Bevacizumab as first-line treatment of ovarian cancer: strengths and weaknesses of subgroup analyses

Sandro Pignata[1], Sabrina Chiara Cecere[1], Massimo Di Maio[2]

Abstract
Bevacizumab is approved as first-line treatment for ovarian cancer, in combination with carboplatin-paclitaxel, in patients with FIGO IIIb or IV disease, and as maintenance therapy. The results of subgroup analyses of data from the ICON7 study have resulted in bevacizumab use being restricted to "high-risk" patients in several countries. While a goal of obtaining as much information as possible from clinical trial data is understandable, there are also a number of limitations associated with subgroup analyses. Interpretation can be limited by the multiplicity of statistical tests (increasing the risk of false positive findings) and by low statistical power (increasing the chance of false negative results). In addition, the credibility of subgroup analyses is improved if they are restricted to the primary study outcome and limited to a small number of predefined patient subgroups. Overall, subgroup findings should be considered exploratory and hypothesis-generating. The subgroup analysis in ICON7 had a number of limitations, including *post hoc* definition of "high-" and "low-risk" groups, and questionable applicability of these definitions to clinical practice. Use of residual disease as a determinant of risk status is also problematic. Excessive emphasis on the role of subgroup analyses may lead to misunderstanding of the true results generated by the overall trial data, and support the incorrect selection of patients to be treated with the new or experimental therapy. Caution should be used in translating results of subgroup analysis to clinical practice guidelines.

Key words: bevacizumab, clinical trials, ovarian cancer, subgroup analysis

Bevacizumab is approved as first-line treatment of ovarian cancer, given concurrently with carboplatin-paclitaxel and continued as maintenance. Approval was based on the results of the GOG218 randomized trial, that enrolled only patients with incompletely resected stage III or stage IV disease [1]. Bevacizumab has been approved by the European Medicines Agency (EMA) without any restriction in terms of baseline patient characteristics, with the exception of disease stage, because approved was granted only for FIGO IIIb to IV patients. In contrast to GOG218, the ICON7 trial investigated the efficacy of adding bevacizumab to standard chemotherapy in a population that included both patients with high-risk early stage disease and those with more advanced disease. As a result, ICON7 included a subgroup of patients with an absence of residual disease after surgery. The primary endpoint of the ICON7 trial, as in GOG218, was progression-free survival (PFS), and the intention-to-treat (ITT) analysis demonstrated a significant PFS prolongation in favor of bevacizumab. Overall survival (OS), a secondary endpoint, was not significantly prolonged in the overall study population [2]. However, the authors have also published subgroup analyses, showing that OS benefit was evident only in the subgroup of "high-risk" patients (defined as those not operated on or with residual disease of >1 cm after primary surgery, or those with stage IV disease) [2]. In several countries, these subgroup analysis results have affected clinical practice, and resulted in the use of bevacizumab being limited to "high-risk" patients. The aim of this paper is to underline the potential usefulness, but also the limitations, of these kinds of subgroup analyses.

It is quite reasonable that investigators and sponsors want to gain as much useful information as possible from clinical trial data. Regulatory agencies also are keen to know whether there are subgroups of trial participants who are more likely to benefit from the intervention under investigation. Several surveys of trials published in leading journals have consistently found that most publications of clinical study results include subgroup analyses [3-5].

The interpretation of subgroup analyses from a random-

[1]Department of Uroginecology, Istituto Tumori di Napoli, Fondazione G. Pascale IRCCS, Napoli, Italy.
[2]Department of Oncology, Università di Torino, AOU San Luigi Gonzaga, Orbassano (Torino), Italy.
Correspondence to: Sandro Pignata MD,
Dipartimento Uro-Ginecologico, Istituto Nazionale Tumori,
via M. Semmola 52, 80131 Napoli, Italy.
Phone: + 39 081 5903637 – Fax: + 39 081 5903861
E-mail: s.pignata@istitutotumori.na.it
CANCER BREAKING NEWS 2016;4(1):45-47
DOI: 10.19156/cbn.2016.0009

ized trial is limited by the multiplicity of statistical tests (increasing the risk of false positive findings) and by low statistical power (increasing the chance of false negative results). Therefore, there is a gap between the reasonable desire to identify heterogeneity in the efficacy of trial interventions in different patient groups and the technical capacity to produce reliable results when exploring subgroups. It has been identified that the reporting of subgroup analyses of clinical trials is characterized by suboptimal quality and methodological pitfalls [3, 5, 6].

Investigators should be cautious when undertaking subgroup analyses. As a general rule, subgroup findings should be exploratory and hypothesis-generating, in order to produce evidence that warrants confirmation in further prospective trials. Consistent with this approach, the results of subgroup analyses should only influence the conclusion of the overall trial when a clear interaction between the absence or presence of treatment efficacy and a specific patient characteristic is evident.

The credibility of subgroup analyses is improved if confined to the primary outcome and to a few predefined subgroups selected based on biologically-plausible hypotheses. Investigators should recognize that their trial is not large enough to detect realistic subgroup effects, and be particularly wary of claiming a treatment difference in a subgroup when the overall treatment comparison is not statistically significant. Such subgroup rescues of globally negative trials should be methodologically discouraged [7].

Statistical tests of interaction should be used to assess the presence of heterogeneity in treatment efficacy for different subgroups, rather than making comparisons between study treatments within each specific subgroup, which is prone to producing inappropriate claims both in terms of false negative and false positive results [8, 9]. Unfortunately, if the study of interaction was not among the original trial endpoints, analysis of interaction will often be characterized by a low statistical power. However, this could be considered and discussed in the study conclusions only if the statistical test of interaction supports a differential effect among subgroups. Even in that case, however, emphasis and claims will depend on many factors, including biological plausibility, the number of subgroup analyses performed, their prespecification and pre-planning in study protocol, and the statistical strength of evidence. How does the acknowledgment of these limitations apply to the correct interpretation of subgroup analyses performed in ICON7?

Ideally, for the reasons described above, subgroups analyses should be defined before starting the trial and should be limited to a small number of clinically important questions. In ICON7, the subgroup analysis was not planned at time of trial design, and the "high-" and "low-risk" subgroups were identified later, after the presentation of the GOG218 primary analysis. Furthermore, the definition of "high-risk" has been changed over time, between the first presentation during a symposium and the final publication, although three different definitions of "high-risk" patients were tested in the final paper with consistent results [2]. Multiple unplanned subgroups were tested, not only "high-" and "low-risk" based on residual disease, but also histology (clear cells, low grade serous), in addition to other stratification variables.

It is common practice to conduct subgroup analyses not only on the primary study endpoint, but also on other secondary endpoints. From this point of view, the ICON7 authors are not the first to present these kinds of subgroup results, and they will probably not be the last. However, once an endpoint is considered acceptable for drug approval, as it often has for PFS in the setting of first-line ovarian cancer treatment, it could be risky to use subgroup analyses of a secondary endpoint to condition the interpretation of the overall study results and to limit use of the drug in clinical practice.

From a clinical point of view, the main variable of the "high-risk" definition in ICON7 is residual disease after surgery. It is worth noting that a relevant proportion of patients (about one-third) enrolled in the GOG218 trial had a residual disease after surgery of ≤1 cm. According to the overall result of GOG218, these patients can also benefit from the addition of bevacizumab to standard chemotherapy, and there was no suggestion of a differential effect in terms of PFS according to residual disease. Of note, the hazard ratio (HR) value for disease progression or death in favor of bevacizumab was 0.717 in the total study population, 0.618 in patients with residual disease ≤1 cm, 0.763 in those with residual disease >1 cm and 0.698 in stage IV patients [1]. These findings were confirmed during a presentation at the ASCO meeting in 2015 [10], where a recent update of the *post hoc* exploratory analysis of subgroups defined by stage and extent of residual disease at diagnosis was performed for the ICON 7 trial. At prolonged follow-up, the PFS benefit associated with bevacizumab treatment observed in the ITT population was seen consistently in all subgroups explored, with a HR of 0.77 (95% confidence interval, 0.59–0.99). These results were obtained irrespective of stage and residual tumor, and therefore, also in patients categorized as having a "low-risk" tumor with the absence of any residual disease at the time of primary surgery.

Another point to consider is the reliability of residual disease as a determinant of risk status. Several groups have shown a significant discrepancy between what is reported by surgeons and what is evident at the time of a post-surgery CT

scan [11]. This is reported even in high quality centres, but it is highly likely that the magnitude of this discrepancy is even greater in a "real-life" setting. Furthermore, several countries were involved in the ICON7 study. Important regional differences in terms of surgical approach have been reported, and it is not clear how this may affect the analysis reported [12]. Taking this into consideration, it is reasonable to conclude that the residual disease variable per se is not reliable as a predictive marker, adding to the multiple methodological limitations related to subgroups analyses. In our opinion, although molecular biomarkers for the selection of patients for bevacizumab treatment are unfortunately not yet available, in the current drive to undertake personalized medicine based on molecular characterization of tumors, the "high-risk" definition is an "old-fashioned" and potentially inefficient method to select patients for any medical therapy.

In conclusion, two large randomized controlled trials have investigated the efficacy of bevacizumab in ovarian cancer patients [1, 2]; both met their primary endpoint (PFS) in the ITT analysis, which has recently been updated for ICON7 [2]. Claiming that "high-risk" patients may have OS benefit and that "low-risk" subjects do not benefit at all from treatment is risky because it is based on the results of an originally unplanned subgroup analysis of a secondary endpoint. Of course, this finding is hypothesis generating, as most subgroup analyses should be. In addition, residual disease is an unreliable variable, and it should not be used to deny an effective treatment, approved for clinical practice, to patients with advanced ovarian cancer. An excessive emphasis on the role of subgroup analyses may lead to misunderstanding of the true results generated by overall trial data and support the incorrect selection of patients to be treated with the new or experimental therapy. Caution should be used in translating results of subgroup analysis to clinical practice guidelines.

Acknowledgments

The authors thank Nicola Ryan, an independent medical writer, who provided native English editing and journal styling on behalf of HPS. This editorial assistance was funded by PharmaMar, Spain.

Conflicts of Interest

The Authors declare there are no conflicts of interest in relation to this article.

References

1. Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. N Engl J Med 2011;365(26):2473-83.
2. Oza AM, Cook AD, Pfisterer J, et al. Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. Lancet 2015;16(8):928-36.
3. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other mis(uses) of baseline data in clinical trials. Lancet 2000;255(9209):1064-9.
4. Pocock SJ, Assmann SE, Enos LE, et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting. Stat Med 2002;21(19):2917-30.
5. Zhang S, Liang F, Li W, et al. Subgroup analyses in reporting of phase III clinical trials in solid tumors. J Clin Oncol 2015;33(15):1697-702.
6. Yusuf S, Wittes J, Probstfield J, et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991;266(1):93-8.
7. Di Maio M, Perrone F. Subgroup analysis: refining a positive result or trying to rescue a negative one? J Clin Oncol 2015;33(36):4310.
8. Brookes ST, Whitley E, Peters TJ, et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess 2001;5(33):1-56.
9. Parker AB, Naylor CD. Subgroups, treatment effects and baseline risks: some lessons from major cardiovascular trials. Am Heart J 2000;139(6):952-61.
10. Gonzalez-Martin A, Oza AM, Embleton AC, et al. Exploratory outcome analyses according to stage and residual disease in the ICON7 trial of front-line carboplatin/paclitaxel (CP) ± bevacizumab (BEV) for ovarian cancer (OC). J Clin Oncol 2015;33(Suppl):abstr 5548.
11. Burger IA, Goldman DA, Vargas HA. Incorporation of postoperative CT data into clinical models to predict 5-year overall and recurrence free survival after primary. cytoreductive surgery for advanced ovarian cancer. Gynecol Oncol 2015;138(3):554-9.
12. Fairfield KM, Lucas FL, Earle CC, et al. Regional variation in cancer-directed surgery and mortality among women with epithelial ovarian cancer in the Medicare population. Cancer 2010;116(20):4840-8.