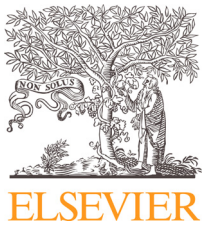
Journal of Plastic, Reconstructive & Aesthetic Surgery 75 (2022) 4096–4105



JPRAS
An International Journal of Surgical Reconstruction
www.JPRASurg.com

Review

The risk of bias of non-randomized observational studies in deep inferior epigastric perforator flap breast reconstruction: A systematic review using ROBINS-I

Morgan Yuan[a], Jeremy Wu[b], James Lee[c], Daniel Cao[a], Minh NQ Huynh[d], Lucas Gallo[d], Anne C O' Neill[e], Stefan O.P. Hofer[e,*]

[a] *Michael G. DeGroote School of Medicine, McMaster University, Hamilton ON, Canada*
[b] *Faculty of Medicine, University of Toronto, Toronto ON, Canada*
[c] *Faculty of Health Sciences, McMaster University, Hamilton ON, Canada*
[d] *Division of Plastic Surgery, Dept. of Surgery, McMaster University, Hamilton, ON, Canada*
[e] *Division of Plastic, Reconstructive & Aesthetic Surgery, Department of Surgery, University of Toronto, Toronto ON, Canada*

Received 24 January 2022; accepted 21 June 2022

KEYWORDS
ROBINS-I;
Risk of bias;
Breast reconstruction;
Deep inferior epigastric perforator;
DIEP;
Observational studies

Summary *Introduction:* The deep inferior epigastric perforator (DIEP) flap is regarded as the gold standard for autologous breast reconstruction. However, due to difficulty designing and conducting randomized controlled trials in surgical interventions, the majority of literature on DIEP flap breast reconstructions are observational studies. As such, it is pivotal that these studies are performed with high internal validity.
Methods: A literature search was performed using MEDLINE, Embase, and CENTRAL from January 1, 2015 to October 23, 2021. Studies identified as observational studies about DIEP breast reconstruction and published in a journal with a Web of Science impact factor above 1.0 were included. Screening and risk of bias (RoB) assessment using the ROBINS-I tool were performed independently and in duplicate by two authors.
Results: From 12,371 studies, 66 observational studies were included. The majority were found at RoB, with 11 at moderate, 26 at serious, and 6 at critical RoB. Only two studies had low RoB.

* Corresponding author at: 200 Elizabeth Street, 8N-865, Toronto, ON M5G 2C4, Canada.
E-mail addresses: jeremytl.wu@mail.utoronto.ca (J. Wu), stefan.hofer@uhn.ca (S.O.P. Hofer).

https://doi.org/10.1016/j.bjps.2022.06.093

The bias most commonly arose due to Domain 1 (confounding variables), Domain 3 (classification of interventions), and Domain 6 (measurement of outcomes).
Conclusions: In this review, we demonstrate the high RoB of observational studies evaluating DIEP breast reconstruction, which may jeopardize the validity of findings. We recommend that authors consult the ROBINS-I tool not only when assessing observational studies for systematic reviews but also when designing or conducting these studies. In our study, we present additional considerations for each domain to provide researchers with guidance on assessing and conducting surgical observational studies.

Contents

Introduction

Breast cancer patients undergoing mastectomies often suffer from persistent psychological disturbances, such as loss of femininity, disruption of body image, and declining self-worth.[1,2] These patients are often offered the option of breast reconstruction, which has been documented to provide significant psychological, emotional, and functional benefits for postmastectomy patients, including improved self-esteem and sexuality and reduced risk of cancer recurrence.[2]

Among the different modalities available for reconstruction, the deep inferior epigastric perforator (DIEP) flap is often regarded as the gold standard for autologous breast reconstruction. A recent study comparing four abdominal-based autologous breast reconstruction methods showed that the DIEP flap was associated with the highest abdominal well-being and lowest donor site morbidity.[3] However, the incidence of complications following this procedure has been reported to be 20–40%, with the majority considered minor complications, such as fat necrosis and infections.[4,5] The incidence of flap loss is estimated to be approximately 2%, with the risk of a reoperation, due to hematomas or wound-healing disturbances, being relatively higher than that of alloplastic reconstructions due to the invasive nature of DIEP flap reconstructions.[4-6] This evidence suggests a considerable risk for patients deciding on reconstructive options following mastectomy. As such, research regarding DIEP flap breast reconstructions must be designed and conducted with high internal validity to ensure these patients can achieve the desired esthetic and psychological benefits while minimizing risks.

Although randomized controlled trials (RCTs) are considered the gold standard for comparing the effects of novel surgical interventions, these studies are difficult to design and conduct within the field of plastic surgery. An article by Hassanein et al. found only 1.83% of studies in three major plastic surgery journals to be RCTs.[7] They postulated that this is due to ethical concerns regarding sham surgeries which provide no therapeutic benefit, inability to provide classic randomization given patient and surgeon preference for a particular surgical option, and differences in a surgeon's proficiency for different surgeries.[7] Particularly, it is well known that breast reconstruction RCTs have difficulties with the recruitment of participants.[8] These reasons lead non-randomized studies to dominate the plastic surgery literature, including the evidence which informs breast reconstruction procedures.

To ensure non-randomized studies are performed with high internal validity, the Risk Of Bias In Non-randomized

Table 1 ROBINS-I criteria.

Criteria	Domain	Description
1	Pre-intervention	Bias due to confounding
2	Pre-intervention	Bias in the selection of participants into the study
3	At intervention	Bias in the measurement of interventions
4	Post-intervention	Bias due to departures from intended interventions
5	Post-intervention	Bias due to missing data
6	Post-intervention	Bias in the measurement of outcomes
7	Post-intervention	Bias in the selection of the reported results

Studies - of Interventions (ROBINS-I) tool has been developed.[9] This tool consists of seven domains, which address bias due to confounding, selection of participants, classification of interventions, deviations from the originally stated intervention, missing data, measurement of outcomes, and selection of the reported result (Table 1).[9] This tool was developed through consensus from both methodological experts and systematic review authors and editors, with substantial revisions through feedback from users.[9] To the best of our knowledge, no previous studies have assessed the internal validity of non-randomized breast reconstruction studies. As such, the primary objective of this study was to evaluate the risk of bias (RoB) in non-randomized studies concerning DIEP flap breast reconstruction using the ROBINS-I tool.

Methods

This systematic review was conducted in adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) reporting guideline. This study was prospectively registered on Open Science Framework (https://osf.io/2d4jh/).

Search strategy

A comprehensive literature search of MEDLINE, Embase, and CENTRAL was performed to identify all non-randomized observational studies published from January 1, 2015 to October 23, 2021 using key terms that pertain to DIEP flap breast reconstructions. Our search strategies for each database can be found in Appendix A.

Studies identified by the search strategy were uploaded to Covidence software for systematic reviews (Veritas Health Innovation Ltd). Two authors (J.L. and D.C.) independently screened titles and abstracts in duplicate to assess for eligibility. Any studies where there was insufficient information in the titles and abstracts to determine eligibility were reviewed by full-text screening. The same two authors assessed studies for final inclusion at the full-text level. All discrepancies were resolved through consensus between the two reviewers and the corresponding senior author (S.H.). A pilot screening assessment was performed on 10 studies prior to the inclusion date (prior to January 1, 2015) to ensure agreement between reviewers.

Eligibility criteria

Articles were included if they:

1. were identified as non-randomized observational studies that had clearly defined groups comparing interventions;
2. have a focus on DIEP flap breast reconstruction;
3. assessed clinical and/or patient-reported outcomes;
4. were published in journals with a 2020 Web of Science impact factor above 1.0 between January 1, 2015 to October 23, 2021.

Studies that were non-English, non-human, randomized controlled trials, case series, case studies, systematic reviews, and other study designs (narrative reviews, expert opinions, editorials, protocols, and conference abstracts) were excluded. Studies were also excluded if they were identified as observational studies that compared multiple variables and did not have clearly defined groups.

Data collection and analysis

Independent data extraction was conducted in duplicate by two authors (M.Y. and J.W.). Discrepancies that arose were resolved through discussion and consensus. The included studies were assessed for their RoB using the ROBINS-I tool. Study characteristics were also extracted including journal, year of publication, study type, and sample size.

Quality assessment

The ROBINS-I tool was used to assess the RoB of included non-randomized observational studies. This tool evaluates seven domains for RoB: two domains at the level of pre-intervention, one domain at the level of the intervention delivery, and four domains at the level of post-intervention (Table 1). Each domain was assigned an RoB judgment among the following options: low, moderate, serious, critical, or no information. A final judgment with the same options was then made for the entire study based on the findings from each domain.

Statistical analysis

The general characteristics of included studies will be summarized using descriptive statistics. The ROBINS-I scores will

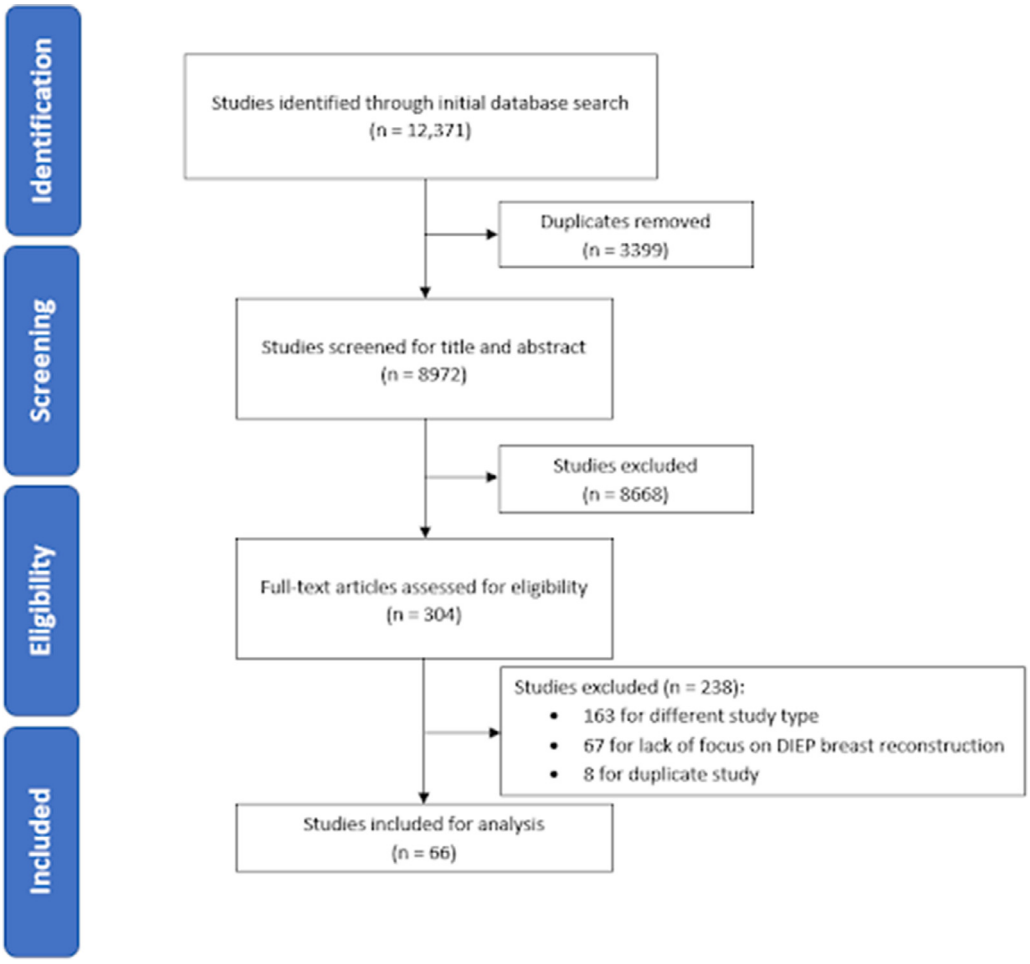


Fig. 1 PRISMA diagram.

be presented based on those at risk for each domain and the overall RoB of studies.

Results

Search results and study selection

The initial search identified 12,371 articles, with 3399 duplicates removed. The remaining 8972 studies were screened for eligibility. Following title and abstract screening, 304 studies moved on to full-text review. A further 225 studies were excluded at the full-text level, with 66 studies meeting our eligibility criteria and being included for analysis (Fig. 1). A list of our included studies can be found in Appendix B.

Study characteristics

The characteristics of included studies are outlined in Table 2. Of the 66 articles included in this review, 51 were retrospective cohort studies (77%), 12 were prospective cohort studies (18%), and 3 were cross-sectional studies (4.5%). The observational studies were published across 17 journals, with the most studies being published in *Journal of Plastic, Reconstructive and Aesthetic Surgery* ($n = 17$, 26%), *Plastic and Reconstructive Surgery* ($n = 14$, 21%), and the *Journal of Reconstructive Microsurgery* ($n = 7$, 11%). A detailed summary of characteristics of individual studies is available in Appendix C.

ROBINS-I assessment

The overall RoB and RoB of each domain, determined by the ROBINS-I, are presented in Table 3. A graphical summary is presented in Fig. 2. Among the included studies, 2 studies had a low RoB, 11 had a moderate RoB, 26 had a serious RoB, 6 had a critical RoB, and 21 were judged to be no information. Most commonly, bias arose due to Domain 1 (confounding variables), Domain 3 (classification of interventions), and Domain 6 (measurement of outcomes). A detailed breakdown of RoB in individual studies is available in Appendix D, and examples of serious or critical RoB in each domain from our included studies with recommendations to authors can be found in Table 4.

Table 2 Characteristics of included studies.

Study characteristic	No. (%)
Number of studies included	66
Year	
2015	10 (15.2%)
2016	14 (21.2%)
2017	6 (9.1%)
2018	9 (13.6%)
2019	4 (6.1%)
2020	11 (16.7%)
2021	12 (18.2%)
Journal	
Journal of Plastic, Reconstructive and Aesthetic Surgery	17 (25.6%)
Plastic and Reconstructive Surgery	14 (21.2%)
Journal of Reconstructive Microsurgery	7 (10.6%)
Microsurgery	6 (9.1%)
Annals of Plastic Surgery	6 (9.1%)
Other	16 (24.2%)
Study type	
Prospective cohort	12 (18.2%)
Retrospective cohort	51 (77.3%)
Cross-sectional	3 (4.5%)
Sample size	Mean: 531.3, range: 20–3926

Table 3 Risk of bias of each domain and overall risk of bias in included studies.

Domains	Number of studies for each risk of bias classification (%)				
	Low	Moderate	Serious	Critical	No information
D1	26 (39.4)	20 (30.3)	16 (24.2)	4 (6.1)	0 (0)
D2	34 (51.5)	5 (7.6)	5 (7.6)	1 (1.5)	21 (31.8)
D3	38 (57.6)	18 (27.3)	1 (1.5)	0 (0)	9 (13.6)
D4	52 (78.8)	7 (10.6)	2 (3.0)	1 (1.5)	4 (6.1)
D5	40 (60.6)	7 (10.6)	6 (9.1)	0 (0)	13 (19.7)
D6	49 (74.2)	12 (18.2)	4 (6.1)	1 (1.5)	0 (0)
D7	53 (80.3)	11 (16.7)	2 (3.0)	0 (0)	0 (0)
Overall	2 (3.0)	11 (16.7)	26 (39.4)	6 (9.1)	21 (31.8)

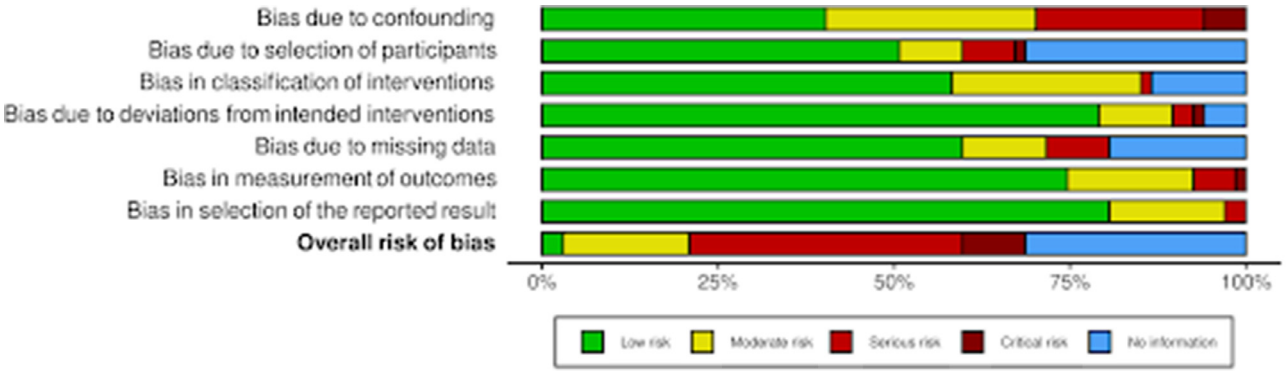


Fig. 2 Risk of bias of each domain and overall risk of bias in included studies.

Discussion

A well-designed and conducted observational study should be comparable to RCTs in the validity of conclusions drawn, apart from randomization. The ROBINS-I tool was designed to identify potential biases in observational studies. The goal of our study was to evaluate the RoB in observational studies evaluating DIEP flap breast reconstructions using the ROBINS-I tool. Our study found that the distribution of RoB judgments across all domains is in keeping

Table 4 Examples of risk of bias (RoB) of each domain from included studies with recommendations to mitigate RoB.

ROBINS-I domain	Example	Rationale	Recommendation
D1 - Bias due to confounding	Beugels et al. 20,18[6]: compared postoperative quality of life via the BREAST-Q questionnaire between immediate and delayed DIEP flap breast reconstructions *RoB Rating: Critical*	Between the immediate and delayed DIEP reconstruction groups, there were significant differences in baseline characteristics, including concurrent chemotherapy and radiation therapy that are known to affect breast reconstruction outcomes. As well, the rate of recipient-site and donor-site complications differed between groups. These discrepancies were not adjusted and controlled for through statistical means when the quality of life was compared between groups and can factor into a patient's perception of their postoperative quality of life.	Researchers should be cognizant of confounding variables that may influence the outcome of interest as they may obscure the true effect of interventions. This includes outlining any differences between groups and adjusting for these variables in analysis plans such as through regression analyses.
D2 - Bias in the selection of participants into the study	Lagares-Borrego et al. 201,6[63]: compared cost and clinical outcomes between two-stage expander/implant reconstruction and DIEP flap *RoB Rating: Serious*	This prospective cohort study outlined different minimum follow-up times between groups: 5 years for alloplastic and 2 years for DIEP reconstruction. Therefore, there are significant selection bias risks as follow-up time was selectively excluded between groups. As well, the start of data collection relative to the surgical course is unclear between the groups. As such, there is not enough information to determine that the beginning of data collection coincided for both groups, which may result in potential lead-time bias.	Selection bias may occur in both the selection of participants and follow-up time. Researchers should explicitly define the allocation of participants (including the rationale of allocation to specific groups) and ensure follow-up times coincide between groups.
D3 - Bias in the measurement of interventions	Bamba et al. 202,1[26]: compared early and standard expectations in DIEP procedures following enhanced recovery after surgery protocols and sought to determine differences in length of stay (LOS) *RoB Rating: Serious*	The standard group was poorly defined as patients were followed by two surgeons, who did not have a standardized expectation for LOS. This variability contradicts the group definitions set by authors as specific patients in the standard group may have had expectations relayed similar to the early group. The early group had a surgeon practice change in the last 6 months of the study, with the expectation changed from postoperative day 3 to 2. Although it is stated this was unrelated to the study, this change may have been influenced by knowledge of outcomes in this surgeon's practice.	Ideally, low RoB in this domain would imply comparability to a well-facilitated RCT, where the classification of intervention is determined at the time of delivery and explicitly defined. Authors must be particularly cognizant of this domain when defining study groups in retrospective observational studies, where intervention classification may be influenced by knowledge of outcomes. Although this is unlikely in surgical contexts, especially in prospective studies where interventions are defined at the time of delivery, this still poses a risk in specific interventions with a strong potential for recall bias.

(*continued on next page*)

Table 4 (*continued*)

ROBINS-I domain	Example	Rationale	Recommendation
D4 - Bias due to departures from intended interventions	Shakir et al. 202,1[29]: assessed postoperative complications between endoscopic, laparoscopic, and robotic-assisted harvest of the DIEP vessels *RoB Rating: Serious*	Among their cohorts, 1/94 endoscopic patients, 3/38 laparoscopic patients, and 3/3 robotic patients underwent a concurrent procedure, such as partial gastrectomy or total abdominal hysterectomy and bilateral salpingo-oophorectomy. These co-interventions were not standardized across groups, and possibly introduced performance bias in outcomes such as length of hospital stay	In the design of observational studies, it is imperative to standardize intervention protocols to prevent the introduction of co-interventions in prospective studies and consider excluding patients with co-interventions in retrospective studies.
D5 - Bias due to missing data	Beugels et al. 201,6[47]: compared complications and surgical outcomes between 322 unilateral and 208 bilateral DIEP flap surgeries *RoB Rating: Serious*	When comparing the intraoperative ischemic time between groups, 129/322 (40%) breasts in the unilateral group and 37/208 (18%) breasts in the bilateral group were not accounted for, without reasons for exclusion. The authors also stated that the missing data were not imputed in the analyses. Without providing more details regarding missing patients, RoB arises since there may be systematic differences between the groups, thereby skewing the treatment effect.	When missing data differs between interventions, it is pivotal to evaluate whether there is a systematic reason for this difference. If such a scenario arises, authors should also account for the differences with appropriate unbiased analyses (i.e., multiple imputation).
D6 - Bias the in measurement of outcomes	Chang et al. 20,20[8]: assessed subjective clinical parameters and patient-reported improvement between patients receiving simultaneous free DIEP flap with vascularized lymph node transfer, with or without lymphovenous anastomosis *RoB Rating: Critical*	The lymphovenous anastomosis group was assessed prospectively for subjective outcomes after surgery, while the control group's outcome data were collected retrospectively. Since outcome assessors were not blinded, there is a high risk for detection bias as there may be a decreased threshold for the detection of subjective outcomes prospectively. There may also be differences in data collection between groups given the difference in methodology.	We recommend authors pay particular attention to bias from the measurement of outcomes when evaluating subjective outcomes. Ideally, this should be conducted prospectively with assessor blinding of intervention status and standardization of collection process.
D7 - Bias in the selection of the reported results	Kerrebijn et al. 20,21[7]: assessed patient-reported outcome via BREAST-Q following bilateral mixed timing breast reconstruction vs. unilateral delayed reconstruction *RoB Rating: Serious*	In their methods section, the authors discuss comparing BREAST-Q scores for 3 quality of life domains and 3 satisfaction domains between groups at 12 months following their reconstruction surgery. However, in their results, they report BREAST-Q scores at the 18-month mark, which was not defined *a priori*. As well, the scores for 2 out of 6 domains were not reported for either group in their results section or tables.	To mitigate bias due to selective reporting, authors should utilize resources, such as clinicaltrials.gov, to publish a priori protocols, thereby increasing transparency to readers and reducing the exaggeration of treatment effects.

Note: RoB: Risk of bias.
DIEP: Deep inferior epigastric perforator.
*Citations included can be found in Appendix B.

with the findings from Igelström et al., which compiled the ROBINS-I assessment of 1344 observational studies from recent systematic reviews.[10] Similar to their study, Domain 1: bias due to confounding was at the highest RoB. Few studies have highlighted the considerations and applicability of ROBINS-I in specific domains.[11] This is necessary in order for researchers and readers to better recognize the limitations of the existing literature and work to improve this in the future study designs. Given the surgical nature of our included studies, there are unique considerations in each of the seven ROBINS-I domains that will be discussed further.

Given the focus of included studies was DIEP reconstruction, many general characteristics should be assessed within every study. Based on a recent study by Thorarinsson et al., smoking was associated with the highest number of early overall complications in DIEP reconstructions, while high body mass index was associated with the highest number of late overall postoperative complications.[12] As such, age, body mass index, smoking status, and comorbidities, such as diabetes mellitus, should be accounted for in all comparative breast reconstruction studies as they may confound the independent and dependent variables being assessed. Additional confounding variables were discussed, including the reception of radiotherapy or chemotherapy, depending on the relevancy to the specific outcome of interest. Notably, many studies did not consider one or more of these characteristics or had significant differences between groups at baseline that were not accounted for in subsequent analyses. These were deemed as moderate or severe RoB ($n = 36$, 55%) depending on the severity of the omission. In severe cases, studies were deemed at critical RoB ($n = 4$, 6.1%) if they did not present any baseline characteristics, or if there were significant differences between groups in characteristics such that they could not be compared. As this domain had the highest RoB among the ROBINS-I domains, we urge researchers to be cognizant of confounding variables that may influence the outcome of interest as they may obscure the true effect of interventions.[13] This includes outlining any differences between groups and adjusting for these variables in analysis plans such as through regression analyses.

Although around half of the included studies were deemed to be low RoB ($n = 34$, 52%) in the selection of participants into the study, 32% of studies were determined to have insufficient information. This was commonly seen when studies did not provide details regarding the allocation of patients, particularly in retrospective studies. As a result, study groups may be inherently prone to outcomes based on factors that led to the group allocation rather than the intervention itself. Important considerations that place this domain at risk include immortal time bias, which is a period of follow-up where the outcome cannot occur. Such concerns arise if the two groups may not necessarily coincide in the initiation of follow-up. For example, in studies comparing mastectomy and delayed reconstruction, the follow-up between mastectomy and delayed reconstruction may be unclear or excluded. In studies that assess for outcomes including disease recurrence, this may introduce a bias toward the safety of delayed reconstructions. To avoid this, authors should explicitly define the allocation of participants and ensure follow-up times coincide between groups.

Over half of the studies were judged to be at a low RoB in the classification of interventions ($n = 38$, 58%). The remaining studies mostly had moderate RoB or not enough information, due to aspects of intervention assignment being determined retrospectively or poorly defined interventions, respectively. For instance, Bamba et al. compared early and standard expectations in DIEP procedures following enhanced recovery after surgery (ERAS) protocols and sought to determine differences in length of stay (LOS).[14] The standard group was poorly defined as patients were followed by two surgeons, who did not have a standardized expectation for LOS. This variability contradicts the group definitions set by the authors as specific patients in the standard group may have had expectations relayed similar to the early group. As well, the early group had a surgeon practice change in the last 6 months of the study, with the expectation changed from postoperative day 3 to 2. Although it is stated this was unrelated to the study, this change may have been influenced by knowledge of outcomes (i.e., LOS) in this surgeon's practice. Ideally, low RoB in this domain would imply comparability to a well-facilitated RCT, where the classification of intervention is determined at the time of delivery and explicitly defined.[15] Authors must be particularly cognizant of this domain when defining study groups in retrospective observational studies, where intervention classification may be influenced by knowledge of outcomes. Although this is unlikely in surgical contexts, especially in prospective studies where interventions are defined at the time of delivery, this still poses a risk in specific interventions with a strong potential for recall bias.

We found that the majority of studies ($n = 52$, 79%) had low RoB due to departures from intended interventions. This is not surprising as the intervention status was determined at the time of delivery (i.e., the type of surgery received). Given the surgical nature of included studies, we only encountered two studies where bias may be introduced due to deviations from the intended "target" protocol. More commonly, bias arose in this domain due to the administration of co-interventions. For example, a study by Shakir et al., which assessed postoperative complications between endoscopic, laparoscopic, and robotic-assisted harvest of the DIEP vessels, was deemed at serious RoB.[16] Among their cohorts, 1/94 endoscopic patients, 3/38 laparoscopic patients, and 3/3 robotic patients underwent a concurrent procedure, such as partial gastrectomy or total abdominal hysterectomy and bilateral salpingo-oophorectomy. These co-interventions were not standardized across groups and possibly introduced performance bias in outcomes, such as length of hospital stay.[15] In the design of observational studies, it is imperative to standardize intervention protocols to prevent the introduction of co-interventions in prospective studies and consider excluding patients with co-interventions in retrospective studies.

Data were completely reported for the majority of studies ($n = 40$, 61%). Uncommonly, studies were rated as either moderate RoB due to missing data ($n = 7$, 11%) if missing data were balanced evenly across all groups or serious ($n = 6$, 9.1%) if missing data were uneven between groups and were not addressed appropriately in the analysis. For example, in a study by Beugels et al., they sought to compare complications and surgical outcomes between 322 unilateral and 208 bilateral DIEP flap surgeries.[17]

However, when comparing the intraoperative ischemic time between groups, 129/322 (40%) breasts in the unilateral group and 37/208 (18%) breasts in the bilateral group were not accounted for, without reasons for exclusion. The authors also stated that the missing data were not imputed in the analyses, placing this study at serious RoB in this domain. Without providing more details regarding missing patients, RoB arises since there may be systematic differences between the groups, thereby skewing the treatment effect. Thus, when missing data differs between interventions, it is pivotal to evaluate whether there is a systematic reason for this difference. If such a scenario arises, authors should also account for the differences with appropriate unbiased analyses (i.e., multiple imputation).[18]

With regard to bias in the measurement of outcomes, the majority of studies ($n = 49$, 74%) were classified as low RoB, with only 5 (7.6%) identified to be serious or critical RoB. When assessing for clinical outcomes, such as complications, there is often low RoB given the objective classification. This is especially true for retrospective studies, which comprised 77% of our included studies, as identification of complications follows routine clinical practice and guidelines. The studies most prone to bias in outcome measurement were those that assessed subjective or patient-reported outcomes between groups. This is evident in a study by Chang et al., which assessed subjective clinical parameters and patient-reported improvement between patients receiving simultaneous free DIEP flap with vascularized lymph node transfer and with or without lymphovenous anastomosis.[19] This study was deemed at critical RoB because the lymphovenous anastomosis group was assessed prospectively for subjective outcomes after surgery, while the control group's outcome data were collected retrospectively. Since outcome assessors were not blinded, there is a high risk for detection bias as there may be a decreased threshold for the detection of subjective outcomes prospectively. There may also be differences in data collection between groups given the difference in methodology. As such, we recommend authors pay particular attention to bias from the measurement of outcomes when evaluating subjective outcomes. Ideally, this should be conducted prospectively with assessor blinding of intervention status and standardization of collection process.

Similarly, the majority of studies ($n = 53$, 80%) were at low RoB in the selection of the reported result. Although many studies reported registration with local ethics boards, this domain was difficult to assess as most studies did not publish their protocols, so bias was determined largely on statistical analysis plans described in the methods sections. Studies with risk in this domain were mostly moderate, as they either reported results of analyses that were not previously mentioned or selectively reported certain significant analyses. This can mislead clinicians into overestimating the benefits of an intervention or underestimating potential safety concerns. To mitigate bias due to selective reporting, authors should utilize resources, such as clinicaltrials.gov, to publish *a priori* protocols, thereby increasing transparency to readers and reducing the exaggeration of treatment effects.

Despite the standardized signaling questions and judgment for RoB within domains, there are limitations in applying ROBINS-I to surgical observational studies. During our analysis, assessors had difficulties coming to a consensus for certain domains, notably Domain 2: bias due to selection of participants and Domain 4: bias due to deviations from intended interventions. This is in keeping with a study by Thomson et al., who found the signaling questions for these domains can be interpreted in different ways to result in varying judgments of RoB. As well, we set an arbitrary cutoff to exclude studies published in journals with an impact factor of less than 1.0 to limit the number of studies captured within our search. Previous studies have shown that publications in lower impact factor journals may be associated with lower methodological quality.[20] By excluding these studies, we may be underestimating the overall RoB in DIEP breast reconstruction studies.

Conclusions

In this review, we demonstrate the high RoB of observational studies evaluating DIEP breast reconstruction, which may jeopardize the validity of findings. Given that these studies comprise the majority of the literature on the topic, observational studies should be designed with adequate transparency and consideration of various biases. We recommend that authors consult the ROBINS-I tool not only when assessing observational studies for systematic reviews but also when designing or conducting these studies. In our present study, we present additional considerations for each domain to provide researchers with guidance on the assessment and conduction of observational studies.

Declaration of Competing Interest

There are no conflicts of interest for any authors.

Funding

The authors have no financial interest to declare in relation to the content of this article. No external funding was received.

Ethical approval

Not required.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.bjps.2022.06.093.

References

1. Doherty C, Pearce S, Baxter N, et al. Trends in immediate breast reconstruction and radiation after mastectomy: a population study. *Breast J* 2020;**26**(3):446–53.

2. Wilkins EG, Cederna PS, Lowery JC, et al. Prospective analysis of psychosocial outcomes in breast reconstruction: one-year postoperative results from the Michigan breast reconstruction outcome study. *Plast Reconstr Surg* 2000;**106**(5):1014–25.
3. Macadam SA, Zhong T, Weichman K, et al. Quality of life and patient-reported outcomes in breast cancer survivors: a multicenter comparison of four abdominally based autologous reconstruction methods. *Plast Reconstr Surg* 2016;**137**(3):758–71.
4. Heidekrueger PI, Moellhoff N, Horch RE, et al. Overall complication rates of DIEP flap breast reconstructions in Germany—a multi-center analysis based on the DGPRÄC prospective national online registry for microsurgical breast reconstructions. *J Clin Med* 2021;**10**(5):1016.
5. Thorarinsson A, Fröjd V, Kölby L, et al. A retrospective review of the incidence of various complications in different delayed breast reconstruction methods. *J Plast Surg Hand Surg* 2016;**50**(1):25–34.
6. Bennett KG, Qi J, Kim HM, Hamill JB, Pusic AL, Wilkins EG. Comparison of 2-year complication rates among common techniques for postmastectomy breast reconstruction. *JAMA Surg* 2018;**153**(10):901–8.
7. Hassanein AH, Herrera FA, Hassanein O. Challenges of randomized controlled trial design in plastic surgery. *Can J Plastic Surg* 2011;**19**(3):e28.
8. Winters ZE, Emson M, Griffin C, et al. Learning from the QUEST multicentre feasibility randomization trials in breast reconstruction after mastectomy. *Br J Surg* 2015;**102**(1):45–56.
9. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;**355**:i4919.
10. Igelström E, Campbell M, Craig P, Katikireddi SV. Cochrane's risk of bias tool for non-randomized studies (ROBINS-I) is frequently misapplied: a methodological systematic review. *J Clin Epidemiol* 2021;**140**:22–32.
11. Thomson H, Craig P, Hilton-Boon M, Campbell M, Katikireddi SV. Applying the ROBINS-I tool to natural experiments: an example from public health. *Syst Rev* 2018;**7**(1):15.
12. Thorarinsson A, Fröjd V, Kölby L, Lidén M, Elander A, Mark H. Patient determinants as independent risk factors for postoperative complications of breast reconstruction. *Gland Surg* 2017;**6**(4):355–67.
13. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J* 2012;**3**(1):9–12.
14. Bamba R, Wiebe JE, Ingersol CA, et al. Do patient expectations of discharge affect length of stay after deep inferior epigastric perforator flap for breast reconstruction? *J Reconstr Microsurg* April 2021.
15. Sterne J.A.C., Hernan M.A., Reeves V.C., et al. Risk of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance, updated 12 October 2016. Available from http://www.riskofbias.info. [Accessed December 21, 2021].
16. Shakir S, Spencer AB, Piper M, Kozak GM, Soriano IS, Kanchwala SK. Laparoscopy allows the harvest of the DIEP flap with shorter fascial incisions as compared to endoscopic harvest: a single surgeon retrospective cohort study. *J Plast Reconstr Aesthet Surg* 2021;**74**(6):1203–12.
17. Beugels J, Hoekstra LT, Tuinder SMH, Heuts EM, van der Hulst RRWJ, Piatkowski AA. Complications in unilateral versus bilateral deep inferior epigastric artery perforator flap breast reconstructions: a multicentre study. *J Plast Reconstr Aesthet Surg* 2016;**69**(9):1291–8.
18. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;**338**:b2393.
19. Chang EI, Ibrahim A, Liu J, et al. Optimizing quality of life for patients with breast cancer-related lymphedema: a prospective study combining DIEP flap breast reconstruction and lymphedema surgery. *Plast Reconstr Surg* 2020;**145**(4):676e.
20. Ahmed Ali U, Reiber BMM, Ten Hove JR, et al. Journal impact factor and methodological quality of surgical randomized controlled trials: an empirical study. *Langenbecks Arch Surg* 2017;**402**(7):1015–22.