Critical Reviews in Oncology/Hematology 104 (2016) 108–114

Contents lists available at ScienceDirect

Critical Reviews in Oncology/Hematology

journal homepage: www.elsevier.com/locate/critrevonc

# Reporting adverse events in cancer surgery randomized trials: A systematic review of published trials in oesophago-gastric and gynecological cancer patients☆

Leila Meghelli[a], Fabrice Narducci[b], Christophe Mariette[c,d], Guillaume Piessen[c], Marie Vanseymortier[e], Eric Leblanc[b], Pierre Collinet[f], Alain Duhamel[a,d], Nicolas Penel[a,d,e,*]

[a] Lille Nord de France University, Medical School (Research Unit EA 2694), Lille, France
[b] Gynecological Cancer Departement, Oscar Lambret Cancer Centre, Lille, France
[c] Department of Digestive and Oncological Surgery, Claude Huriez University Hospital, Lille, France
[d] SIRIC OncoLille Consortium, Lille, France
[e] Clinical Research Unit, Oscar Lambret Cancer Centre, Lille, France
[f] Department of Gynecology, Hopital Jeanne de Flandre, Lille, France

**Contents**



ARTICLE INFO

*Article history:*
Received 13 January 2016
Received in revised form 11 April 2016
Accepted 25 May 2016

*Keywords:*
Adverse events reporting
Cancer surgery

ABSTRACT

*Background:* Few reports describe how adverse events (AEs) are reported in cancer surgery trials.
*Materials and methods:* We systematically reviewed 179 consecutive study reports issued between January 1, 1990 and November 15, 2014, which investigated surgery in oesophago-gastric (OG) or gynecologic (GY) cancer patients. Based on the reviewed reports, we assessed how AEs were reported according to CONSORT statement.
*Results:* Morbidity assessment was the primary objective of 56 studies (31.3%). Postoperative AEs were described in 161 studies (90%). Definition of AEs and grading scale (NCI-CTC AE, Dindo-Clavien scale, etc ...) were given in 27.3% and 16.8% of studies, respectively. AEs were

☆ Partly presented at the 18th ECCO/40th ESMO meeting, in Vienna, Austria – 25 Sep–29 Sep 2015 (As oral presentation; Abstract 901).

* Corresponding author at: Clinical Research Unit, Centre Oscar Lambret, 3 rue F Combemale, 59020, Lille, France.
*E-mail addresses:* n-penel@o-lambret.fr, scientifique@o-lambret.fr (N. Penel).

http://dx.doi.org/10.1016/j.critrevonc.2016.05.017

Clinical trials
Post-operative complications
Toxicity

reported by event and grade in 8.3% of studies. Definition of expectedness, seriousness, causality and safety population were present in 0.5%, 1.1%, 7.8%, and 7.2% of the studies, respectively. Reporting of AEs did not improve over time nor better in high-impact factor journals.
*Conclusion:* The reporting of AEs in cancer trials investigating surgery needs to be improved.


## 1. Introduction

Surgery remains the cornerstone of curative-intent treatment of the majority of solid tumors, even if surgery is associated with neo-adjuvant or adjuvant treatment such as, chemotherapy, hormonal therapy, external radiation therapy, brachytherapy, etc... Current randomized clinical trials (RCTs) in cancer surgery are mainly assessing the role of mini-invasive surgery and the impact of combination treatments (Diaz-Nieto et al., 2013; Galaal et al., 2012).

RCTs provide the highest level of evidence and may lead to changes in practice. The primary objective of most RCTs is the assessment of the effectiveness of the "investigational" treatment compared to recognized standard of care. However, the accurate measurement of both benefits and harms of the "investigational" treatment, whether pharmacological or not, are of major importance to properly weight the benefit-risk balance (Moher et al., 2001; Ioannidis et al., 2004). Thus, the reporting of adverse events (AEs) should be standardized, objective and reproducible in order to compare the different therapeutic approaches. The CONSORT (The Consolidated Standards of Reporting Trials) issued recommendations concerning the report of AEs in pharmacological clinical trials (Ioannidis et al., 2004). However, the description of surgical AEs significantly differs from the description of AEs in pharmacological clinical trials. To date there are no recommendations for the report of surgical AEs. To the best of our knowledge, few reports assessed the quality of AEs reporting in cancer surgery randomized trials (Blencowe et al., 2012).

Based on these facts, we have conducted a systematic literature review to measure how AEs are reported in two different clinical settings: oesophago-gastric cancer surgery and gynecological cancer surgery, both of which represent our fields of expertise. The main objective of our work was to analyze the quality of the reporting of surgical AEs in cancer surgery clinical trials. Our secondary objective was to identify factors influencing the description of AEs.

## 2. Materiel and methods

### 2.1. Selection of relevant publications and data extraction

We conducted a systematic review of the literature using the PubMed database. Selection criteria of relevant publications were as follow: randomized clinical trials (RCTs) assessing surgery in two clinical settings (of special interest for the study coordinators of our institutions) gynecological cancers patients (endometrium, cervix, ovary and vulva) and in oesophago-gastric cancers patients, with surgical procedure(s) in at least one of the treatment arms, including more than 50 patients, fully published in English and issued between 01st January 1990 and 15th November 2014. We have used several research equations taken into account the different primary locations (see supplemental Figs. 1–6 in the On-line appendix). Limits used were: randomized clinical trials, English, published between January 1, 1990 and November 15, 2014. Two expert surgeons (Pr C. Mariette and Dr F. Narducci) have reviewed, and validated the list of relevant publications to ensure its completeness. All consecutive fully published reports had been selected.

For each publication the following variables were collected: the year of publication, the journal and its impact factor (IF), the number of patients included, the study sponsor (academic or industrial), the continent of the study coordinator, type of cancer (gynecologic or oesophago-gastric), the study design (surgery A against surgery B or multimodal treatment) and the main objective of the trial (efficiency or morbidity).

### 2.2. Systematic analysis of the reporting of AEs

We have developed a grid composed of 18 items derived from the CONSORT recommendations for the report of AEs in clinical trials (Ioannidis et al., 2004). Each item was as objective as possible and coded in a binary fashion (present or absent; see Table 1).

We have analyzed each and every publication. For each publication, we have assessed the quality of the AEs reporting according the 18-items grid. Validation of the coding was done by a double blind reading of a random sample of articles by Dr. N. Penel and L. Meghelli (total of 28 articles, 5 articles by primary locations, except for vulvar cancer, all published articles (3 articles) have been reviewed by Dr N. Penel and L. Meghelli). A reconciliation was performed for any discrepancy in coding. Consensual rules were established after the double blind reading, then subsequently applied to all the publications, read and scored by L. Meghelli. The established consensual rules are as follow:

Surgical AEs were considered as described if at least an overall number of post-operative morbidity events was given in the results section of the publication.

Surgical AEs were considered as defined if there was a clear definition given by the authors (for example, definition of cellulitis or delayed gastric emptying), or if a standardized or a recognized classification (for example, NCI-CT) was given in "Materials and Methods" section or "Results" section.

"Unexpectedness" was considered as defined if there was a list of expected or usual AEs for the surgical intervention.

"Seriousness" was considered as defined if a precise definition of serious AEs was present.

Causality was considered as defined if there was at least one sentence giving a time limit to consider AEs as related to surgery (for example: AEs were considered related to the surgery if they occurred within 30 days post-operative).

Per-operative AEs were considered as described separately if they were clearly stated in the results or if the author clearly stated that there were no per-operative AEs.

The early and late AEs should be described separately in the text (with a timeframe given in the text).

Regarding the modality of AEs collection, the authors should specify who collected the information about peri operative period (clinical research nurse, blinded investigator, the surgeon ...), when the information was collected (prospective or retrospective), and the postoperative monitoring modalities.

The severity of surgical AEs was considered as "graded" if a grading scale described by the authors in the "Materials and Methods" section or a recognized grading scale was used (ex: NCI-CT, Dindo-

**Table 1**
Consort recommendations with the corresponding 18 items of the reading grid weighted by experts.

| CONSORT Recommendations | | Items of the reading grid | S | M | A |
|---|---|---|---|---|---|
| 1 | If the study collected data on harms and benefits, the title or abstract should so state | **1.** AEs mentioned in the title | 2 | 2.2 | **2.1** |
| | | **2.** AEs mentioned in the abstract | 3.3 | 3.4 | **3.3** |
| 2 | If the trial addresses both harms and benefits, the introduction should so state | **3.** AEs mentioned in the introduction | 1.9 | 2.2 | **2** |
| 3 | List addresses AEs with definitions for each (with attention, when relevant, to grading, expected vs unexpected events, reference to standardized and validated definitions, and description of new definitions) | **4.** Precise definition of AEs (NCI-CT, Dindo-Clavien. . .) | 4 | 3.5 | **3.8** |
| | | **5.** Definition of ¨(un)expectedness" | 2.6 | 3.5 | **3** |
| | | **6.** Definition of "seriousness" | 3.2 | 3.8 | **3.5** |
| | | **7.** Definition of causality | 3 | 3.7 | **3.3** |
| | | **8.** Intra operative AEs described separately | 3 | 3.8 | **3.4** |
| | | **9.** Early and late AEs reported | 3.2 | 3.8 | **3.5** |
| 4 | Clarify how harms-related information was collected (mode of data collection, timing, attribution methods, intensity of ascertainment, and harms-related monitoring and stopping rules, if pertinent) | **10**. Description of AEs collection modalities | 3 | 3.2 | **3.1** |
| 5 | Describe plans for presenting and analyzing information on harms (including coding, handling of recurrent events, specification of timing issues, handling of continuous measures, and any statistical analyses) | **11.** Use of a validated severity grading scale (NCI-CT, Dindo-Clavien. . ..) | 3.9 | 4 | **3.9** |
| | | **12.** Description of management of recurrent AEs | 2.8 | 3.3 | **3.1** |
| 6 | Describe for each arm the participant withdrawals that are a result of harms and their experiences with the allocated treatment | Not included in this analysis | | | |
| 7 | Provide the denominators for analyses on harms | **13.** AEs summarized in table(s) | 2.9 | 3.7 | **3.2** |
| | | **14.** Definition of the "safety population" | 2.9 | 4 | **3.5** |
| 8 | Present the absolute risk per arm and per AE type, grade, and seriousness, and present appropriate metrics for recurrent events, continuous variables and scale variables, whenever pertinent. | **15.** AEs reported by arm | 3.6 | 3.6 | **3.6** |
| | | **16.** AEs reported by grade | 3.6 | 3.6 | **3.6** |
| | | **17.** Exhaustive list of AEs | 3.8 | 3.4 | **3.6** |
| 9 | Describe any subgroup analyses and exploratory analyses for harms | Not included in this analysis | | | |
| 10 | Provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability, and other sources of information on harms | **18.** Discussion of harms/benefit balance | 3.4 | 4 | **3.7** |
| **Maximum achievable score** | | | 51.2 | 68.8 | **56.5** |

S: the means of items weighted by surgeons and physicians (from 0 to 4) – M: the means of items weighted by statisticians and methodologists (from 0 to 4) – A: the mean of two prior means.

Clavien or Chassagne et al.) (Dindo et al., 2004; Chassagne et al., 1993; DCTD et al., 2010).

The "safety population" was considered as defined if the authors specified on which population risks of AEs were calculated (the population who actually had surgery or not (per protocol analysis for the "safety")).

The report was considered exhaustive if a detailed and complete list of AEs was present (classified by grade or not).

A publication was considered a "pure surgical trial" if it compared two surgical techniques without neo-adjuvant or adjuvant treatment(s) (chemotherapy or radiotherapy).

We have used classical descriptive method to depict the reporting of events: numbers, percentages, 95%-confidence intervals in cases of categorical data and median, extreme values or mean and standard derivation in cases of continuous data.

### 2.3. Potential explicative factors

In order to identify potential explicative factors, we have built a score derived from the 18-items grid. Because all the 18 items could not be regarded similarly, we have submitted the grid to a committee of experts who have weighted the importance of each of the 18 items using a 4-point Likert scale: not very relevant (1 point); rather relevant (2 points); relevant (3 points) or very relevant (4 points). This panel of experts included two expert groups: (i) surgeons or physicians (experienced investigators or study coordinators), and (ii) methodologists or experts in pharmaco-vigilance. They ranked the grid independently of each other during a single round.

For each item, we calculated the average score per expert group and then the average of the two averages. For each article, the score has been calculated according to the presence of 18 potential items weighted by the experts. Using these values par item, the maximum achievable score per publication was 56.5 (Table 1).

The score was described by its mean and standard deviation and its median and extreme values.

We explored the normality of the distribution of the score graphically and by the Shapiro-Wilk test. Because the distribution of the score was not normal ($p < 0.001$),we used nonparametric tests for comparisons. We explored the link between the score and continuous data (years, IF, number of patients included) graphically and by calculating the correlation coefficient of Spearman. For

**Table 2**
Characteristics of the 179 analyzed trials.

| Characteristics | n (%) |
|---|---|
| **Years of publication** | |
| 1990–1994 | 18 (10.0) |
| 1995–1999 | 25 (14.0) |
| 2000–2004 | 35 (19.5) |
| 2005–2009 | 46 (25.8) |
| 2010–2014 | 55 (30.7) |
| **Continent** | |
| Europe | 76 (42.5) |
| Asia | 78 (43.6) |
| North America | 17 (9.5) |
| South America | 4 (2.2) |
| Oceania | 4 (2.2) |
| **Sponsor** | |
| Industrial | 5 (2.8) |
| Academic | 174 (97.2) |
| **Type of trial** | |
| "Pure" surgical trials | 101 (56.4) |
| Combined strategies | 78 (43.6) |
| **Primary objective of the trial** | |
| Morbidity | 56 (31.3) |
| **Tumor site** | |
| Oesophago-gastric | 133 (74.3) |
| Gynecological | 46 (25.7) |
| Cervix | 24 (13.4) |
| Endometrium | 14 (7.8) |
| Ovary | 7 (3.9) |
| Vulva | 3 (1.7) |
| **Number of patients included** | |
| Median (interquartile range) | 138 0-2,62) |

categorical data (type of sponsor, continent of the principal investigator, type of cancer, type of treatment and primary objective of the trial), an average comparison was performed by the Mann-Whitney *U* test.

Because hypothesized explanatory factors were tested, a Bonferroni adjusted significance level of 0.00625 was calculated to account for the increased possibility of a type-I error. Statistical analyzes were performed using SPSS 22.0 statistics software.

## 3. Results

### 3.1. Description of the relevant articles

We have selected 205 publications after reading the abstract and then 179 after reading the full text. The selection process is shown in supplemental Figs. 1–6 (see On-line appendix). The main characteristics of the publications included in the analysis are listed in Table 2.

### 3.2. Reporting of the AEs

Surgical AEs were described in 161 articles (89.9%). A standardized definition of AEs was used in 27.3% of the publications (Table 3).

**Table 4**
Quality of reporting of Adverse Events (AEs) in cancer surgical trial according to the CONSORT recommendations.

| Items of the reading grid | n (%) |
|---|---|
| **Where are mentioned AEs** | |
| AEs described in the results section | 161 (89.9) |
| AEs collection stated in the title | 27 (15.1) |
| AEs collection stated in the abstract | 128 (71.5) |
| AEs mentioned in the introduction | 103 (57.4) |
| Harm/benefit ratio in the discussion | 128 (71.5) |
| **Precise information about AEs** | |
| Definition of expectedness | 1 (0.0) |
| Definition of seriousness | 2 (0.0) |
| Definition of causality | 14 (7.8) |
| Definition of safety population | 13 (7.2) |
| Description of AE collection modalities | 39 (21.8) |
| Management of recurrent AEs | 0 |
| Intraoperative AEs reported separately | 38 (21.2) |
| **Reporting of AEs** | |
| Early and late AEs reported | 19 (10.6) |
| AEs summarized in table(s) | 129 (72.1) |
| AEs clearly reported by arm | 154 (86.0) |
| AEs reported by grade | 15 (8.3) |
| Serious AEs reported by arm | 1 (0.0) |
| Exhaustive list of AEs | 64 (35.7) |

Regarding the grading of the severity, a grading scale was used in only 16.8% of cases and description of AEs per grade was available in 8.3%.

The definitions of expectedness and seriousness were almost never present (1 and 2 trials respectively). Causality was defined only in 7.8% of the articles, the safety population in 7.2% of articles. AEs collection modalities were described in 21.8% of publications. Intraoperative AEs were separately reported in 38 publications (21.2%). Finally, AEs seemed exhaustively described in only 35.7% of the publications. (Table 4)

### 3.3. Weighted reading grid

Both experts groups ranked the 18 items of reading grid (physicians versus methodologists). We compared the distribution of the overall scores given by two expert groups; there was no significant difference ($p = 0.120$ with the Mann-Whitney *U* test). We used the Mann-Whitney *U* test for comparing the averages obtained by each item of the grid. Only the distribution of the averages for the item "definition of the safety population" was significantly different between the 2 groups: expert methodologists tend to be more stringent ($p = 0.001$).

Among the 18 items of our reading grid, we have defined the items of major importance according to the ranking done by the experts: major items are those of the 4th quartile according to the expert ranking. These major items and the rates of studies providing appropriate data for each major item were as follows: the use of a validated severity grading scale (16.8%), the use of a standardized definition (26.8%), discussion of the harms/benefit balance (71.5%), AEs reported by arm (86%), AEs reported by grade (8.4%) and exhaustive list of AEs (35.8%).

**Table 3**
Severity grading scale.

| | Trials with clear definition of used terms | Trials with specified severity grading scale |
|---|---|---|
| n (%) | 49/179 (27.3) | 30/179 (16.8) |
| NCI-CTC | 22/49 (44.9) | 15/30 (50.0) |
| Specified for events of interest | 14/49 (28.6) | 2/30 (6.7) |
| Chassagne classification | 5/49 (10.2) | 4/30 (13.3) |
| Dindo-Clavien classification | 5/49 (10.2) | 6/30 (20.0) |
| Other | 3/49 (6.1) | 3/30 (10.0) |

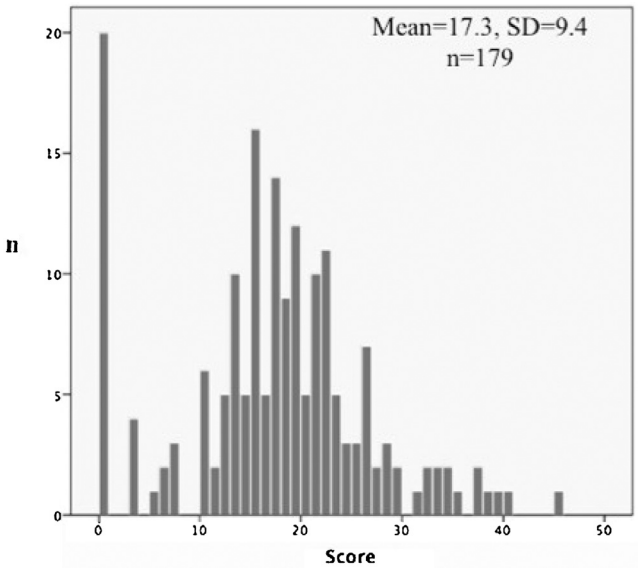


**Fig. 1.** Score distribution (SD: standard deviation).

### 3.4. Identification of explicative factors

The mean score calculated for the 179 publications was 17.3 (SD: 9.4) and the median score was 17.6 (range 0 to 45.6). Distribution of the score is presented in Fig. 1.

There was no link between the score and the publication year ($r = 0.19$, $p = 0.009$), the impact factor ($r = 0.017$, $p = 0.820$), the number of patients included ($r = 0.09$; $p = 0.210$), the continent of principal investigator or tumor location (mean score = 19.1 (DS: 10.7) for gynecological cancer versus 16.7 (DS: 8.9) for oesophago-gastric cancer $p = 0.170$)

The score was significantly higher for "pure surgical trials" compared to other types of trials (average score = 20.2 vs 13.6, $p = 0.001$) and when morbidity was the primary objective (Mean = 23.9 vs 14.3, $p < 0.001$). (Fig. 2)

### 3.5. Subgroup analysis: pure surgical trials and trials whose primary objective was morbidity

We analyzed all parameters previously tested in these 2 subgroups. Only the impact factor and the number of patients included were associated with better report of AEs in the subgroup: morbidity as primary objective ($p = 0.003$). (Table 5)

## 4. Discussion

In total, we analyzed 179 published trial reports, 133 in oesophago-gastric cancers and 46 gynecological cancers, including 43.6% of publications evaluating multimodal treatment(s). The evaluation of morbidity was the main objective in 31.3% of the published trials. We have found that 88.9% described perioperative AEs. A precise definition of AEs was present in only 26.8% of the publications. A severity grading scale was used in only 16.8% of the articles.

The average score obtained by the publications was 17.35 (SD: 9.45) for a theoretical maximum score of 56.5. There was no link between the score and the year of article issue, the journal impact factor, the number of patients included, the continent of the study coordinator or the primary tumor site. The score was significantly higher for "pure surgical trials" (20.22 vs 13.63, $p = 0.001$) and when morbidity was the primary objective (23.95 vs 14.34, $p < 0.001$).

To our knowledge, there are very few articles in the literature assessing the quality of AEs reporting in surgical RCTs in oncology. We identified only one publication assessing the quality of harms reporting in surgical oncology. This review published in 2012 by Blencowe et al. included surgical publications on esophageal cancer. They included 122 publications, including 17 RCTs. They found that 27.6% of the articles defined at least one specific surgical complication and only one publication provided a definition of all the reported complications. Only 5.1% of the analyzed publications used a standardized classification of severity (Blencowe et al., 2012).

Other studies assess harms reporting quality in surgery whatever the surgical specialty. Their results are similar to our findings: surgical AEs are poorly described in surgical trials. Sinha and al published a critical review about harm reporting in clinical trials published in highest impact factor surgical journals in 2005 (Annals of Surgery, American Journal of Transplantation and Annals of Surgical Oncology). They included 42 publications and analyzed

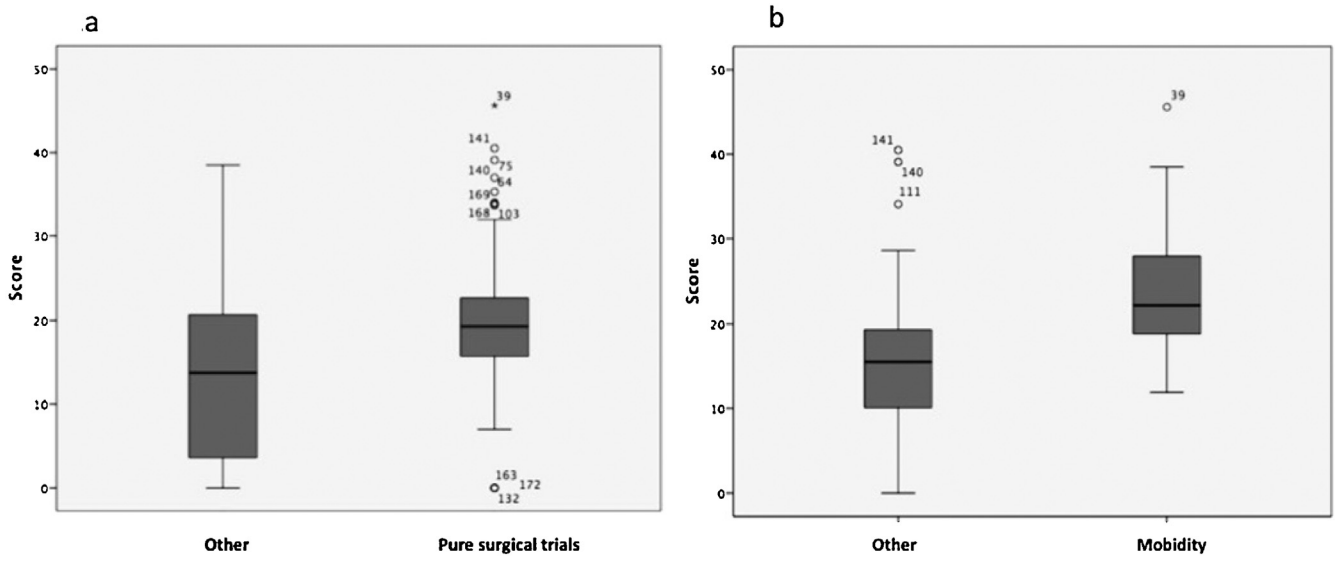


**Fig. 2.** Score distribution by type of trial (2.a) and main objective (2.b).

**Table 5**
Relation between score of articles and potential explicative factors.

| Potential explanatory factors | Pure surgical trial | Morbidity as primary objective |
|---|---|---|
| **Continuous data: correlation coefficient r (p) (Spearman test)** | | |
| Publication year | r = 0.11 (p = 0.220) | r = 0.2 (p = 0.140) |
| Impact factor | r = 0.21 (p = 0.032) | r = 0.39 **(p = 0.003)** |
| Number of patients included | r = 0.24 (p = 0.017) | r = 0.39 **(p = 0.003)** |
| **Categorical data: mean (n, p) (Mann-Whitney test)** | | |
| **Continent of investigator** | | |
| Europe (reference) | 21.14 (n = 39) | 25.13 (n = 24) |
| Asia | 19.03 (n = 54, p = 0.180) | 22.61 (n = 24, p = 0.230) |
| North America | 26.00 (n = 4, p = 0.280) | 26.00 (n = 4, p = 0.830) |
| South America | 19.30 (n = 2, p = 0.110) | 18.77 (n = 3, p = 0.150) |
| South Sea Islands | 23.65 (n = 2, p = 0.860) | 35.3 (n = 1) |
| **Tumor site** | | |
| Oesophago-gastric | 19.59 (n = 80) | 23.57 (n = 34) |
| Gynecological | 21.61 (n = 21, p = 0.110) | 24.60 (n = 21, p = 0.610) |

their compliance to the CONSORT recommendations, and calculated their Jadad Score (assessing the quality of the trial). They found that 26.8% of their publications provided a validated definition of harms. The quality of reporting was poor; 5 out of 21 non-pharmacological RCTs adequately report the adverse events (Sinha et al., 2009).

Another publication by Rosenthal et al. in 2014 included 46 consecutive clinical trials published in 2010 in the three highest impact factor journals in the fields (Annals of Surgery, surgery JAMA and the British Journal of Surgery). Rosenthal et al. have assessed the report of both per and post operative AEs in RCTs. They found that 4.3% of the articles did not describe at all AEs and 37% did not describe per-operative complications. Moreover, precise definitions of AEs were provided in 13.0% of the publications for per-operative and 50.0% for postoperative complications. A classification of severity was used in respectively 9.0% and 54.0% of the publications (Rosenthal et al., 2015).

In the urological field, few reviews asses the report of adverse events in surgical trials (Breau et al., 2010; Donat, 2007; Martin et al., 2002; Mitropoulos et al., 2012). However their results are also similar to our findings. For example, Breau and al performed a systematic literature search of all RCTs including surgical procedure(s) published in three journals of urology between 1996 and 2004. They included 152 RCT, 71.7% reported AEs, only 21.7% reported the severity of AEs, and 34.2% of the publications assessed the benefit-risk balance in the discussion (Breau et al., 2010).

We also identified one review on the report of AEs in breast reconstruction after cancer surgery published by Potter and al in 2011. They included 134 studies published between 1995 and 2009 (11 RCT, 74 cohort studies and 49 case studies). They found that 35.1% of their publications did not provide at least one complication definition and only 23.1% of their publications defined more of 75% of the complications. Severity grading system was used in 58.2% of the publications (Potter et al., 2011).

Lastly, in orthopedic literature, one systematic review published in 2009 by Goldhahn and al assessed the quality of complication reporting in 122 RCTs published between 2006 and 2007. AEs were reported in 67.0% of the publications, severity was reported in 9.3% and definitions were provided in 7.1% of RCTs (Goldhahn et al., 2009).

All theses reviews constantly found that AEs were poorly reported in RCTs including surgical treatment, whatever the medical field.

Our study had some limitations. We included publications about two tumor sites only (pelvic gynecologic oncology and gastro-esophageal oncology), which may not be representative of the rest of surgical oncology specialties. However, these two specialties are quite different and the quality of the surgical AEs reporting did not differ significantly between these two subgroups. Besides, the quality of reporting of AEs in other specialties (literature search on urologic, orthopedic and breast surgery) is similar to our review.

We used a single database for our systematic literature review (PubMed) and search by key words was difficult because of a poor indexing of the publications. We have thus used several research equations; two experts checked the completeness of our review; finally; references of each publication were also checked.

The coding of EA reporting was not blind about journals and authors. However, this seems hardly feasible and we had no conflicts of interest. Moreover, we did not see any improvement of the score in high-impact factor journals. Only one reviewer proceeded to data extraction; this could be seen as a limitation, nevertheless, the 28 first studies have been assessed by 2 reviewers in a blinded fashion, with a secondary reconciliation in case of discrepancy. Furthermore, the items are binary (present and absent) and could be regarded as much as objective as possible.

We included in our analysis several types of publications. Some assessed the effectiveness of treatments, other morbidity; some evaluated multimodal therapies and other only surgery. This causes a great heterogeneity that may affect AEs reporting. To overcome this bias, subgroup analyzes were performed. We have limited our analysis to two different surgical settings, but the literature review was exhaustive.

Appropriate reporting of surgical AEs in cancer RCTs needs a collective initiative of both physicians and methodologists. Some traditional items used in pharmacological trials needs to be formally defined in surgical trials, if relevant: definition of safety population in surgical trials. The use of appropriate definition of surgical AEs and inherent grading scale is mandatory. In our review, we found that several grading systems and definitions are used (Dindo-Clavien, NCI-CTC, Chassagne.) without consensus on the better method for defining the AEs. We found that in most trials, the description of AEs was truncated and focused on some events of interest that are rarely formally defined. In these cases, the assessment of the harm could not be properly done.

## 5. Conclusion

According to the CONSORT recommendations, AEs are poorly reported in surgery cancer trials. This reporting did not improve over time. This reporting is inaccurate even if the study is published in a high impact factor journal.

Specific recommendations about surgical AEs reporting should be developed. An international and collective effort is urgently needed.

## Conflicts of interest

There are no financial disclosures from the authors for the manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.critrevonc.2016.05.017.

## References


Blencowe, N.S., Strong, S., McNair, A.G.K., Brookes, S.T., Crosby, T., Griffin, S.M., et al., 2012. Reporting of short-term clinical outcomes after esophagectomy: a systematic review. Ann. Surg. 255 (April (4)), 658–666.

Breau, R.H., Gaboury, I., Scales, C.D., Fesperman, S.F., Watterson, J.D., Dahm, P., 2010. Reporting of harm in randomized controlled trials published in the urological literature. J. Urol. 183 (May (5)), 1693–1697.

Chassagne, D., Sismondi, P., Horiot, J.C., Sinistrero, G., Bey, P., Zola, P., et al., 1993. A glossary for reporting complications of treatment in gynecological cancers. Radiother. Oncol. 26 (March (3)), 195–202.

Diaz-Nieto, R., Orti-Rodríguez, R., Winslet, M., 2013. Post-surgical chemotherapy versus surgery alone for resectable gastric cancer. In: Cochrane Database of Systematic Reviews [Internet]. John Wiley & Sons, Ltd, Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008415.pub2/abstract [cited 2015 Aug 31].

Dindo, D., Demartines, N., Clavien, P.A., 2004. Classification of surgical complications – a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann. Surg. 240 (August (2)), 205–213.

Donat, S.M., 2007. Standards for surgical complication reporting in urologic oncology: time for a change. Urology 69 (February (2)), 221–225.

DCTD, NCI, NIH, DHHS. Common Terminology Criteria for Adverse Events, Version 4.03. 2010.

Galaal, K., Bryant, A., Fisher, A.D., Al-Khaduri, M., Kew, F., Lopes, A.D., 2012. Laparoscopy versus laparotomy for the management of early stage endometrial cancer. In: Cochrane Database of Systematic Reviews [Internet]. John Wiley & Sons, Ltd [cited 2015 Aug 31].

Goldhahn, S., Sawaguchi, T., Audigé, L., Mundi, R., Hanson, B., Bhandari, M., et al., 2009. Complication reporting in orthopaedic trials: a systematic review of randomized controlled trials. J Bone Joint Surg. Am. 91 (August (8)), 1847–1853.

Ioannidis, J.P.A., Evans, S.J.W., Gøtzsche, P.C., O'Neill, R.T., Altman, D.G., Schulz, K., et al., 2004. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann. Intern. Med. 141 (November (10)), 781–788.

Moher, D., Schulz, K.F., Altman, D., 2001. The consort statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 285 (April (15)), 1987–1991.

Martin, R.C.G., Brennan, M.F., Jaques, D.P., 2002. Quality of complication reporting in the surgical literature. Ann. Surg. 235 (June (6)), 803–813.

Mitropoulos, D., Artibani, W., Graefen, M., Remzi, M., Rouprêt, M., Truss, M., et al., 2012. Reporting and grading of complications after urologic surgical procedures: an ad hoc EAU guidelines panel assessment and recommendations. Eur. Urol. 61 (February (2)), 341–349.

Potter, S., Brigic, A., Whiting, P.F., Cawthorn, S.J., Avery, K.N.L., Donovan, J.L., et al., 2011. Reporting clinical outcomes of breast reconstruction: a systematic review. J. Natl. Cancer Inst. 103 (January (1)), 31–46.

Rosenthal, R., Hoffmann, H., Dwan, K., Clavien, P.-A., Bucher, H.C., 2015. Reporting of adverse events in surgical trials: critical appraisal of current practice. World J. Surg. 39 (January (1)), 80–87.

Sinha, S., Sinha, S., Ashby, E., Jayaram, R., Grocott, M.P.W., 2009. Quality of reporting in randomized trials published in high-quality surgical journals. J. Am. Coll. Surg. 209 (November (5)), 565–71.e1.