



# Evidence-based recommendations for blinding in surgical trials

Pascal Probst<sup>1,2</sup> · Steffen Zschke<sup>1,2</sup> · Patrick Heger<sup>1,2</sup> · Julian C. Harnoss<sup>1,2</sup> · Felix J. Hüttner<sup>1,2</sup> · André L. Mihaljevic<sup>1,2</sup> · Phillip Knebel<sup>1,2</sup> · Markus K. Diener<sup>1,2</sup>

Received: 17 December 2018 / Accepted: 5 February 2019 / Published online: 1 March 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

**Rationale** Blinding reduces performance and detection bias in randomized controlled trials (RCT). There is evidence that lack of blinding leads to overestimation of treatment effects in pharmacological trials. Since surgical trials use interventions with a physical component, blinding is often complicated. The aim of this study was to analyze, in general and abdominal surgery RCT, the status of blinding, the potential for blinding, and the influence of blinding on outcomes.

**Methods** A systematic search of the literature in CENTRAL, MEDLINE, and Web of Science was conducted to identify RCT with a surgical intervention, starting in 1996, the year when the first CONSORT statement was published. Information on general study characteristics and blinding methods was extracted. The presence or absence of blinding of the study contributors—patients, surgeons, data collectors, outcome assessors, and data analysts—was analyzed. The association of blinding with the trial outcome was investigated for every study contributor.

**Results** Out of 29,119 articles, 378 RCT were included in the analysis. These investigated a total of 62,522 patients, of whom 15,025 were blinded (24.0%). Contributors could have been blinded in far more trials, as the potential for blinding measures ranged from 69% for outcome assessors to 98% for data analysts. If blinding of surgeons would have been possible but had not been performed, this was associated with more significant trial outcomes (OR 13.670; 95% CI 1.308 to 142.840;  $p = 0.0289$ ).

**Discussion** The potential of blinding, an important quality measure in surgical RCT, has not been exhausted. This study summarizes the existing evidence on blinding in surgical trials and gives evidence-based recommendations for the use of blinding in future surgical trials.

**Systematic review registration** PROSPERO 2015:CRD42015026837.

**Keywords** Blinding · Bias · Performance bias · Detection bias · Surgery · Risk of bias · Randomized controlled trial · Systematic review · Cochrane risk of bias assessment tool

## Background

The aim of evidence-based medicine is to identify the optimal treatment for patients. This process is based on the expertise of

the treating practitioner, the patients' characteristics, and the best available external evidence [1]. Properly performed randomized controlled trials (RCT) are considered to provide the best external evidence by virtue of measures to minimize bias and to ensure the validity of the conclusions drawn [2, 3]. Bias is a systematic error that leads to deviation of the measured effect from the true effect of an intervention [4]. The CONSORT statement (Consolidated Standards of Reporting Trials), a guideline on the reporting of outcomes in RCT, declared information enabling evaluation of bias to be mandatory [5].

One measure that can be used to reduce bias is blinding of study contributors. Generally, RCT have five kinds of contributors: patients, practitioners (here surgeons), data collectors, outcome assessors, and data analysts (Fig. 1). The risk that awareness of the applied intervention influences its effects is called performance or detection bias. A patient or surgeon

---

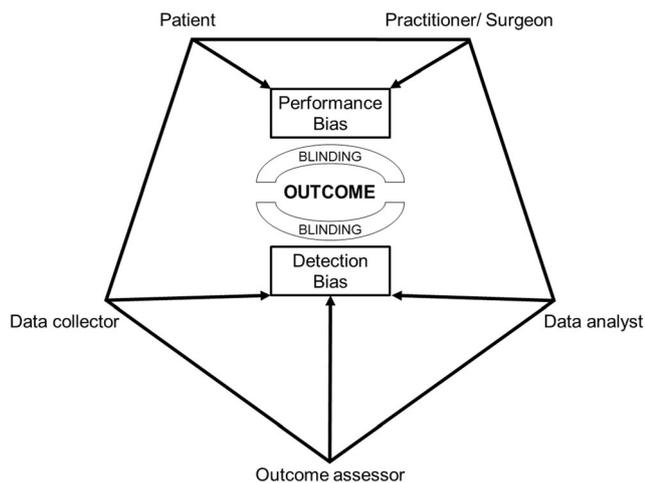
**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00423-019-01761-6>) contains supplementary material, which is available to authorized users.

---

✉ Pascal Probst  
pascal.probst@med.uni-heidelberg.de

<sup>1</sup> Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany

<sup>2</sup> The Study Center of the German Surgical Society (SDGC), University of Heidelberg, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany



**Fig. 1** Study contributors in a trial may generate performance and detection bias, which can be avoided through blinding

who trusts in the efficacy of a specific intervention may unconsciously or intentionally perceive or detect an enhanced treatment effect [4]. Blinding of both participants and personnel is commonly referred to as “double blinding” [6]. Detection bias refers to the risk of how the evaluation of the outcomes biases effects. Outcome assessors and data analysts (investigators or statisticians) who are aware of the treatment used may unconsciously or intentionally alter their assessment. Particularly in the case of soft endpoints, e.g., pain or quality of life, blinding of outcome assessors is important. For objective outcomes, such as mortality, detection bias is irrelevant [4, 7].

The physical component of the interventions in surgical RCT means that blinding of the operating surgeon is usually not possible. Blinding of patients and outcome assessors is also not readily achieved [3, 4, 8, 9]. Various blinding methods have been investigated for a number of non-pharmacological treatments, yielding detailed methodological information on the potential extent of blinding in surgery [10]. However, it remains unclear whether ornate blinding measures for surgical interventions are justified by improved quality of evidence. It is very probable that measures to blind patients are negated by the physical component of the treatment, and this aspect has never been systematically investigated.

Therefore, the objective of this systematic review and empirical study was to investigate the impact of blinding in general and abdominal surgery RCT and give evidence-based recommendations for its future use in surgical trials.

## Methods

This systematic review and empirical study was reported as specified in the PRISMA statement [11]. The study is

registered under PROSPERO 2015:CRD42015026837 and was conducted according to a predefined protocol [12].

## Research question

This study aimed to determine the present status of blinding in general and abdominal surgery RCT, its potential application in such trials, and its influence on outcomes. The present status of blinding in general and abdominal surgery was evaluated by quantification of blinding measures and their reporting in RCT since 1996. The potential for blinding in surgical RCT was determined by comparing the possible blinding measures with those actually used. Finally, the influence of blinding on outcomes was evaluated by investigating trials for the association of blinded study contributors with significant trial outcomes.

## Systematic literature search

A research question was formulated according to the PICO model (Participants, Interventions, Comparisons, Outcomes) [4]: P: adult human patients undergoing general and abdominal surgery; I: any surgical intervention; C: any surgical control; O: method of blinding and respective outcome of primary endpoint. The following search strategy was used:

```
(((((("General Surgery"[Mesh] OR "general surgery"[tiab] OR "Colorectal Surgery"[Mesh] OR "colorectal surgery"[tiab] OR "Digestive System Surgical Procedures"[Mesh] OR "gastrointestinal surgery"[tiab] OR "rectum surgery"[tiab] OR "rectal surgery"[tiab] OR "visceral surgery"[tiab] OR "Liver/surgery"[Mesh] OR "liver surgery"[tiab] OR "hepatic surgery"[tiab])) OR ((pancreatic OR pancreas) AND (surgery[tiab] OR surgeries[tiab] OR surgic*[tiab]))) OR ((("Esophagus/surgery"[Mesh]) OR (esophagus AND (surgery[tiab] OR surgeries[tiab] OR surgic*[tiab]))) OR ((transplant*[tiab] AND (surgery[tiab] OR surgeries[tiab] OR surgic*[tiab]))) OR ("hepatobiliary-pancreatic"[tiab] AND (surgery[tiab] OR surgeries[tiab] OR surgic*[tiab]))) OR ("Retroperitoneal Neoplasms/surgery"[Mesh] OR "Abdominal Neoplasms/surgery"[Mesh] OR "abdominal surgery"[tiab] OR "gastric surgery"[tiab] OR "bariatric surgery"[tiab] OR "bariatric surgery"[mesh] OR "splenectomy"[mesh] OR splenectomy[tiab] OR splenectomies[tiab] OR "Thyroidectomy"[Mesh] OR thyroidectomy[tiab] OR thyroidectomies[tiab] OR "Cholecystectomy"[Mesh] OR cholecystectomy[tiab] OR cholecystectomies[tiab]))) AND ((randomized controlled trial [pt]) OR random*[tiab]). The search strategy is described in full in the
```

## Box 1 Recommendations for future use of blinding in surgical trials

**Recommendations for future use of blinding in surgical trials**

There is an unused high potential for blinding in surgical trials. Therefore, during the planning phase of a trial the following questions should be answered for every study contributor (patient, surgeon, data collector, outcome assessor and data analyst):

- **Is the primary endpoint influenceable by the [study contributor] or is the endpoint completely robust against unconscious or intentional influence?**

Endpoints like pain or quality of life are prone to unconscious or intentional altering by unblinded patients. Outcome assessors may unconsciously or intentionally more manipulate their judgement the less exact the assessment of an endpoint is defined.

- **How can the [study contributor] be blinded to the intervention until the primary endpoint is assessed?**

The blinding measures taken should be explained for every study contributor in the trial protocol.

- **How ornate are the blinding methods that would be applied in relation to the expected effect to avoid bias?**

At some point avoidance of bias cannot be justified by very ornate blinding measures which interrupt the general conduct of the trial or even arouse desires to “outsmart” the blinding. If this is the case it should be described in detail in the trial protocol.

- **The operating surgeon seems to be the gateway to bias. If not blinding, the operating surgeon should not be involved in any decisions that could influence the primary endpoint or in judgement of the primary endpoint.** Whether this is the case and how it is achieved should be described in the trial protocol.
- In the presence of a predefined and published data analysis plan, blinding of the data analyst may be dispensable. However, there are not many conceivable circumstances in which blinding of the data analyst would pose any major problems.
- **The efficacy of blinding measures in surgical trials should be assessed.** All blinded study contributors should be asked whether unintentional unblinding occurred. If they are still blinded, they should be asked which arm of the study they think they are part of, giving their reasons. These guesses should be compared to the actual allocation.

published protocol [12]. The following databases were searched: CENTRAL, MEDLINE (via PubMed), and Web of Science [13]. A search strategy based on vocabulary thesaurus (e.g., MeSH) in combination with text words was used. The survey was limited to the period from August 1996, when the first CONSORT statement [14] was published. The last search was performed on December 2015. No language restrictions were applied.

## Study selection

Abstracts and full texts were screened independently by two reviewers following the recommendations of the Cochrane Collaboration [4]. Articles identified by the systematic literature search were screened for eligibility. RCT from general and abdominal surgery with a surgical intervention in adult human patients were included in full-text screening. A surgical intervention is characterized by a skin incision and a dominant physical component and/or modification of anatomy. Good examples of surgical trials are comparisons of two surgical accesses (e.g., open vs. laparoscopic), surgical strategies (e.g., Bassini vs. Shouldice in groin surgery), or two ways of performing a specific resection (e.g., hand vs. stapler anastomosis of bowels). Trials that investigated surgical training, compared a surgical with a nonsurgical intervention (e.g., pharmacology or irradiation), compared early and late surgery, or comprised more than two study arms were excluded.

Full-text screening was performed for all articles eligible after abstract screening. To ensure homogeneity of the sample [12], all trials were checked for proper randomization and allocation (selection bias) according to the recommendations of the Cochrane Collaboration [4]. Furthermore, only RCT with an a priori defined primary endpoint (reporting bias) and an intention-to-treat (ITT) analysis or without unexplained drop-out (attrition bias) were included in quantitative analysis. Exclusion of trials without proper randomization and allocation prevents a high risk of selection bias [4]. An a priori-defined primary endpoint is necessary in order to obtain a sample free of trials that feature changed, newly introduced, or omitted primary endpoints and are therefore at high risk of reporting bias [15]. An endpoint is defined a priori if there is an open accessible protocol or registration defining the endpoint as reported or if the primary endpoint is based on a sample size calculation. Including only trials with ITT analysis (with less than 10% imputed/drop-out data) or without unexplained drop-out prevents high risk of bias due to incomplete outcome data (attrition bias) [4].

## Data extraction

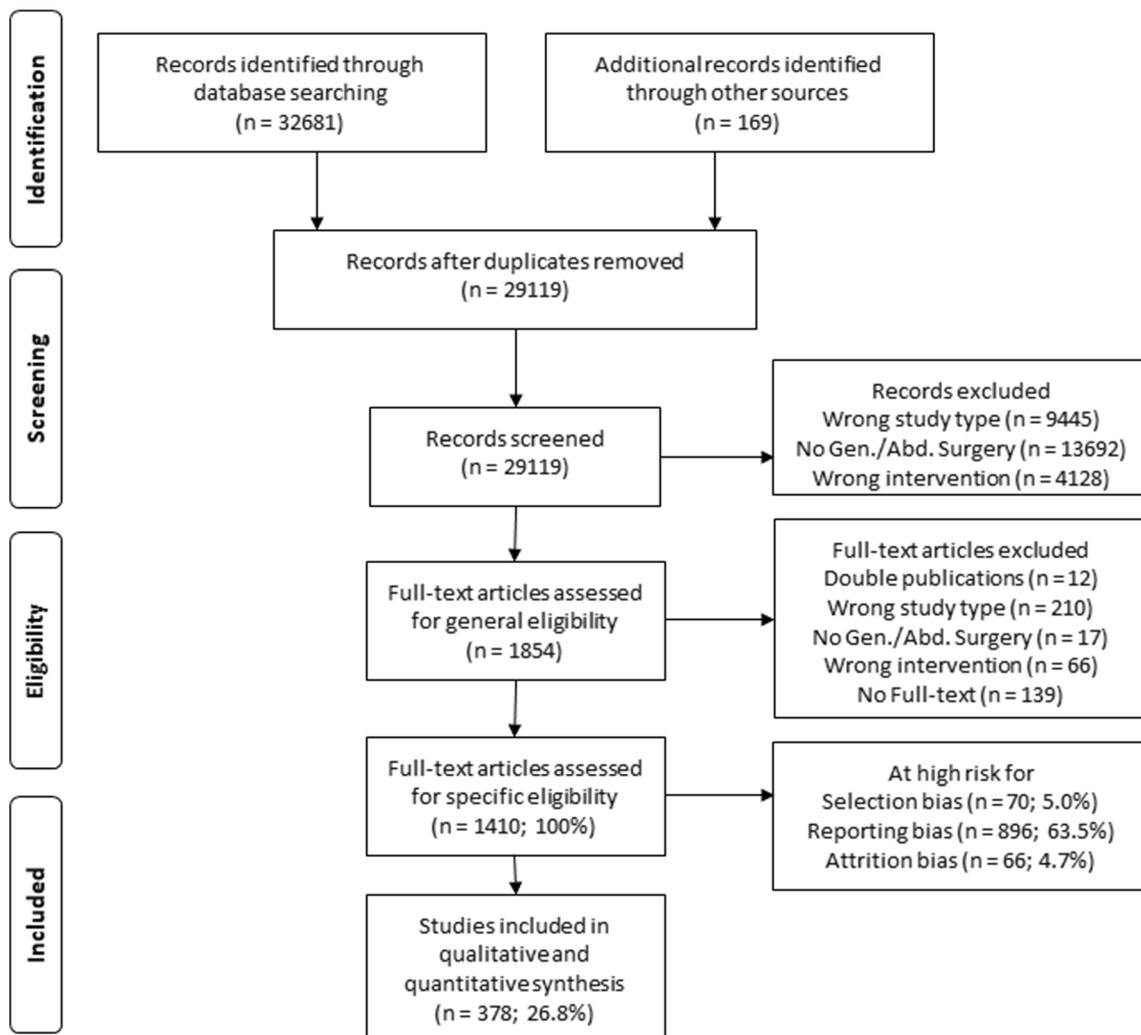
The following study characteristics were extracted: title, author, year of publication, journal, surgical speciality (upper gastrointestinal surgery, hepato-pancreatico-biliary surgery, colorectal surgery/proctology, endocrine surgery, hernia, mixed, other), study intervention and control intervention, primary endpoint, and outcome (subjective/objective). Details of blinding were extracted from the published trials or protocols: position in article where blinding was mentioned (title, abstract, methods section, or separate protocol/ registration); presence, absence, or non-reporting of blinding of trial contributors (patients, practitioners (surgeons), data collectors, outcome assessors, data analysts); feasibility of blinding trial contributors according to Boutron et al. [10]; whether the influence of lack of blinding was discussed; and whether possible unblinding was assessed during the trial. Further sources of bias, such as industrial funding [16], number of study centers, and number of nations included, were also extracted [16]. The original data extraction sheet is published with the protocol [12].

Data extraction was performed by two reviewers independently for quality assurance purposes [17]. Discrepancies between the two reviewers were resolved by a third reviewer, and a final extraction sheet was determined for database entry. After the last extraction sheet was entered into the database, the latter was closed and made available for statistical analysis.

## Data synthesis and statistical analysis

To evaluate the status of blinding in surgical RCT, the rates of blinded trials and blinded study contributors were expressed descriptively overall and for three periods (1996–2000; 2001–2009; 2010–2015). These periods were determined by the publication of the different versions of the CONSORT statements in 1996, 2001, and 2010. Moreover, for each study contributor, it was investigated whether blinding was feasible and not done, feasible and done, not feasible, or unclear. For these groups, a two-sample test for equality of proportions was performed, comparing significant and non-significant trial results with the level of significance set at 5%. The potential for blinding was expressed as the comparison of trials in which blinding was actually carried out with trials where blinding would have been feasible. To determine whether blinding of a study contributor would have been theoretically possible, the study by Boutron et al. [10] was used.

It is known from other medical fields that lack of blinding leads to more significant trial results. Therefore, the association of blinding status with



significant trial outcome was tested in a multivariable logistic regression model at a level of significance of 5%. The target factor was the outcome of the included RCT (significant vs not significant outcome for primary endpoint). The influencing factor was blinding status (blinded vs unblinded) of the study contributors. Further factors and variables were the international vs national nature of the included RCT, the time period

(1996–2000 vs 2010–2015, 2001–2009 vs 2010–2015), and other potential threats to validity (non-industry vs industry, baseline imbalance vs balanced, number of study centers, and sample size). The associations were expressed as odds ratio (OR) with 95% confidence interval. Within the multivariable logistic regression model, a significantly raised OR would indicate an influence of lack of blinding on trial outcomes in surgery.

**Table 1** Blinded study contributors overall and in the three separate study periods

Blinded study contributors	Study period			
	1996 to 2000 (n = 34)	2001 to 2009 (n = 139)	2010 to 2015 (n = 205)	Total (n = 378)
Patients	2 (5.9%)	25 (18.0%)	66 (2.2%)	93 (24.6%)
Surgeons	0 (0.0%)	2 (1.4%)	5 (2.4%)	7 (1.9%)
Data collectors	0 (0.0%)	15 (10.8%)	19 (9.3%)	34 (9.0%)
Outcome assessors	2 (5.9%)	38 (27.3%)	57 (27.8%)	97 (25.7%)
Data analysts	0 (0.0%)	4 (2.9%)	4 (2.0%)	8 (2.1%)

**Table 2** Feasibility for blinding and actual blinding of study contributors

	<i>n</i> ; %	Significant (%)	Non-significant (%)	<i>p</i> value
<b>Patients</b>				
Feasible and not done	63 (17%)	30 (48%)	33 (52%)	0.792
Feasible and done	93 (25%)	48 (52%)	45 (48%)	0.825
Not feasible	34 (9%)	15 (44%)	19 (56%)	0.599
Blinding unclear	188 (50%)	91 (48%)	97 (52%)	0.674
<b>Surgeons</b>				
Feasible and not done	8 (2%)	4 (50%)	4 (50%)	> 0.999
Feasible and done	7 (2%)	0 (0%)	7 (100%)	0.023
Not feasible	153 (40%)	71 (46%)	82 (54%)	0.365
Blinding unclear	210 (56%)	109 (52%)	101 (48%)	0.570
<b>Data collectors</b>				
Feasible and not done	76 (20%)	36 (47%)	40 (53%)	0.717
Feasible and done	34 (9%)	11 (32%)	23 (68%)	0.054
Not feasible	37 (10%)	17 (46%)	20 (54%)	0.736
Blinding unclear	231 (61%)	120 (52%)	111 (48%)	0.528
<b>Outcome assessors</b>				
Feasible and not done	73 (19%)	36 (49%)	37 (51%)	> 0.999
Feasible and done	97 (26%)	46 (47%)	51 (53%)	0.664
Not feasible	27 (7%)	13 (48%)	14 (52%)	> 0.999
Blinding unclear	181 (48%)	89 (49%)	92 (51%)	0.865
<b>Data analysts</b>				
Feasible and not done	115 (30%)	52 (45%)	63 (55%)	0.311
Feasible and done	8 (2%)	4 (50%)	4 (50%)	> 0.999
Not feasible	4 (1%)	3 (75%)	1 (25%)	0.616
Blinding unclear	251 (66%)	125 (50%)	126 (50%)	0.938

All statistical analysis was performed with the program R [18].

## Results

After deleting duplicates, 29,119 articles were identified by the initial search. After abstract screening, 1854 publications remained for full-text screening. Of these, 444 articles were excluded on grounds of duplicate publication, wrong study type, or wrong intervention. This resulted in a comprehensive set of 1410 surgical RCT (100%).

Seventy (5.0%) of the full texts were excluded because of selection bias, 896 (63.5%) due to reporting bias, and 66 (4.7%) for reasons of attrition bias. Finally, 378 RCT (26.8%) were included in the qualitative and quantitative analysis. Figure 2 shows the PRISMA flow chart.

### Characteristics of the included RCT

A total of 62,522 patients were investigated in the 378 RCT, of whom 15,025 were blinded (24.0%). The highest number of trials investigated hepato-pancreatico-biliary surgery (121 of 378 RCT; 32.0%) followed by colorectal

**Table 3** Potential for blinding of study contributors

	Theoretically blindable (of 378 RCT)	Actually blinded (if blindable)	Potential <sup>a</sup> (not blinded)
Patients	315 (83%)	93 of 315 (30%)	70%
Surgeons	28 (7%)	7 of 28 (25%)	75%
Data collectors	276 (73%)	34 of 276 (12%)	88%
Outcome assessors	313 (83%)	97 of 313 (31%)	69%
Data analysts	370 (98%)	8 of 370 (2%)	98%

<sup>a</sup> Difference between 100% and rate of actually blinded patients

**Table 4** Multivariable analysis of the impact of blinding on significant trial results

Target factor	Significant study outcome	
	Odds ratio (95% confidence interval)	<i>p</i> value
Factors		
All studies at low risk of selection bias		
All studies at low risk of attrition bias		
All studies at low risk of reporting bias		
Patients		
Unblinded vs blinded	0.697 (0.332 to 1.466)	0.3416
Surgeons		
Unblinded vs blinded	13.670 (1.308 to 142.840)	0.0289
Data collectors		
Unblinded vs blinded	2.267 (0.578 to 8.882)	0.2403
Outcome assessors		
Unblinded vs blinded	1.296 (0.527 to 3.192)	0.5722
Data analysts		
Unblinded vs blinded	0.593 (0.099 to 3.540)	0.5661
Geographical scope of trials		
International vs single country	2.790 (1.078 to 7.220)	0.0345
Year		
1996–2000 vs 2010–2015	0.592 (0.256 to 1.372)	0.2216
2001–2009 vs 2010–2015	0.802 (0.486 to 1.323)	0.3876
Other potential threats to validity		
Non-industry vs. industry	1.277 (0.564 to 2.893)	0.558
Baseline imbalances vs balanced	0.799 (0.396 to 1.609)	0.529
Variables		
Number of study centers	0.950 (0.904 to 0.998)	0.0395
Sample size	0.998 (0.997 to 1.000)	0.0614

surgery or proctology (101 of 378 RCT, 26.7%), upper gastrointestinal surgery (80 of 378 RCT; 21.2%), endocrine surgery (30 of 378 RCT, 8.0%), mixed operations (24 of 378 RCT; 6.4%), and hernia surgery (16 of 378 RCT; 4.2%). Almost half of the trials investigated surgical strategies (183 of 378 RCT; 48.4%). The other half compared surgical instruments (118 of 378 RCT; 31.2%) or different surgical accesses (76 of 378 RCT; 20.1%). The primary endpoints of the included trials were considered to be subjective in 61.1% of cases (231 of 378 RCT) and objective in 38.9% (147 of 378 RCT).

### Reporting of blinding

Blinding was mentioned in the titles of 20 of 378 RCT (5.3%), in the abstracts of 54 of 378 RCT (14.7%) and in the methods sections of 190 of 378 RCT (50.3%). A protocol or registration was available in 128 of the 378 RCT (33.9%), and blinding was mentioned in the protocol in 112 of 128 RCT (87.5%). In 41 of 378 RCT (10.8%), the absence or presence of blinding measures was mentioned

only in the protocol, not in the publication giving the results.

The blinding was characterized in different ways: single-blinded (34 of 378 RCT; 9.0%), double-blinded (36 of 378 RCT; 9.5%), and blinded or masked (71 of 378 RCT; 18.8%). “Double-blinded” meant either that at least two types of contributors were blinded or that both detection and performance bias were avoided. In 93 of 378 RCT (24.6%), various expressions were used to state that the trial was not blinded, e.g., open-label, not masked, informed, or aware. In 14 (15.1%) of these 93 “open-label” trials, the lack of blinding measures was discussed as a source of bias. In one of 378 RCT, the occurrence of unblinding of patients was investigated (0.3%).

### Status of blinding

Of the 378 RCT, 34 (9.0%) were published between 1996 and 2000, 139 (36.8%) between 2001 and 2009, and 205 (54.0%) between 2010 and 2015. The proportion of all types of study contributors who were blinded

increased between the period 1996 to 2000 and the period 2001 to 2009. However, comparing the periods of 2001 to 2009 and 2010 to 2015, only the proportion of blinded patients increased substantially, from 18 to 32.2%. For all other study contributors, the proportions remained similar or decreased. The proportions of overall blinded study contributors and for the defined periods are shown in Table 1.

Blinding of patients was considered not feasible in 34 (9.0%) of the 378 included RCT, and reporting was unclear in 188 of 378 RCT (49.7%). Blinding was feasible and done in 93 RCT (24.6%) and feasible but not done in 63 RCT (16.7%). For each of these groups, no difference regarding significant trial outcomes was present. For surgeons, blinding was considered not feasible in

153 of 378 RCT (40.5%), and reporting was unclear in 210 of 378 RCT (55.6%). Only in 7 of 378 RCT (1.9%) blinding was feasible and done, and in 8 of 378 (2.1%) RCT, blinding was feasible but not done. All trials in which the surgeon was blinded for the intervention resulted in non-significant results for the primary endpoint. As for the blinding of data collectors, this was considered feasible and done in 9.0% of trials (34 of 378 RCT) and feasible but not done in 20.1% of trials (76 of 378 RCT). In the majority of trials (231 of 378 RCT; 61.1%), the blinding status of data collectors remained unclear, and in 9.8% of trials (37 of 378 RCT), blinding of data collectors was not feasible. Blinding of outcome assessors was considered not feasible in 27 of 378 RCT (7.1%), and reporting was unclear in 181 of 378 RCT

#### Box 2 Recommendations for reporting of blinding in surgical trials

##### **Recommendations for reporting of blinding in surgical trials**

- **Terms like “double-blinded” should be avoided, as they are inadequately defined, both in general and specifically for surgical trials.** The simple term blinded in the title and the abstract section with a comprehensive description of blinding measures in the methods section should be favored to imprecise catchwords.
- **For every study contributor (patient, surgeon, data collector, outcome assessor and data analyst) it should be reported whether blinding measures were applied.** To avoid an unclear status regarding blinding measures a short and simple statement for every study contributor is necessary. If there were reasons for not applying blinding methods (please see: recommendations for future use of blinding in surgical trials) this should be mentioned in the limitations section referencing the trial protocol. Moreover, the potential deviation of the measured effect from the true effect by lack of blinding should be discussed.
- **If not blindingable, the role of the operating surgeon in the postoperative care should be described in detail. For outcome assessors and data analysts it should be outlined whether there were either predefined definitions for assessment or an a priori defined analysis plan.** This information will allow the reader to estimate the potential impact of respective study contributors on trial outcomes by performance or detection bias.

## Box 3 Recommendations for assessing blinding in surgical systematic reviews

**Recommendations for assessing blinding in surgical systematic reviews**

- There are some specifics when assessing blinding in surgical trials for a systematic review compared to other medical fields. These specifics are all related to the **outstanding role of the operating surgeon** applying the intervention. For none of the other study contributors an introduction of bias on trial outcomes could be shown by this study. Therefore, important information is lost by summarizing the absence of blinding in performance and detection bias only. **For surgical systematic reviews it is recommended to describe the blinding measures and their impact on endpoints for every study contributor individually.**
- **The plausibility of maintaining the study contributors blinded should be assessed.** This includes time from the index operation to the assessment of the primary endpoint and how ornate blinding measures were. If authors present an analysis of efficacy of blinding measures (please see: recommendations for future use of blinding in surgical trials), this is a strong argument that no bias due to deviations from intended interventions is present.
- Risk of bias should be assessed as high irrespective of the possibility of blinding of the intervention. More important is the potential of each study contributor to introduce bias into the study by lack of blinding. **If a surgical trial uses objective endpoints not prone to performance or detection bias the risk of bias for these domains should generally be judged as low.**

(47.9%). In 73 of 378 RCT (19.3%), blinding was feasible but not done, and in 97 of 378 (25.7%), feasible and done. For data analysts, blinding was considered not feasible in 4 of 378 RCT (1.1%), and reporting was unclear in 251 of 378 RCT (66.4%). Blinding was feasible and done in 8 RCT (2.1%) and feasible but not done in 115 RCT (30.4%). Comparing the subgroups for data collectors, outcome assessors, and data analysts, no differences regarding significant trial outcomes were present. Table 2 summarizes the results for the different types of study contributors.

**Potential of blinding**

Using the methods described by Boutron et al. [10], blinding of the study contributors would have been possible in most of the included RCT, except for practitioners, i.e., the operating

surgeons (28 of 378 RCT; 7.4%). Patients, data collectors, outcome assessors, and data analysts were theoretically blinding to the surgical intervention in 73 to 98% of the RCT (Table 3). Among these RCT with theoretically blinding surgical interventions, the lowest rate of actual blinding was seen for data analysts (8 of 370 RCT; 2.1%) and the highest rate for outcome assessors (97 of 313 RCT; 31.0%). The unused potential for blinding in surgical RCT was 70% for patients, 75% for surgeons, 88% for data collectors, 69% for outcome assessors, and 98% for data analysts.

**Influence of blinding on outcomes**

The influence of blinding on the outcomes of the primary endpoint of the included RCT was calculated as the association with blinded study contributors (Table 4). Of the five types of study contributor, only the practitioner, i.e., the

unblinded surgeon, was independently associated with significant trial results (OR 13.670; 95% CI 1.308 to 142.840;  $p = 0.0289$ ). Of the other factors with a possible influence on trial results, a higher number of study centers within an RCT were independently associated with less significant outcomes (OR 0.950; 95% CI 0.904 to 0.998;  $p = 0.0395$ ).

## Discussion

This systematic review and empirical study provides evidence on the present status of blinding, the potential for blinding, and the influence of blinding on outcomes in general and abdominal surgery RCT. The systematic literature survey yielded 378 RCT with a surgical intervention between 1996 and 2015, otherwise free of selection, reporting, and attrition bias. The present status of blinding in general and abdominal surgery RCT is that while blinding is used for every fourth patient and outcome assessor, only low proportions of the other types of study contributors are blinded. Only the rate of patients blinded in surgical RCT increased constantly throughout the survey period (Table 1). Consequently, there is an unexploited potential for blinding study contributors in general and abdominal surgery RCT (Table 3). As stated earlier, blinding to surgical interventions is sometimes impossible; however, in a large number of the RCT, blinding measures would have been feasible but were not applied. Blinding of the operating surgeon seems to influence the trial outcome more than blinding of any other study contributors (Table 4). One reason could be that non-blinding of the surgeon leads to unintended unblinding of other study contributors in surgical RCT because of the close cooperation in daily practice. However, unintended unblinding of study contributors was almost never assessed. Overall, stricter adherence to the CONSORT statement [11] should be demanded by journal editors.

Based on the presented evidence, the Study Center of the German Surgical Society formulated recommendations for future use of blinding in surgical trials (Box 1), reporting of blinding in surgical trials (Box 2), and assessing blinding in surgical systematic reviews (Box 3).

As an illustration of the different scenarios and potential for blinding in surgical RCT, Park et al. [19] investigated the effect of abdominal drainage on postoperative outcome. An abdominal drain is placed by the operating surgeon at the end of an intervention, so the surgeon cannot be blinded. In Park's study, every patient received a drain, the randomization was performed directly after the operation, and the drains were removed by a study assistant, leaving the operating surgeon blinded. Another example of a situation where the operating surgeon can readily be blinded is when different yet similar-looking devices are used, e.g., coated suture materials [20]. Generally, it has to be assumed that not all theoretically

blinding interventions are actually pragmatically blinding. The fact that an intervention is theoretically blinding may lead to the use of ornate blinding methods, resulting in unintentional unblinding. Therefore, the success of blinding the contributors should be assessed.

Further, the results of this study have to be compared with the existing literature. Table 5 (see Supplementary material) gives an overview of the existing evidence on blinding in the medical literature [7, 10, 16, 21–43]. These studies investigated different aspects of blinding and associated bias in various fields of medicine. Those investigating the association of lack of blinding with trial results all concluded that it leads to exaggeration of treatment effects [7, 21, 24, 25, 31, 33, 34]. The statistical analysis of the present study reproduced this association only for the operating surgeon. Therefore, further discussion is needed to explore discrepancies. The field of medicine investigated plays an important role, and differences among the results of studies may reflect differences among the fields. It is possible that surgical interventions, owing to their strong physical and visible component, are unintentionally unblinded more often than interventions in pharmacological trials. Another possible explanation is the growing standardization of endpoints in surgical trials. In pancreatic surgery, for example, the assessment of a postoperative pancreatic fistula is nowadays highly standardized, whereas up to around 10 years ago, each study defined this outcome individually [44]. Consequently, the assessment of outcomes based on predefined criteria minimizes the subjective influence of outcome assessors in the surgical field leading to a different risk of bias assessment.

This study has limitations that need to be considered when interpreting its results. First, there is a difference between the published protocol [12] and the presented results. Initially, it was planned to test the effect of performance and detection bias according to the Cochrane Handbook on significant trial results [4]. However, during the extraction and analysis stage, it became clear that combining different types of study contributors leads to loss of important information. For example, performance bias combines the bias resulting from lack of blinding of patients and surgeons. The blinding of these contributors places specific demands on surgical RCT, and by combining the resulting bias, the effect of patient blinding may confound the measured effect of surgeon blinding or vice versa. Therefore, it was decided during the analysis stage to evaluate the presence and absence of blinding of the various study contributors and present above all the planned secondary outcomes of the study protocol. Further, the significant association of lack of a blinded surgeon and significant trial results is mainly based on seven trials in which a blinding of the surgeon was possible and none of these trials had a significant primary endpoint. Therefore, this result should be interpreted with caution. Furthermore, the lack of reporting of blinding led to a high rate of unclear risk of bias according

to the Cochrane Handbook, which made the planned analysis even more futile [4].

In conclusion, some aspects of blinding are specific to surgical RCT. Generally, its potential has not yet been exhausted. If blinding of the operating surgeon is feasible but not done, this has a measurable effect on the outcome of an RCT. Reporting guidelines for RCT and systematic reviews and specific recommendations from this study should be considered during the planning, conduct, and reporting of surgical studies.

**Acknowledgements** Furthermore, the authors thank the following people for translation of full-text articles: Anja Truong (Japanese), Defan Chen (Chinese), and Solveig Tenckhoff (Italian and Spanish). Especially, the authors thank the staff of the university library for acquisition of full-text articles. Large parts of the “Background” and “Methods” section formed part of the protocol [12] and were published under the terms of the Creative Commons Attribution 4.0 International License.

**Author contributions** PP and SZ developed the study concept, acquired, analyzed, interpreted the data, and drafted the manuscript. PH, JCH, and FJH acquired literature and revised the manuscript. ALM, PK, and MKD developed the study concept, interpreted data, and revised the manuscript. All authors read and approved the final manuscript. Furthermore, all authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding information** No additional funding source is available. However, the resources and facilities of the University of Heidelberg were used in conducting this review.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no competing interests.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Abbreviations** CI, confidence interval; ITT, intention to treat; OR, odds ratio; PICO, Participants, Interventions, Comparisons, and Outcomes; RCT, randomized controlled trials

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312(7023):71–72
- Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M (2009) Oxford centre for evidence-based medicine—levels of evidence. Available from: [www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/](http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/). Last accessed 30.09.2016
- Solomon MJ, McLeod RS (1995) Should we be performing more randomized controlled trials in evaluating surgical procedures? *Surgery* 118:459–467
- Cochrane Handbook, Higgins JPT (2011) Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org). Accessed 30.09.2016
- Schulz KF, Altman DG, Moher D, CONSORT Group (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials* 11:32
- Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Mouton VM et al (2001) Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 285:2000–2003
- Jüni P, Altman DG, Egger M (2001) Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 323(7303):42–46
- Probst P, Grummich K, Harnoss JC, Hüttner FJ, Jensen K, Braun S, Kieser M, Ulrich A, Büchler MW, Diener MK (2016) Placebo-controlled trials in surgery: a systematic review and meta-analysis. *Medicine (Baltimore)* 95(17):e3516
- Boutron I, Tubach F, Giraudeau B, Ravaud P (2004) Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 57(6):543–550
- Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P (2007) Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 4(2):e1
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2010) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 8(5):336–341
- Probst P, Grummich K, Heger P, Zschke S, Knebel P, Ulrich A, Büchler MW, Diener MK (2016) Blinding in randomized controlled trials in general and abdominal surgery: protocol for a systematic review and empirical study. *Syst Rev* 5:48
- Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Büchler MW, Diener MK (2018) Optimal literature search for systematic reviews in surgery. *Langenbeck's Arch Surg* 403(1):119–129
- Begg CB, Cho MK, Eastwood S et al (1996) Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 276:637–639
- Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 291(20):2457–2465
- Probst P, Grummich K, Tenckhoff S, Ulrich A, Büchler MW, Knebel P, Diener MK (2016) Industry bias in randomized controlled trials in general and abdominal surgery: an empirical study. *Ann Surg* 264(1):87–92
- Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP (2006) Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol* 59(7):697–703
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna <http://www.R-project.org>. Accessed 30.09.2016
- Park JS, Kim JH, Kim JK, Yoon DS (2015) The role of abdominal drainage to prevent of intra-abdominal complications after laparoscopic cholecystectomy for acute cholecystitis: prospective randomized trial. *Surg Endosc* 29(2):453–457
- Diener MK, Knebel P, Kieser M, Schüler P, Schiergens TS, Atanassov V, Neudecker J, Stein E, Thielemann H, Kunz R, von Frankenberg M, Schernikau U, Bunse J, Jansen-Winkel B, Partecke LI, Prechtel G, Pochhammer J, Bouchard R, Hodina R, Beckurts KT, Leißner L, Lemmens HP, Kallinowski F, Thomusch O, Seehofer D, Simon T, Hyhlik-Dürr A, Seiler CM, Hackert T,

- Reissfelder C, Hennig R, Doerr-Harim C, Klose C, Ulrich A, Büchler MW (2014) Effectiveness of triclosan-coated PDS Plus versus uncoated PDS II sutures for prevention of surgical site infection after abdominal wall closure: the randomised controlled PROUD trial. *Lancet* 384(9938):142–152
21. Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273:408–412
  22. Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH (2002) In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol* 55:787–790
  23. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hrobjartsson A, Ravaud P (2006) Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 3:e425. <https://doi.org/10.1371/journal.pmed.0030425>
  24. Poolman RW, Struijs PA, Krips R, Siersevelt IN, Marti RK, Farrokhyar F, Bhandari M (2007) Reporting of outcomes in orthopaedic randomized trials: does blinding of outcome assessors matter? *J Bone Joint Surg Am* 89:550–558. <https://doi.org/10.2106/JBJS.F.00683>
  25. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 336:601–605. <https://doi.org/10.1136/bmj.39465.451748.AD>
  26. Karanicolas PJ, Bhandari M, Taromi B, Akl EA, Bassler D, Alonso-Coello P, Rigau D, Bryant D, Smith SE, Walter SD u G, H G (2008) Blinding of outcomes in trials of orthopaedic trauma: an opportunity to enhance the validity of clinical trials. *J Bone Joint Surg Am* 90:1026–1033. <https://doi.org/10.2106/JBJS.G.00963>
  27. Karanicolas PJ, Farrokhyar F, Bhandari M (2010) Practical tips for surgical research: blinding: who, what, when, why, how? *Can J Surg* 53:345–348
  28. Farrokhyar F, Bajammal S, Kahn moui K, Bhandari M (2010) Practical tips for surgical research. Ensuring balanced groups in surgical trials. *Can J Surg* 53:418–423
  29. Miller LE, Stewart ME (2011) The blind leading the blind: use and misuse of blinding in randomized controlled trials. *Contemp Clin Trials* 32:240–243. <https://doi.org/10.1016/j.cct.2010.11.004>
  30. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA, Cochrane Bias Methods G, Cochrane Statistical Methods G (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928. <https://doi.org/10.1136/bmj.d5928>
  31. Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S (2012) Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 344:e1119. <https://doi.org/10.1136/bmj.e1119>
  32. Boulind CE, Ewings P, Bulley SH, Reid JM, Jenkins JT, Blazeby JM, Francis NK (2013) Feasibility study of analgesia via epidural versus continuous wound infusion after laparoscopic colorectal resection. *Br J Surg* 100:395–402. <https://doi.org/10.1002/bjs.8999>
  33. Baethge C, Assall OP, Baldessarini RJ (2013) Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000–2010. *Psychother Psychosom* 82:152–160. <https://doi.org/10.1159/000346144>
  34. Berkman ND, Santaguida PL, Viswanathan M, Morton SC (2014) In: The empirical evidence of Bias in trials measuring treatment differences, Rockville (MD): Agency for healthcare research and quality (US), Report No.: 14-EHC050-EF. AHRQ Methods for effective health care.
  35. Feys F, Bekkering GE, Singh K, Devroey D (2014) Do randomized clinical trials with inadequate blinding report enhanced placebo effects for intervention groups and nocebo effects for placebo groups? *Syst Rev* 3:14. <https://doi.org/10.1186/2046-4053-3-14>
  36. Cuijpers P, Karyotaki E, Andersson G, Li J, Mergl R, Hegerl U (2015) The effects of blinding on the outcomes of psychotherapy and pharmacotherapy for adult depression: a meta-analysis. *Eur Psychiatry* 30:685–693. <https://doi.org/10.1016/j.eurpsy.2015.06.005>
  37. Kahan BC, Rehal S, Cro S (2015) Blinded outcome assessment was infrequently used and poorly reported in open trials. *PLoS One* 10:e0131926. <https://doi.org/10.1371/journal.pone.0131926>
  38. Metcalfe C, Avery K, Berrisford R, Barham P, Noble SM, Fernandez AM, Hanna G, Goldin R, Elliott J, Wheatley T, Sanders G, Hollowood A, Falk S, Titcomb D, Streets C, Donovan JL, Blazeby JM (2016) Comparing open and minimally invasive surgical procedures for oesophagectomy in the treatment of cancer: the ROMIO (Randomised Oesophagectomy: Minimally Invasive or Open) feasibility study and pilot trial. *Health Technol Assess* 20:1–68. <https://doi.org/10.3310/hta20480>
  39. Blencowe NS, Cook JA, Pinkney T, Rogers C, Reeves BC, Blazeby JM (2017) Delivering successful randomized controlled trials in surgery: Methods to optimize collaboration and study design. *Clin Trials*. <https://doi.org/10.1177/1740774516687272>
  40. Sim J (2017) Outcome-adaptive randomization: some ethical issues. *Trials* 18(Suppl 1):O42
  41. Elbers R, Savovic J, Blencowe N, Higgins JPT, Sterne JAC (2017) Performance bias in trials that cannot blind participants and healthcare providers to assigned interventions: implications for trial conduct. *Trials* 18(Suppl 1):P478
  42. Cousins S, Chalmers K, Avery K, Blencowe N, Brookes S, Blazeby J, Kobetic M, Munder T (2017) Minimising bias in surgical RCTs through blinding: a systematic review. *Trials* 18(Suppl 1):P377
  43. Blencowe N, Main BG, Blazeby JM (2017) Beyond blinding: a systematic review to explore performance bias in surgical RCTs. *Trials* 18(Suppl 1):P224
  44. Bassi C, Marchegiani G, Dervenis C, Sarr M, Abu Hilal M, Adham M, Allen P, Andersson R, Asbun HJ, Besselink MG, Conlon K, Del Chiaro M, Falconi M, Fernandez-Cruz L, Fernandez-Del Castillo C, Fingerhut A, Friess H, Gouma DJ, Hackert T, Izbicki J, Lillemoe KD, Neoptolemos JP, Olah A, Schulick R, Shrikhande SV, Takada T, Takaori K, Traverso W, Vollmer CR, Wolfgang CL, Yeo CJ, Salvia R, Buchler M, International Study Group on Pancreatic Surgery (ISGPS) (2017) The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 years after. *Surgery* 161(3):584–591