size of an abstract limits the details involved in a study. The quality of the abstract presentation influences the acceptance rate^{1,2}.

Generally, studies in progress are presented, but only some are published³⁻⁵. Oral presentations (OP)⁵, the region of origin of the study⁶ or the institutions involved, the sample size,

the presence of positive results, or the level of evidence have a positive impact on publication⁷. Several factors can influence the selection and classification of studies, including the quality of a study, the form of presentation of its abstract,² and its methodology. Regarding the evaluation committee, relevant factors include the training of evaluators (basic science or clinical practice)⁸, the criteria used⁸, the area of the study⁹, the form of analysis (structured evaluation)¹⁰, the blinding of the evaluators¹¹, the methodology used during agreements or disagreements among the evaluators⁹, and the concordance between the evaluators⁹.

There are criteria for the selection of the studies and criteria for ranking them. For initial selection, simple criteria can

¹Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu, Programa de Pós-Graduação em Tocoginecologia – Botucatu (SP), Brazil.

²Universidade Federal de Goiás – Goiania (GO), Brazil.

³Instituto do Câncer Dr. Arnaldo Vieira de Carvalho – São Paulo (SP), Brazil.

⁴Hospital Alemão Oswaldo Cruz – São Paulo (SP), Brazil.

⁵Universidade Federal de São Paulo – São Paulo (SP), Brazil.

*Corresponding author: reneacv@gmail.com

Conflict of interest: the authors declare there is no conflicts of interest. Funding: none.

Received on November 20, 2022. Accepted on December 19, 2022.

be used, such as a Likert scale ranking (minor=worst; maximum=best)¹⁰, adding criteria for accepting (e.g., reject, unsure, accept)^{9,10}, or adding abstract items (abstract's clarity, significance of learning objectives, relevance in clinical practice, grouping it in a Likert Scale)¹². Limited studies presented the criteria used for selecting abstracts at scientific events^{12,13,8}, justifying the present study.

METHODS

The present study did not involve human beings, so it did not require evaluation by the Brazilian Research Ethics Committee. We conducted an observational, prospective, and blind study to evaluate the influence of the criteria used for the evaluation of studies and their relationship with the order of their classification in scientific events. We used an event in the area of mastology as our basis. The criteria that were used for the evaluation of the abstracts followed a recently published model⁸.

The study was conducted at the 2021 Brazilian Breast Cancer Symposium (BBCS). The members of the scientific committee were invited to participate in the study. Only members who completed the evaluation of all studies and evaluated them according to the three criteria participated in this research. To the committee, the studies were presented in a Microsoft Excel[®] spreadsheet in a blind manner, and three columns were presented for their evaluation of the different criteria. Clinical studies were separated from case reports. The first criterion (CR1) was based on six criteria (method, ethics, design, originality, promotion, and social contribution). The sum for each study resulted in 10 points, representing the event pattern⁸. For each factor, the score was predefined. In the second criterion (CR2), the evaluator rated the study from 0 to 10. The third criterion (CR3) was based on previous mastology meetings, considering 5 criteria (presentation, method, originality, scientific and social contribution), scored from 0 to 10 (free score), with the sum divided by 5, representing the final score based on the mean evaluation. If the evaluator participated in one research, we opted to use the mean of the other evaluators. Table 1 specifies the criteria used. Subsequently, the classification of the studies was evaluated using the criterion used in the event (CR1) as a standard, and the 10 best potential studies were identified. For analysis, we consider the best studies considered by the study evaluators.

Statistics

We sought to evaluate whether the medians of the tests were equal. For this purpose, the Kruskal-Wallis test was used. When there were changes between the tests, a *post hoc* evaluation was

performed using the Dunn test to determine the tests where there was a difference. To demonstrate whether there was a difference in the classification position between the studies, the Friedman test was used.

When there were changes between the tests, a *post hoc* evaluation was performed using Nemenyi's all-pairs comparisons test to identify which tests showed differences. In these evaluations, we used the program "R".

We attempted to quantify the relationship between the results of these methods using multiple evaluation methodologies (CR1 and CR3) using Cronbach's alpha. To determine the best methodology and the smallest number of items that could determine the same results, factor analysis was performed (Table 2). The minimum sample size necessary for factorial analysis was 100 patients¹⁴. To compare the evaluation methodologies, the IBM SPSS[®] software for Mac[®] was used.

RESULTS

All 20 professors on the scientific committee were invited to participate in this research, allowing us to observe the adherence of five members in all evaluations, all senior professionals, and from different services. Among the evaluators, the mean age was 58 years (range 49–71 years), with an average of 25.4 years (range 18–36 years) of activity in mastology and 13.8 years (range 7–20 years) of participation on scientific congress committees. All of them had completed medical residencies, with four doctors (Ph.D.) and one master (M.Sc.). All of them advocated the separation of studies into clinical articles, molecular biology research, and case reports. When asked about the criteria, 4 of 5 evaluators considered it important to use predefined criteria in the evaluation of studies. The researchers were unanimous in their identification of study design, method, originality, ethics, and clinical relevance as important criteria.

Approximately 122 studies were evaluated, including 94 original studies and 28 case reports. Regarding the criteria used in the event (mean±standard deviation), CR1 indicates that original studies (5.62±0.92) received better scores than case reports (3.64±0.84). Evaluating the original studies and the type of criteria, CR2 presented higher scores [CR2 (6.43±0.72) > CR1 (5.62±0.92) > CR3 (4.61±0.84)].

Binary comparisons (CR1/CR3, CR1/CR2, and CR2/CR3) were performed. The Kruskal–Wallis test showed that there was a difference in the medians between the tests ($p < 0.001$). There was a difference between the CR1–CR2 criteria ($p < 0.001$), CR1–CR3 criteria ($p < 0.001$), and CR2–CR3 criteria ($p = 0.004$). According to the Friedman test, there was also a significant difference in the classification of the studies. Nemenyi's all-pairs

Table 1. Criteria used in the study.

| Criterion 1 – Single criteria adapted from Mastology ⁸ | |
|---|--|
| Value | Criteria |
| | Study method |
| 2.8 | Systematic review of randomized studies with or without a meta-analysis. Molecular: Omics studies (genomics, transcriptomics, proteomics) |
| 2.4 | Randomized experimental studies. Molecular - Functional studies (<i>in vitro/in vivo</i>) |
| 2.0 | Cohort Studies. Molecular: The identification of biomarkers (with validation methodology) |
| 1.6 | Case control studies |
| 1.2 | Case series. Molecular: Descriptive studies without validation or without a control group |
| 0.8 | Case report |
| 0.4 | Expert opinions |
| Ethics in research | |
| 1.0 | Approval from the ethics committee |
| 1.0 | No need for a Research Ethics Committee under Resolution No. 466 |
| 0.0 | No description or evaluation by the ethics committee |
| Design/Study presentation | |
| 2.5 | Adequate description of the study with clear, reproducible methodology, consistent results and adequate conclusion that is compatible with the data presented. Approved through ClinicalTrials/ReBEC or something similar. |
| 2.0 | Adequate description of the study with clear, reproducible methodology, consistent results and adequate conclusion that is compatible with the data presented. |
| 1.5 | Adequate description of the study, however the methodology is weak (not reproducible), consistent results and adequate conclusion that is compatible with the data presented. |
| 1.0 | Adequate description of the study, however the methodology is weak (not reproducible), and the results and/or conclusions were not adequate for the data presented. |
| 0.5 | Severe failures in the introduction, methodology, results and conclusions. |
| 0.0 | Does not apply. Methodology and results not described. |
| Originality | |
| 1.7 | Unprecedented - new interpretation of the concept |
| 1.2 | Ratifies a known concept that is optional |
| 0.7 | Ratifies a classic concept that is used everyday |
| 0.4 | Does not introduce a new concept |
| Grant/Promotion | |
| 1.0 | Promotion from a public agency |
| 0.5 | Promotion from a private agency |
| 0.0 | Self-promotion or no promotion |
| Clinical/Social contribution | |
| 1.0 | In clinical practice and/or social practice |
| 0.5 | In the study subgroup |
| 0.0 | No clinical applicability or does not fit |
| Criterion 2 – Simple assessment of Abstracts – Grade from 0 to 10 for the study. | |
| Criterion 3 – Criteria used in past Congresses – Grade from 0 to 10 for each of the item. | |
| Presentation – Method – Originality – Social Contribution – Scientific Contribution | |

Table 2. Factor analysis results*.

| Order | Criteria 1 | | Criteria 3 | |
|-------|---------------------|-------|-------------------------|-------|
| | Category | Value | Category | Value |
| 1 | Method | 0.868 | Scientific contribution | 0.944 |
| 2 | Study design | 0.843 | Presentation | 0.934 |
| 3 | Social contribution | 0.741 | Originality | 0.922 |
| 4 | Originality | 0.588 | Method | 0.914 |
| 5 | Ethics | 0.449 | Social contribution | 0.836 |
| 6 | Grant | 0.438 | - | - |

*In the criteria 1, the scores were pre-defined for each type of assessment, and in criterion 3, the evaluator was free to give a score within the question.

comparisons *post hoc* test indicated differences ($p < 0.01$) for all comparison pairs. When exclusively evaluating the top 10 (Figure 1), the Friedman test showed a difference between the rank of the abstract in relation to the different criteria used ($p < 0.001$), and Nemenyi's all-pairs test showed no difference between the CR1×CR2 criteria ($p = 0.17$) and CR1×CR3 criteria ($p = 0.06$), although the CR2×CR3 criteria were very different ($p = 0.01$).

The correlation between the items observed in CR1 was evaluated, demonstrating a good correlation ($\alpha = 0.730$), whereas factorial analysis showed that only three criteria would best represent the evaluation, namely, study method, study design, and social relevance (Table 2). The correlation between the items observed in CR3 was evaluated, showing an excellent correlation ($\alpha = 0.937$), whereas factor analysis showed that only one criterion would best represent the evaluation, namely, scientific contribution (Table 2).

DISCUSSION

Studies at scientific events have greater flexibility since they constitute a space where several professionals from the same specialty are gathered, and there is also space for the presentation of institutions or infrequent situations. The type of scientific event influences the quality of the studies presented, with international specialty events at the top of this hierarchy, associated with abstracts in high-impact factor journals, followed by other international events and then national, regional, and local events. Depending on the event, these selected studies are published in impact factor journals or event supplements.

Between the presentation of a study at a medical event and its final publication, there is a period that should be considered. This path is long, and many studies have never been published. The publication rate varies from 7.9%⁴ to 57%⁵, a fact influenced by the regularity and hierarchy of medical events, the occurrence of concurrent events with the same specialty and type of specialty, or the acceptance of previously published studies⁶.

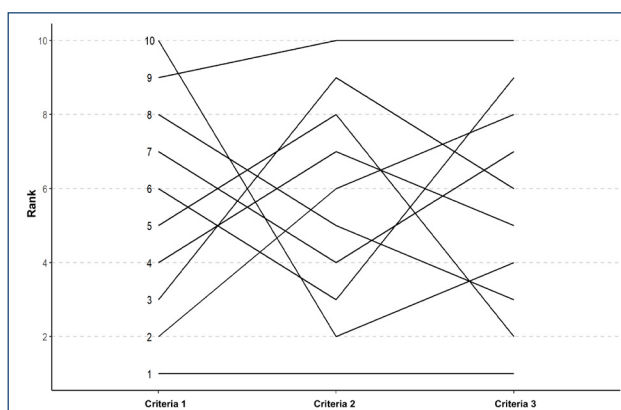


Figure 1. Adjusted variation of position among the 10 best studies.

The average time for publication is 2 years^{3,5}, and most studies are published in 5 years⁵. In Brazil, the number of studies published in public services is higher^{4,6}, with most of these published in national and specialty journals^{4,7}. Among the reasons for nonpublication are time, reluctance to publish incomplete findings, no attempt at publication, the need to increase casuistry, the responsibility of another author, and rejected study³.

To select the best abstracts, we evaluated three different methodologies. After evaluating the scores, a factor analysis was performed to determine the main factors associated with the studies by employing methodology, design/presentation, and social relevance as multiple criteria and scientific contribution as a single criterion. This method should be considered during the future selection and classification of abstracts. There is less description in the literature regarding the criteria used in the selection of abstracts. It is important to consider the description of clinical applicability, innovation, clarity in the description of findings (objective, hypothesis, description of findings, and discussion), and quality of the method¹⁵. Based on the quality of study evidence, one study suggested quality scores² and the other suggested criteria to increase the study description, given transparency, and integrity for publications submitted

to conference¹. We previously created clear and reproducible criteria⁸. However, these were grouped to facilitate the evaluation of clinical and basic studies, and therefore a lower score was used for case reports.

In the literature, the number of evaluators reported was 3², 5¹³, 9⁹, or 10¹⁰. But we have to evaluate the number of abstracts evaluated by each evaluator, and the evaluator's characteristics. One publication reviewed 938 abstracts, 70–100 members, and 20–30 abstracts/evaluators¹². Another considered from 17,205 abstracts, 1,000 were selected, and 100 were evaluated in the study by three researchers, creating a potential criteria for quality². New members in the association evaluated the abstracts (n=194) of one study¹³. Although reliability generally increases with the number of reviewers, the annual increase in abstracts submitted may require a decrease in the number of reviewers for each abstract, a fact that difficult studies in this area¹⁰. Our study included five senior professionals (four PhDs and one MSc).

Different methods were used to choose the abstract, varying the Likert scale with or without other criteria^{9,10,12}. Likert scales have different ranges (-6 to +6¹⁰, 1 to 7¹², and 1 to 5⁹), and we chose 0–10. For quality criteria, the literature is not uniform in relation to abstract items,¹² and one publication suggests 14 important items for evaluation². We used three models, namely, criteria+scale (CR1), Likert (CR2), and criteria+Likert (CR3). Model CR1 used six predetermined criteria,⁸ and CR3 used a Likert scale (0–10) in five situations. Factorial analysis selected the main items for CR1 that were method, design, and social relevance, so the quality of the study and relevance in clinical practice were important for the abstract evaluation. When the evaluation of CR3's scientific contribution was considered, multiple criteria reduce the subjectivity of the evaluation and help the evaluator.

The BBCS has established itself as the main event for clinical and basic research in mastology in Brazil, and it is currently in its tenth edition. For the selection of the studies, we took some care to avoid possible bias, as the form of evaluation (blind or not blind) interferes with the acceptance of studies and programming¹¹. To avoid biases related to the participation of the study (detection bias), we opted to use the mean of the other four evaluators, a fact that occurred in 8 abstracts (6.5%). To prevent biases among the evaluators, national researchers with different services and high experience (senior) in the field of mastology were invited. In our study, the high experience of the committee may have entailed a certain bias in the evaluations; when using CR1, three factors were found to be representative because the committee was comprised of professionals with primarily clinical activity. To avoid biases related to the

evaluators (attrition bias), we chose to evaluate similar situations and compare the results, and we stopped the evaluation based on the five evaluators, using the same methodology in the same sample.

We do not compare our results with the final classification of the BBCS, as 20 researchers classify the abstracts. We evaluated the results related to our five reviewers. Concerning only the 10 top-ranked studies selected from the event (CR1), there was a difference in their order of classification in relation to the other criteria (Figure 1). We thus observed that the top study, due to its quality, remained in the first place, independent of classification.

The use of a structured questionnaire can be useful in the objective evaluation of abstracts during a scientific meeting and can facilitate the comparison of abstracts. The meritocratic distribution of abstracts among reviewers is thus advocated, and more studies are necessary to improve the reliability of their classification, justifying future studies.

CONCLUSION

The original studies received better scores. Methodologies that used many criteria showed a good correlation, which was the preference of the evaluators. The methodology used in the evaluation of studies thus influences the classification of the best studies. In the selection of criteria, a detailed method, study design, and scientific contribution were relevant.

ACKNOWLEDGMENTS

We thank Flavio Ferraz Vieira for his help in the statistical evaluation.

AUTHORS' CONTRIBUTIONS

RACC: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **RRP:** Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **FFOR:** Data curation, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **MARM:** Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **RC:** Data curation, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **ECP:** Conceptualization, Visualization, Writing – original draft, Writing – review & editing. **RMSR:** Resources, Visualization, Writing – original draft, Writing – review &

editing. **GF:** Conceptualization, Methodology, Validation, Visualization, Writing – original draft, Writing – review &

editing. **RFJ:** Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing.

REFERENCES

1. Foster C, Wager E, Marchington J, Patel M, Banner S, Kennard NC, et al. Good practice for conference abstracts and presentations: GPCAP. *Res Integr Peer Rev.* 2019;4:11. <https://doi.org/10.1186/s41073-019-0070-x>
2. Timmer A, Sutherland LR, Hilsden RJ. Development and evaluation of a quality score for abstracts. *BMC Med Res Methodol.* 2003;3:2. <https://doi.org/10.1186/1471-2288-3-2>
3. Oliveira LR, Figueiredo AA, Choi M, Ferrarez CE, Bastos AN, Netto JM. The publication rate of abstracts presented at the 2003 urological Brazilian meeting. *Clinics (Sao Paulo).* 2009;64(4):345-9. <https://doi.org/10.1590/s1807-59322009000400013>
4. Rahal RMS, Nascimento S, Soares LR, Freitas-Junior R. Publication rate of abstracts on breast cancer presented at different scientific events in Brazil. *Mastology.* 2020;30(1):e20200048. <https://doi.org/10.29289/2594539420202020200048>
5. Gürses iA, Gayretli Ö, Gürtekin B, Öztürk A. Publication rates and inconsistencies of the abstracts presented at the national anatomy congresses in 2007 and 2008. *Balkan Med J.* 2017;34(1):64-70. <https://doi.org/10.4274/balkanmedj.2016.0360>
6. Brito MV, Botelho NM, Yasojima EY, Teixeira RK, Yamaki VN, Feijó DH, et al. Publication rate of abstracts presented in a Brazilian experimental surgery congress. *Acta Cir Bras.* 2016;31(10):694-97. <https://doi.org/10.1590/S0102-865020160100000009>
7. Forlin E, Fedato RA, Junior WA. Publication of studies presented as free papers at a Brazilian national orthopedics meeting. *Rev Bras Ortop.* 2013;48(3):216-20. <https://doi.org/10.1016/j.rboe.2012.10.004>
8. Vieira RAC, Bonetti TCV, Marques MMC, Facina G. Criteria for evaluating studies at scientific medical events. *Mastology.* 2020;30(1):e20200031. <https://doi.org/10.29289/25945394202020200031>
9. Bhandari M, Templeman D, Tornetta P. Interrater reliability in grading abstracts for the orthopaedic trauma association. *Clin Orthop Relat Res.* 2004;(423):217-21. <https://doi.org/10.1097/01.blo.0000127584.02606.00>
10. van der Steen LP, Hage JJ, Kon M, Mazzola R. Reliability of a structured method of selecting abstracts for a plastic surgical scientific meeting. *Plast Reconstr Surg.* 2003;111(7):2215-22. <https://doi.org/10.1097/01.PRS.0000061092.88629.82>
11. Smith J, Nixon R, Bueschen AJ, Venable DD, Henry HH. Impact of blinded versus unblinded abstract review on scientific program content. *J Urol.* 2002;168(5):2123-5. <https://doi.org/10.1097/01.ju.0000034385.02340.99>
12. Newsom J, Estrada CA, Panisko D, Willett L. Selecting the best clinical vignettes for academic meetings: should the scoring tool criteria be modified?. *J Gen Intern Med.* 2012;27(2):202-6. <https://doi.org/10.1007/s11606-011-1879-2>
13. van der Steen LP, Hage JJ, Kon M, Monstrey SJ. Validity of a structured method of selecting abstracts for a plastic surgical scientific meeting. *Plast Reconstr Surg.* 2004;113(1):353-9. <https://doi.org/10.1097/01.PRS.0000097461.50999.D4>
14. Matos DAS, Rodrigues EC. *Análise fatorial.* 1st ed. Brasília: ENAP Escola Nacional de Administração Pública; 2019.
15. Deveugele M, Silverman J. Peer-review for selection of oral presentations for conferences: are we reliable? *Patient Educ Couns.* 2017;100(11):2147-50. <https://doi.org/10.1016/j.pec.2017.06.007>

