Reporting quality evaluation of propensity score method of the published articles in the gastric cancer filed: a systematic review

Jingao Zhang

Qingdao University School of Public Health

Hong Xu

the Affiliated Hospital of Qingdao University

Xiaobin Zhou (✉ xiaobin_zhou@126.com)

Qingdao University School of Public Health

Shuangyang Dai

Qingdao University School of Public Health

Beibei Li

Qingdao University School of Public Health

Jing Zhou

Qingdao University School of Public Health

Review

Keywords: Reporting quality, Evaluation, Propensity score, Gastric cancer, Propensity score matching

Posted Date: February 26th, 2020

DOI: <https://doi.org/10.21203/rs.2.24608/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background The inability to reproduce principal results of some published studies was often due to the poor reporting quality. Propensity score(PS) method has been increasingly employed to balance confounders in observational researches. A few studies showed poor reporting quality of PS method in some medical fields, this would contribute the misleading interpretation of the results and effect clinicians to determine treatment measures. The study of reporting quality of published articles applied PS method in the gastric cancer field had not been available. The aim of this study was to assess the reporting quality of PS method of the published articles in the gastric cancer field and provide critical recommendations for the investigators who would like to conduct and report PS method.

Methods The published articles applied PS method in gastric cancer field were searched in PubMed from inception to July 2019. Two reviewers independently extracted information and evaluated the reporting quality of PS method of the included articles.

Results A total of 143 eligible articles were identified by the inclusion and exclusion criteria. These articles were published from 2007 to 2019 and increased over time roughly. 112 articles(78.3%) clearly listed out the variables and 15articles(10.5%) indicate the variables selection justification for PS models. 34 articles(23.8%) reported interaction between variables or subgroup analysis. Propensity score matching(PSM) was the most used method(124 articles, 86.7%), followed by weighting(8 articles, 5.6%), stratification(4 articles, 2.8%) and regression adjustment(3 articles, 2.1%), using more than one methods was 4(2.8%). In PSM, 34 articles (26.6%) had an sufficient description about the matched algorithm and caliper width, 32 articles(25%) used standardized differences to check the balance, the reporting of replacement was poor(30 articles,21%). 10 articles(7.9%) utilized all subjects and 121articles(94.5%) did not discuss the influence of incomplete matching.

Conclusions There were methodological deficiencies in the reporting and conducting of PS method of the published articles in the gastric cancer field. The researchers should report more details for PS method so that the authors, journal editors and peer-reviewers judge reliability and authenticity of the results.

1. Background

The reproduction of primary results depends on the high quality of reporting of papers in medical researches. Problems with inadequate reporting using propensity score method in the medical researches caused the growing concerns. Description with adequate of crucial components in papers could relieve these concerns about the inability to reproduce.

Propensity score method firstly entered our eyesight in 1983 by Rosenbaum and Rubin[1]. As a widely used statistical method, the PS may control confounding factors by conditioning the probability of receiving the treatment of each participants. It performed better than multivariable regression in terms of randomized design, because the PS precluded those patients who have no similar distributions in groups[1] and it was not limited by the total sample size[2]. The PS also allows researchers to better

understand the potential impact of medical interventions and complement the findings of observational studies. Thus, the PS is a more practical tool for researchers to accurately assess the treatment effect in researches where many biases could exist. However, despite a growing number of publications using PS method, the quality of the reporting of the PS did not increase as desired[3, 4]. Prior literatures showed that most researchers did not report enough details regarding the balance of covariates between the treatment groups and the choice of covariates included in PS models[5–7].

This poor quality of report may directly cause clinicians to choose suboptimal treatment measures that would delay the recovery of the patient[3], and an inadequate report often makes it difficult for other researchers to confidently judge the appropriateness of reported analyses to conduct and reproduce published results[4]. The authors should describe adequate essential details to allow readers and other researchers to validate the findings, which has been considered as an imperative role for high-quality researches.

Incidence of gastric cancer ranks fifth among common cancers, with a higher rates in east Asian countries[8, 9]. We found the PS has been used in many literatures in the gastric cancer. Despite the widespread use of the PS in gastric cancer field, the reporting quality of the PS has not been evaluated and lacks the proper report guidelines for the PS.

The aim of this study was to assess the reporting quality of PS method of the published articles in the gastric cancer field and provide critical recommendations for the investigators who would like to conduct and report PS method.

2. Method

2.1 Search strategy

A well-designed search strategy was conducted in PubMed to identify the published articles using PS in the field of gastric cancer. The search strategy for PubMed was outlined in Figure1. The search was performed from inception of journal to July 8th,2019, and the language was limited to English. The specific search procedure for PubMed was showed in Figure2.

2.2 Criteria for literature inclusion and exclusion

Criteria for literature inclusion: 1)the title, abstract or key words described gastric cancer; 2) the title, abstract or key words described PS method; 3)published in English; 4) the object of study was human.

Criteria for literature exclusion: 1)non-observational studies, such as systematic reviews, meta-analysis, randomized controlled trials (RCTs), quasi-randomized trials, other interventional studies, case series analysis, case reports, meetings, guidelines; 2)studies not related gastric cancer; 3)full text were not available; 4)the object of study was animals.

2.3 Data extraction

The reviewers received training from professionals before extracting data, and then the data from included articles was independently extracted by two of the authors. When some discrepancies existed, we resolved them by making discussions or consulting with a third author. The items based on previously literatures[2, 3, 10] were critically adopted to extract information. The general characteristics of included articles contained year of publication, name of journal, origin region of first author, author's affiliations, participation of statistician or epidemiologist in author (identified from author's affiliations or the acknowledgements part), international cooperation, journal source of Science Citation Index(SCI), impact factors(IF), number of citations, number of pages, number of authors, funding of support, the way to determine the sample size and number of patients engaged.

2.4 Evaluation of report quality of PS methods of included papers

These items of reporting of PS method were recorded.

2.4.1 For the assessment of variables in PS method, we extracted the information about the variables, for example, justification of the variables chosen; the number of variables in the PS model; the inclusion of interaction or polynomial terms of variables in the PS model were also extracted.

2.4.2 For the aspect of how to construct the PS, the reporting of the type of regression model used to estimate the PS was recorded.

2.4.3 The type of PS methods was also extracted. In fact, there are 4 main methods of PS analyses: PS matching, PS stratification, PS adjustment and PS weighting.

2.4.4 The comparability of baseline characteristics in PS analyses was extracted.

2.4.5 In the aspect of propensity score matching(PSM), matching ratio (1/1, 1/n, n/1, etc.), the matching algorithm and distance metric, balance check, replacement or not, the proportion of matched sample size, whether discussed the influence of incomplete matching were recorded.

2.4.6 In the aspect of weighting and stratification, the type of weighting and the number and definition of stratification were abstracted.

2.4.7 The information about whether reported the way to address potential sources of bias was abstracted. Some authors recommend performing sensitivity analyses and subgroups analyses to determine how susceptible the data are to bias unmeasured by the investigators.

If these information could be find in the section of method or result, then we would believe this item was reported. We applied the extracted items to assess the sufficiency of reporting of the PS.

2.5. Data analysis

Categorized variables of characteristics and reporting were described with frequencies and percentages. Continuous variable were described with rang, mean, interquartile range(Q1 to Q3) and median. The

degree of agreement was examined using Kappa coefficient. All the statistics analyses were conducted by SPSS 19.0.

3. Result

3.1. Literature search

325 articles using PS methods published in gastric cancer field were identified. After screening the titles and abstracts, 182 papers were excluded. Ultimately, this procedure yielded 143 eligible articles published from 2007 to 2019. The degree of agreement between data extractors was acceptable (Kappa coefficient = 0.86, $P < 0.01$).

3.2. The General characteristics of selected Articles

The primary features of these articles were outlined in table 1.

In the past five years, the number of articles published has grown rapidly, especially in 2018, with 41 articles (see Figure 3). However, because the deadline for retrieval was the middle of 2019, the number of articles could be underestimated in this year. 38 articles (26.6%) were from Japan, 37 articles (25.9%) from China, 33 articles (23.1%) from South Korea, 9 articles (6.3%) from USA. The articles for participation of statistician or epidemiologist in author were 32 (22.4%). What's more, 140 articles (98.0%) didn't explain how to determine the sample size.

3.3. Reporting quality of PS methods

The characters of PS method of included articles were showed in table 2.

3.3.1. 128 articles (89.5%) did not indicate the covariates selection justification for PS models. All articles used demographic and clinical variables to perform PS analyses. 112 articles (78.3%) clearly listed out the covariates in PS models, the number of assessed covariates ranged from 3 to 37, and the median was 7. 34 articles (23.8%) reported interaction between variables or subgroup analysis.

3.3.2. 103 articles (72.0%) reported estimation of PS models and all of them adopted a logistic regression to construct the PS model. Probit regression, discriminate analysis, regression tree and other methods based on data mining algorithm didn't be used in the included articles.

3.3.3. PSM was the most used method (124 articles, 86.7%), followed by PS weighting (8 articles, 5.6%), PS stratification (4 articles, 2.8%) and PS adjustment (3 articles, 2.1%), using more than one methods was 4 (2.8%), the PS matching and PS weighting accounted for 92.3% (132 articles).

3.3.4. 41 articles (28.7%) didn't report the comparability of measured baseline covariates.

3.3.5. For PSM, 116 articles (90.6%) reported the type of matching ratio. 1:1 matching accounted for 81.3% ($n=104$), following 1:n was 7.0% ($n=9$), other matching ratios, such as $n:1$ and $1:1:n$, was 2.3% ($n=3$). 58

articles(45.3%) applied greedy matching. 40 articles(31.2%) used caliper, the width of caliper ranged from 0.01 to 0.30 standard deviations(SDs). 34 articles(26.6%) had an sufficient description about the matched algorithm and caliper width. 32 articles(25%) used standardized differences to check the balance in PSM. 30 articles(21%) applied matching without replacement. None article used matching with replacement.10 articles(7.9%) utilized all subjects, 58 articles(45.3%) used more than 50% of sample size. 121articles(94.5%) did not discuss the influence of incomplete matching. The detail of PSM report was shown in table3.

3.3.6. For PS weighting, 4 articles(50%) used inverse probability of treatment weighting, one article [11] used the Toolkit for Weighting and Analysis of Nonequivalent Groups method to conduct the research. And for the PS stratification, all of articles about PS stratification reported their own strata and identification.

3.3.7. The way to address potential sources of bias were sensitivity analyses and subgroups analyses, which were reported in 15 articles(10.5%) and 30articles(21%), respectively.

4. Discussion

Our study demonstrated that the quality of these included papers was unsatisfactory, the result was line with prior systematic reviews[7, 10, 12, 13], despite some guidelines about the PS method had found in recent years. Many articles ignored the essential details and adopted the inappropriate methods, causing the misleading interpretation of the treatments in published article. Thus, we would mainly discuss the following aspects.

For the variables in PS models, all included articles used demographic and clinical variables to conduct the PS, which is a good practice. 112 articles(78.3%) clearly listed out the variables, and not all variables were incorporated into the ultimate PS models, this omission of variables implicitly means that potential biases from these variables are considered negligible in theory, meanwhile we should pay more attention to the number of variables, these models, such as PS models with few variable and PS models with many variables but smaller sample size, usually do not produce unbiased causal reference[14]. On the other hand, only 15 articles(10.5%) reported how variables were selected to construct the PS, and the non-parsimonious that include all variables is the most typically ways to choose variables, this method could be deleterious if the contain variables associated without the prognosis[15]. When the justification of the variables included in the PS method is not clearly stated, it is possible that important variables are not included in the analysis and limits the reproducibility of the results. In addition, some studies[16, 17] showed that adequate professional knowledge and practical clinical experience were the key factors for determining included variables, interaction, and/or higher order term. The variables eventually incorporated into PS models should be associated with the outcome[4, 18], but some quality reports of PS methods ignored this part[3, 13].

Another concern was the choice of the estimation of PS model, like T. L. Zakrisson and colleagues reported that most researchers would like to adopt a logistic regression[13], however, in other fields, boosted

regression trees and neural networks were proposed to construct a PS model, these ways based on data mining algorithm are rapidly spread by the development of computer statistical software in recent years[19], one of the advantages is that these methods could automatically find any nonlinear terms and take them into the estimation model of the PS. Although the logistic regression model could meet many circumstances, it would introduce us to select the 'optimal' model that contains the expected results displaying in front of you[4, 20]. Therefore, we encourage researchers to employ statistical methods that based on data mining algorithm, because these methods could yield more precise estimates of treatment effect in conditions of both non-additivity and non-linearity[21–23].

For propensity score methods, PS matching was the most popular method, similar with other reporting[3, 13], a focused analysis on PSM methodology was undertaken. Only 34 articles(26.6%) had a sufficient description about the matched algorithm and caliper width. The caliper width of 0.2 of the SDs of the logit of the propensity score was considered as an ideal choice, and the recommendation was used in most included articles(50%) in our study, because this value could eliminate bias as much as possible[7, 24]. When matches were hard to find, a looser caliper might be acceptable to avoid loss of sample size[18]. For reports with matching ratios, our reporting rate was lower than other studies[3], but there were some matching ratio such as 1:n and n:1 in our study, these matching ratios could maximize the use of the included participants to ensure the precision of the estimation and the generalizability of results[6]. However we found that little report guideline about PS method suggested to discussed the effects about incomplete matching, especially these subjects who were excluded after matching in the treatment group, this inadequate guideline could cause the waste of valuable information and the less reliable results. Other PS methods were also used, like Monte Carlo simulations study indicated that matching and weighting eliminates better systemic differences between treated groups than stratification and covariate adjustment[20, 25]. In our study, the number of researches using matching and weighting was much more than stratification and covariate adjustment. Meanwhile, we found the Toolkit for Weighting and Analysis of Nonequivalent Groups method that has not used in other quality reporting of the PS, different from other methods, because this algorithm(not user) determines the most appropriate model for the propensity score[26].

When checking for covariate balance, 28.7% articles didn't report the comparability of measured baseline covariates in our study. Xiaoxin Yao and colleagues found 21.9% cancer studies and 15.6% cancer surgical studies didn't report them[3], and the literature[15] showed that 20% of 97 surgical studies analyzed included an assessment of covariate balance using standardized differences. In present study, the reporting of checking balance had not improved, 41 articles(28.7%) didn't report the comparability of measured baseline covariates, and most articles(120, 82.0%) used test of significance instead of standardized differences to check the covariates balance. We don't recommend test of significance, the reason is that the method is susceptible to the sample size[6, 20] and might ignore the imbalance because of lower statistical power. It's necessary to encourage authors to report the comparability of measure baseline covariates with appropriate methods. We suggest using standardized differences to check balance of baseline data measured between treated groups in PSM, because it is not confused with other factors [4]. A cutoff value of a standard deviation of less than 0.1 that indicates negligible was

approved by most experts.[13]. Therefore the standardized differences should be encouraged[27]. If the results did not achieve the intended purpose, we could repeat the process by adding more variables or interactions based on the existing model to balance covariates.

Furthermore, a caveat of PS method is that it could only account for measured confounders, those potential biases caused by unmeasured confounding variables could influence authors to obtain an accurate estimation of treatment effect. The optimal ways to avoid the effects of unmeasured covariates is to implement sensitivity analyses or subgroup analyses. In our study, 5 articles(31.5%) reported using these methods, which is consistent with other reports on PS methods[28].

This article was the first evaluation about the reporting quality of PS method of the published articles in this filed, and we hope that our study would be valued by researchers who want to apply PS methods to medical research in the gastric cancer. Meanwhile, there are limitations in our study. Firstly, we didn't discuss the details of covariate adjustment and weighting, because our study contained only a small part of articles, readers could refer elsewhere for these methods[29, 30]. Secondly, some advanced software systems have made it easier for researchers to use the PS method, which might lead them to have no explicate description in their articles. Thirdly, our study only reflected the field of gastric cancer, articles from other fields were excluded.

5. Conclusion

Although many studies have used propensity score methods in the gastric cancer literatures, there were some flaws in the reporting and use of PS method and the quality of the report was suboptimal. It is time to take measurements to improve the reporting quality of PS method of the published articles, thus we propose authors to adopt rational reporting guidelines about PS methods to promote transparency and consistency.

Abbreviations

PS

Propensity Score, PSM:Propensity Score Matching, RCTs:Randomized Controlled Trials, SCI:Science Citation Index, IF:impact factors, SDs:Standard Deviations, Q1:1st Quartile Q3:3nd Quartile.

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and materials:

The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Competing interests:

The authors declare that they have no competing interests.

Funding:

Supported by the Improvement Program for the Education of Graduate Students in Shandong Province, China (grant number: SDYAL18047); the National Steering Committee for Education of Medical Degree Postgraduate (grant number: B2-YX20180203-01); 2018 Qingdao University Graduate Case Database Construction Project.

The funding body support in literature collection, paper publishing, technical training, and expert guidance.

Authors' contributions:

Study concepts: Z.J.G., Z.X.B.; Study design: Z.J.G., Z.X.B.; Manuscript writing Z.J.G.,Z.X.B.; Manuscript editing: Z.J.G., Z.X.B., L.B.B.; Data extraction: Z.J.G., Z.J.; Data elaboration and interpretation: Z.J.G., Z.X.B., X.H., D.S.Y, L.B.B.; Statistical analysis: Z.J.G., Z.X.B.; Manuscript revision and approval of submission in its present form: Z.J.G., Z.X.B., D.S.Y, L.B.B. All authors read and approved the final manuscript.

Acknowledgements:

The authors thank Guo-Yi Yu, director of the editorial board of Journal of Qingdao University Medical College and Jian-Qiang Li, director of the editorial board of Journal of Precision Medicine for providing helpful comments on early draft of this article. The authors also thank the reviewers for their insightful comments and suggestions.

References

1.
Rosenbaum PR. R.D., The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55.
2.
Grose E, et al., Use of Propensity Score Methodology in Contemporary High-Impact Surgical Literature. *J Am Coll Surg*, 2019.
- 3.

- Yao XI, et al., Reporting and Guidelines in Propensity Score Analysis: A Systematic Review of Cancer and Cancer Surgical Studies. *J Natl Cancer Inst*. 2017. 109(8).
- 4.
- Deb S, et al. A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. *Can J Cardiol*. 2016;32(2):259–65.
- 5.
- Weitzen S, et al. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf*. 2004;13(12):841–53.
- 6.
- Ali MS, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *J Clin Epidemiol*. 2015;68(2):112–21.
- 7.
- Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *J Thorac Cardiovasc Surg*. 2007;134(5):1128–35.
- 8.
- Warschkow R, et al., Selective survival advantage associated with primary tumor resection for metastatic gastric cancer in a Western population. 2018. 21(2): p. 324–33.
- 9.
- Jemal A, et al. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69–90.
- 10.
- Gayat E, et al. Propensity scores in intensive care and anaesthesiology literature: a systematic review. *Intensive Care Med*. 2010;36(12):1993–2003.
- 11.
- Lee IS, et al. Prognostic impact of extranodal extension in stage 1B gastric carcinomas. *Surg Oncol*. 2018;27(2):299–305.
- 12.
- D'Ascenzo F, et al. Use and misuse of multivariable approaches in interventional cardiology studies on drug-eluting stents: a systematic review. *J Interv Cardiol*. 2012;25(6):611–21.
- 13.
- Zakrisson TL, Austin PC, McCredie VA. A systematic review of propensity score methods in the acute care surgery literature: avoiding the pitfalls and proposing a set of reporting guidelines. *Eur J Trauma Emerg Surg*. 2018;44(3):385–95.
- 14.
- Thoemmes FJ, Kim ES. A Systematic Review of Propensity Score Methods in the Social Sciences. *Multivariate Behav Res*. 2011;46(1):90–118.
- 15.
- Lonjon G, et al. Potential Pitfalls of Reporting and Bias in Observational Studies With Propensity Score Analysis Assessing a Surgical Procedure: A Methodological Systematic Review. *Ann Surg*. 2017;265(5):901–9.
- 16.

- Myers JA, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*. 2011;174(11):1213–22.
- 17.
- Myers JA, Gagne RJ, Huybrechts JJ, Schneeweiss KF, Rothman S. KJ, et al., Myers et al. Respond to “understanding bias amplification” *Am J Epidemiol* 174(1228–9).
- 18.
- Austin PC, Tutorial A. and Case Study in Propensity Score Analysis: An Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality. *Multivariate Behav Res*. 2011;46(1):119–51.
- 19.
- Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J Clin Epidemiol*. 2010;63(8):826–33.
- 20.
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399–424.
- 21.
- McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychol Methods*. 2004;9(4):403–25.
- 22.
- Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PLoS One*. 2011;6(3):e18174.
- 23.
- Setoguchi S, et al. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiol Drug Saf*. 2008;17(6):546–55.
- 24.
- Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat*. 2011;10(2):150–61.
- 25.
- Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Making*. 2009;29(6):661–77.
- 26.
- BA
Griffin
GR,AR
Morral
2014
Griffin BA, Morral GR,AR, et al., Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG), 2014. <http://www.rand.org/statistics/twang>. 2014.
- 27.

Rosenbaum PR. R.D., Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. American Statistician, 1985(39): p. 33–38.

28.

Luo Z, Gardiner JC, Bradley CJ. Applying propensity score methods in medical research: pitfalls and prospects. Med Care Res Rev. 2010;67(5):528–54.

29.

Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015;34(28):3661–79.

30.

Elze MC, et al. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. J Am Coll Cardiol. 2017;69(3):345–57.

Tables

Table1. The General Characteristics of Included Articles(n=143)

	No. of articles(%)
Journal	
Gastric Cancer.	19(13.4)
Surge Endosc.	13(9.2)
Ann Oncol.	11(7.7)
Medicine (Baltimore).	10(7.0)
Eur J Surg Oncol.	8(5.6)
World J Gastroenterol.	7(4.9)
Gastrointest Endosc.	7(4.9)
Other	68(47.6)
Origin region of first author	
Japan	38(26.6)
China	37(25.9)
South Korea	33(23.1)
USA	9(6.3)
Other	26(18.2)
Affiliation of author	
university	13(9.1)
Hospital	130(90.9)
Participation of statistician or epidemiologist in author	
No	111(77.6)
Yes	32(22.4)
International collaborative authorship	
No	128(89.5)
Yes	15(10.5)
Journal source of SCI	
No	3(2.1)
Yes	140(97.9)
IF of SCI	
0-4	99(70.7)
>4 ^a	41(29.3)
Citations(No. of citations)	
1-9	100(70.0)
>9 ^b	33(23.0)
missing	10(7.0)
Manuscript length (No. of pages)	
1-9	91(63.6)
>9 ^c	52(36.4)
Number of authors	
1-8 ^d	72(50.3)
>8	71(49.7)
Funding of support	
No	73(51)
yes	70(49)
Whether reported how to determine the sample size	
No	140(98)
Yes	3(2)
No. of patients enrolled	
Median (Q1 to Q3)	259.5(147 to 725.75)
minimum to maximum	44-63397

a: The median of IF of SCI is 4 that is regarded as division value.

b: The median of citations is 9.

c: The median of manuscript length is 9.

d: The median of number of authors is 8.

IF: Impact Factor, SCI: Science Citation Index, No.: Number, Q1: 1st Quartile, Q3: 3rd Quartile.

Table2. Reporting of Propensity Score Analysis in Included Studies

	No. of articles(%)
Whether reported the justification of the covariates chosen	
No	128(89.5)
Yes	15(10.5)
Reported the covariates included in PS analysis	
No	31(21.7)
Yes	112(78.3)
No. of covariates included in PS analysis	
Mean	8.15
Median (Q1 to Q3)	7(6 to 10)
minimum to maximum	3-37
Whether reported the inclusion of interaction or polynomial terms of covariates in the PS model	
No	109(76.2)
Yes	34(23.8)
The estimation of PS model	
Logistic regression	103(72)
Not reported	40(28)
Type of PS methods	
PS matching	124(86.7)
PS weighting	8(5.6)
PS stratification	4(2.8)
PS adjustment	3(2.1)
More than one methods use	4(2.8)
Comparability of baseline characters	
No	41(28.7)
Yes	102(71.3)
Type of PS weighting(n=8)	
Inverse probability of treatment weighting	4(50)
The Toolkit for Weighting and Analysis of Nonequivalent Groups method	1(12.5)
Other	3(37.5)
Whether reported the number of strata and definition in PS stratification(n =6)	
No	0
Yes	6(100)
Whether reported the way to address potential sources of bias	
Report a sensitivity analyses	
No	128(89.5)
Yes	15(10.5)
Report a subgroups analysis	
No	113(79.0)
Yes	30(21.0)

Table 3 Raw Frequencies and Percentages of Categorical Variables for the 128 Studies That Used Matching

Variable	No. of articles(%)
Ratio of matching	
1:1	104(81.3)
1:n	9(7.0)
Other	3(2.3)
Not reported	12(9.4)
Type of matching algorithm	
Greedy matching	58(45.3)
Greedy nearest neighbor	12(9.3)
Greedy within specified caliper distance	40(31.3)
Greedy matching by digit	6(4.7)
Optimal matching	3(2.3)
Other (ie, kernel, exact)	5(3.9)
Not reported	62(48.5)
Caliper width(X of the SD of the logit of the PS)	
0.01	2(5)
0.02	3(7.5)
0.05	1(2.5)
0.10	8(20)
0.15	1(2.5)
0.20	20(50)
0.25	3(7.5)
0.30	2(5.0)
Type of balance check	
Test of significance	80(56)
Standard difference	32(25)
Not reported	31(21.6)
The proportion of matched sample size	
<25%	17(13.3)
25%-	49(38.3)
50%-	41(32)
75%-	17(13.3)
Not reported	4(3.1)
Use of replacement	
No	30(21)
Yes	0(0)
Not reported	113(79)
Whether discussed the influence of incomplete matching	
No	136(95.1)
Yes	7(4.9)

SD: Standard Deviation

Figures

#1 Stomach Neoplasms⁺
 #2 Neoplasm, Stomach⁺
 #3 Stomach Neoplasm⁺
 #4 Neoplasms, Stomach⁺
 #5 Gastric Neoplasms⁺
 #6 Gastric Neoplasm⁺
 #7 Neoplasm, Gastric⁺
 #8 Neoplasms, Gastric⁺
 #9 Cancer of Stomach ⁺
 #10 Stomach Cancers ⁺
 #11 Gastric Cancer ⁺
 #12 Cancer, Gastric⁺
 #13 Cancers, Gastric ⁺
 #14 Gastric Cancers ⁺
 #15 Stomach Cancer ⁺
 #16 Cancer of the Stomach⁺
 #17#1OR#2OR#3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12
 OR #13 OR #14 OR #15 OR #16⁺
 #18 Propensity Score⁺
 #19 PS ⁺
 #20#18OR#19⁺
 #21#17AND#20⁺

Figure 1

PubMed search strategy.

Figure 2

Flow Diagram of Articles Identified.

Figure 3

Publication Trends in Gastric Cancer Reporting Use of Propensity Score Analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PRISMA2009.doc](#)