Directed acyclic graphs in perioperative observational research – a systematic review and critique against best practice recommendations

Assessing directed acyclic graph use and construction in perioperative literature

Dr Matthew Lamont Watson[1], Mr Sebastian H. M. Hickman[2,3], Ms Kaya Marlen Dreesbeimdiek[4], Dr Katharina Kohler[5] and Dr Daniel J. Stubbs[4,6]*

1: Clinical School of Medicine, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ

2:Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW

3: The Alan Turing Institute, 96 Euston Road, London, NW1 2DB

4: Health Systems Design Group, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ

5: Clinical Lecturer, University Division of Anaesthesia, University of Cambridge, Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ

6: Wellcome Trust Research Fellow, University Division of Anaesthesia, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ

*Corresponding author

Email: djs225@cam.ac.uk

Abstract

The Directed Acyclic Graph (DAG) is a graph representing causal pathways for informing the conduct of an observational study. The use of DAGs allows transparent communication of a causal model between researchers and can prevent over-adjustment biases when conducting causal inference, permitting greater confidence and transparency in reported causal estimates. In the era of 'big data' and increasing number of observational studies, the role of the DAG is becoming more important.

Recent best-practice guidance for constructing a DAG with reference to the literature has been published in the 'Evidence synthesis for constructing DAGs' (ESC-DAG) protocol. We aimed to assess adherence to these principles for DAGs constructed within perioperative literature.

Following registration on the International Prospective Register of Systematic Reviews (PROSPERO) and with adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting framework for systematic reviews, we searched the Excerpta Medica dataBASE (Embase), the Medical Literature Analysis and Retrieval System Online (MEDLINE) and Cochrane databases for perioperative observational research incorporating a DAG. Nineteen studies were included in the final synthesis. No studies demonstrated any evidence of following the mapping stage of the protocol. Fifteen (79%) fulfilled over half of the translation and integration one stages of the protocol. Adherence with one stage did not guarantee fulfilment of the other. Two studies (11%) undertook the integration two stage. Unmeasured variables were handled

inconsistently between studies. Only three (16%) studies included unmeasured variables within their DAG and acknowledged their implication within the main text.

Overall, DAGs that were constructed for use in perioperative observational literature did not consistently adhere to best practice, potentially limiting the benefits of subsequent causal inference. Further work should focus on exploring reasons for this deviation and increasing methodological transparency around DAG construction.

Introduction

The ability to ask causal questions of observational data is one of the core tasks of data analysis, alongside description and prediction [1]. Historically, healthcare researchers have refrained from explicitly seeking such findings [2], with an emphasis on the randomised controlled trial (RCT) as the only methodological framework capable of supporting such claims [1]. In many ways this was an understandable response to the potential for observational data analysis to suffer from various biases, including confounding and selection bias [3].

Since the late 1990s, frameworks for conducting robust, transparent causal research using observational data has emerged and become increasingly formalised across fields [4]. One of these frameworks represents variables of interest in a graphical format, underpinned by a robust mathematical framework [5]. These diagrams are termed directed acyclic graphs (DAGs) and consist of nodes (which represent the variables of interest) connected by directed edges (arrows). DAGs can be useful in a plethora of settings within healthcare research. This includes their use within interventional studies (such as RCTs), and observational research.. To be valid for causal inference, DAGs must denote all causal relationships between nodes, specifically the root causes of any variable pair, even if these are unmeasured in available data [4]. Such unmeasured variables are termed ‘latent variables’. The use of DAGs allows transparent communication of a putative causal model between researchers and can prevent over-adjustment biases when conducting multivariable modelling. This could permit a greater degree of accuracy and confidence in calculated causal estimates.

Such a framework has clear benefits for perioperative and surgical researchers as healthcare data is continuing to grow [6], and perioperative care specifically generates a wealth of data pertaining to physiological, pharmacological, and broader perioperative

events. A robust framework to seek causal inferences from this data is of huge benefit when studying relationships not amenable to an RCT [7]. An introduction to DAGs, their relevance to perioperative and surgical care, and the potential biases they can be used to identify has already been published [8,9]. However, to our knowledge no study has yet sought to systematically catalogue the methods used to generate DAGs used in perioperative research and what problems they have been used to address.

This is, however, vital, as DAGs are increasingly used in other areas of health research [10]. The need for transparency and robustness in their generation is of increasing importance. A best practice framework for constructing DAGs has been recently published. The 'Evidence synthesis for constructing DAGs' (ESC-DAG) [11] protocol provides a recommended methodological approach to DAG construction. It consists of four main stages to build robust DAGs, drawing on previously published studies and containing clear instructions on how to synthesise this knowledge into a unifying causal model of the relationship under study. The ESC-DAG protocol contains four distinct stages. These stages include mapping (where relevant variables are identified from other literature studies), translation (where the relationships between variables are identified), integration 1 (where a single DAG is constructed with only the relevant relationships included) and integration 2 (where similar variables within the DAG are grouped together).

The primary aim of this systematic review is to provide the first comprehensive assessment of how DAGs are used and constructed within the perioperative literature, cataloguing the questions they are being used to address, their construction, and thus transparency. This was done by assessing DAG construction against the ESC-DAG protocol.

Our study also had two major secondary aims. Firstly, we examined how included studies handled unmeasured variables. This is vital as a DAG's validity relies on its accurate inclusion of such latent variables [10]. Additionally, variable relationships within DAGs can also be modelled over time. This is becoming of increasing relevance given that electronic records can permit the repeated measurements of variables (such as blood results or physiological observations) through time. Therefore, our other secondary aim involved assessing how authors were incorporating time as a variable into their DAGs.

Methods

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [12]. The 2020 PRISMA checklist can be found in the supplementary information. The aims and methods of our systematic review were registered prospectively on PROSPERO (CRD42021279183). No ethical approval was required for this study.

Published observational studies in the surgical and perioperative literature that had constructed a DAG were included in the study. This included observational studies that discussed any aspect of surgery or perioperative medicine.

We searched Embase, MEDLINE and Cochrane databases. The search protocol for each database is available as S1 Appendix but broadly consisted of keywords relating to 'Surgery', 'Anaesthesia', 'Perioperative', and 'Directed Acyclic Graph'. Following an initial search on the 7th of September 2021, we updated our search on 30th June 2022. The PRISMA flow chart is shown in Fig 1. All abstracts were screened independently by DS and MW. Abstracts were eligible for further full text screening if the title, abstract, or related keywords on the citation, suggested the study was a primary observational research study containing a DAG. This involved searching the abstract and titles for words such as 'DAG', 'graph', 'acyclic' and other language pertaining to causal inference. If the author was unsure, they were asked to include the abstract so the full text and supplementary material could be screened. If there were any disagreements, a third author (KK) was asked to adjudicate.

Fig 1. PRISMA flowchart of studies included in the final synthesis of the systematic review.

A modified inductive reasoning approach was taken to assess the nineteen studies included in the review [18]. This is outlined in Table 1. This approach was undertaken so we could follow a clear and logical framework to reach our final conclusions as a group. We used this method to qualitatively assess our studies in a transparent and repeatable way. The individual observation phase involved two authors scoring each study independently against the steps laid out in the ESC-DAG protocol. These stages and steps are depicted in Figs 2 and 3. This involved scoring the study either a 1 or 0 depending on whether the study had provided evidence that they had fulfilled that specific step. An initial agreement rate between the authors after individual observations was quantified using Cohen's kappa statistic [19]. Group calibration was then undertaken to discuss the evidence for scoring each paper with a 1 or 0 against the specific steps of DAG construction. Unanimous group decisions about disagreements were made at this phase. A final group conclusion about the construction of DAGs within the surgical and perioperative literature was then made.

Fig 2. An outline of the stages of the ESC-DAG protocol [11], including an example construction of a perioperative DAG.

Fig 3. Stages, purpose, and the steps of DAG construction outlined in the ESC-DAG protocol [11] that the reviewers scored the final papers against. IG: implied graph (a DAG that acts as a structural template to build the DAG).

The specific stages and steps we scored DAG construction against, as well as the purpose of each stage, are outlined in Figs 2 and 3. They are also discussed in detail in the ESC-DAG protocol [11].

A data collection form was constructed and reviewed by all authors before data extraction. The data collection form and individual observations of all nineteen papers can be found in S2 Appendix.

		PROCESS FOLLOWED BY THE AUTHORS
STAGE OF MODIFIED INDUCTIVE REASONING	**INDIVIDUAL OBSERVATION**	Each DAG in the literature was scored against the ESC-DAG protocol with a 1 or a 0 depending on whether it had met the steps laid out in the ESC-DAG protocol.
	GROUP CALIBRATION	We then discussed findings, and a consensus was reached between all authors about the patterns of DAG construction between the studies included. Any disagreements were also debated.
	FINAL DECISION	We made a conclusion on the general construction of DAGs in this group of papers when scored against the ESC-DAG protocol.

Table 1. Process followed by the authors to collect the data and score the DAGs in the studies against the ESC-DAG protocol [11]**.**

We also extracted key study characteristics including publication date relative to the ESC-DAG protocol and whether the study had included and discussed their handling of latent nodes. If the study incorporated time as a variable into their DAG, this was also recorded. This was so we could comment on how these aspects of DAG construction were undertaken, with reference to the ESC-DAG protocol.

Finally, each study was assessed by two authors independently against the Newcastle-Ottawa scale to assess potential risks of bias and to assess the quality of the final studies

included [20]. Scores can be found in Table 2 and the full scoring can be found in S3 Appendix.

Results

A. Characteristics of the final studies included

We identified 638 citations after the removal of duplicates. After abstract and full text screening, nineteen (3%) were included in our final synthesis. We included one study published in German which was read and reviewed by two native speakers on our study team (KMD, KK). We did not include studies that sought to generate a Bayesian network.

Once abstracts had been screened; fifty-four studies remained and were assessed for eligibility by screening the full text and supplementary material for a DAG. After this round of full text screening, nineteen studies were included in the final review. Five of these nineteen studies [13–17] were identified during the updated search in June 2022. References of the final nineteen studies were also screened for relevant studies. One additional study was included via this method (Fig 1).

Although the terms Bayesian network and DAG can both be used in the literature to refer to a causal graphical model, Bayesian networks do not necessarily describe causal relationships and typically do not solely use expert opinion in their construction [21], they are therefore not explicitly covered by the ESC-DAG framework. However, one screened study [22] that constructed a Bayesian network was included, as the graphical core of their diagram describes causal relationships and was constructed solely from expert knowledge derived from literature and clinical practice. Therefore, we decided to include this study as the construction of this graph could have adhered to ESC-DAG protocol.

Of the final nineteen studies reviewed by two reviewers, the median number of stars obtained on the Newcastle-Ottawa scale was seven and the interquartile range was one. The lowest score on the Newcastle-Ottawa scale was five. The score of each paper is included in Table 2 and full scoring of all the nineteen papers can be found in S3 Appendix.

Table 2 shows the journals and years of publication for the studies included in this systematic review. Fifteen (79%) of the nineteen studies included were published after the ESC-DAG protocol [11] was available online on the 19th of July 2019. No study referenced the ESC-DAG protocol.

Twelve studies (63%) included variables on their DAG with a temporal element. For instance, variables with relevant time dimensions such as 'preoperative use of gabapentin' [16].

Study	Year Published	Journal	Study Objective	Published after the ESC-DAG protocol?	Did the study include nodes with a temporal element?	Newcastle Ottawa Scale scoring for each study
Ferrando et al. [13]	2022	Acta Anaesthesiologica Scandinavica	To investigate whether an intraoperative open lung condition reduces the risk of developing a composite of postoperative pulmonary complications.	Y	Y	7
Kalkan et al. [15]	2022	Angiology	To determine whether the uric acid/albumin ratio is a predictor of mortality in STEMI patients.	Y	N	7
Lam et al. [23]	2022	Anaesthesia	To clarify the association between glycated haemoglobin and postoperative outcomes in people without an existing diagnosis of diabetes.	Y	N	7

Nimmaanrat et al. [16]	2022	BMC Anaesthesiology	To find the factors for predicting the amount of opioid requirement post-surgery.	Y	Y	7
Qian et al. [14]	2022	International journal of clinical practice	To clarify the efficiency and outcomes of suctioning ureteral access sheath during flexible ureteroscopic lithotripsy for the management of renal stones.	Y	Y	8
Steele et al. [17]	2022	Frontiers in human neuroscience	To quantify the causal effects of altered motor control and other impairments on gait, before and after single-event multi-level orthopaedic surgery.	Y	Y	7
Bedir et al. [24]	2021	Journal of Cancer Research and Clinical Oncology	To quantify the effect of socioeconomic inequality on head and neck cancer survival.	Y	N	8
Cagigas et al. [22]	2021	International Journal of Colorectal Disease	To predict intra-abdominal infections after colorectal surgery.	Y	Y	5

Laitinen et al. [25]	2021	Acta Orthopaedica	To investigate implant survival and complications of different surgical strategies in the treatment of proximal tibia pathological fractures.	Y	Y	6
Tetta et al. [26]	2021	Surgical Oncology	To predict lung recurrence and disease-specific mortality after pulmonary metastasectomy for soft tissue sarcoma.	Y	Y	6
Wang et al. [27]	2021	BMJ Open	To analyse the clinical value of primary site surgery in improving the cancer-specific survival and overall survival of initial metastatic cervical cancer patients.	Y	N	8
Duprey et al. [28]	2020	Journal of the American Geriatrics Society	To provide researchers with guidance on the methodological tools to use data from clinical cohorts to better understand medication risk factors and outcomes.	Y	N	8
Pollmann et al. [29]	2020	Acta Orthopaedica	To identify the contribution of early deep surgical site infection to mortality after hip fracture surgery.	Y	Y	8

Sittivarakul et al. [30]	2020	Medicine	To determine the surgical outcomes and prognostic factors of cytomegalovirus retinitis-related retinal detachment in acquired immune deficiency syndrome patients following vitrectomy.	Y	Y	8
Hoorntje et al. [31]	2019	Orthopaedic Journal of Sports Medicine	To investigate the extent and timing of return to work in the largest high tibial osteotomy cohort investigated for return to work and to identify prognostic factors for return to work after high tibial osteotomy.	Y	Y	8
Kerkhoffs et al. [32]	2019	The American Journal of Sports Medicine	To investigate the extent and timing of return to sport after high tibial osteotomy in the largest cohort investigated for return to sport to date and to identify prognostic factors for successful return to sport.	N	Y	7
Pathak et al. [33]	2018	Hospital paediatrics	To identify predictors of oophorectomy in girls hospitalized throughout Texas with ovarian torsion.	N	N	7

Asgari et al. [34]	2011	Archives of Dermatology	To identify preoperative, intraoperative, and postoperative variables that predict higher short and long-term patient satisfaction with Mohs surgery.	N	Y	7
Sehrndt et al. [35]	2011	Gesundheitswesen	To investigate the influence of socioeconomic status on health-related quality of life in patients before and after aortocoronary bypass surgery.	N	N	7

247

248 **Table 2. Table depicting the year, journal, study objective and publication date relative to the ESC-DAG protocol of the final nineteen**
249 **studies. Whether the DAG within the study included a temporal variable is also recorded. A study including temporal variables means**
250 **this study acknowledged or specified that a certain variable e.g., drug usage was measured at a specific time point in the preoperative**
251 **pathway.** STEMI, ST-elevation myocardial infarction.

252

We also assessed the geographical distribution of the final studies included. The studies were written by four authors affiliated in the USA (21%), two in China (11%), two in Germany (11%), two in Spain (11%), two in the Netherlands (11%), two in Thailand (11%), one in Turkey (5%), one in the United Kingdom (5%), one in Norway (5%), one in Finland (5%) and one in Italy (5%).

B. Assessing DAG construction against the ESC-DAG protocol

Following individual observation and review of each paper, the initial agreement rate between reviewers was 0.86 (as calculated using a Cohen's Kappa statistic). This suggests a high level of initial agreement between all reviewers when independently scoring the papers during the individual observation phase of our analysis [36].

Table 3 contains the individual scoring of each paper against the ESC-DAG protocol after a consensus was reached. It also demonstrates the patterns identified and general conclusion about the whole cohort of studies assessed. Even though no studies referenced the ESC-DAG protocol, studies did show evidence of fulfilling some stages and steps of the protocol. No studies showed evidence of completing any steps of mapping. Fifteen (79%) of the studies completed 50% or more of the translation and integration one steps. There was a large variation between the studies as to the number of steps fulfilled within each of these two stages. Two (11%) of the final studies fully completed integration stage two; no other studies completed any steps within this stag

275

			PAPERS (Lead Author)																		
			Asgari [34]	Bedir [24]	Cagigas [22]	Duprey [28]	Ferrando [13]	Hoorntje [31]	Kalkan [15]	Kerkhoffs [32]	Laitinen [25]	Lam [23]	Nimmaanrat [16]	Pathak [33]	Pollmann [29]	Qian [14]	Sehrndt [35]	Sittivarakul [30]	Steele [17]	Tetta [26]	Wang [27]
STAGE OF MODIFIED INDUCTIVE REASONING	**INDIVIDUAL OBSERVATIONS**	**Percentage of steps in the ESC-DAG protocol each study has fulfilled during DAG construction**	Mapping (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Translation (%)	75	100	75	75	75	75	25	62.5	0	50	50	50	25	75	50	75	62.5	0	50
			Integration 1 (%)	60	100	0	40	90	80	80	80	100	70	60	0	90	100	70	60	100	20	80
			Integration 2 (%)	0	100	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0

	GROUP CALLIBRATION	**Key group observations on the construction of the DAG in all studies using the ESC-DAG protocol as guidance**	• No studies demonstrated any evidence of completing any steps of mapping. • Fifteen studies (79%) carried out 50% or more of the steps within the translation stage. The steps were discussed by some papers, but with strong variation amongst the information presented. Only one study showed evidence of performing a counterfactual thought experiment (one of the steps within this stage). • Fifteen studies (79%) undertook 50% or more of the steps of integration one; but did not discuss this in their text. • Only two studies (11%) undertook and discussed integration two.
	FINAL DECISION	**Overall conclusion of DAG construction from scoring against the ESC-DAG protocol**	No study displayed evidence of completely following the ESC-DAG protocol. Mapping (structured synthesis of previously published studies) was never documented. Overall, DAGs constructed in this selection of papers did not consistently adhere to the ESC-DAG protocol.

Table 3. Table demonstrating the results from the screening of all nineteen papers when assessed against the ESC-DAG protocol [11]**.**

C. Unmeasured variables

Table 4 demonstrates which studies acknowledged and discussed their methods for including unmeasured variables. Six studies (32%) did not include, acknowledge, or discuss any unmeasured variables. Seven studies (37%) acknowledged unmeasured variables within their study. However, none of these studies included these unmeasured variables in their DAG or discussed how unmeasured variables were managed or might impact on the interpretation of their findings. Two studies (11%) included unmeasured variables within their DAG. No study in these categories commented on how they tried to mitigate for any unmeasured variables, despite the acknowledgement that potentially key variables were not measured or incorporated in their DAG.

Three studies (16%) included and discussed how unmeasured variables would affect their study. For instance, one study [24] aimed to quantify the effect of socioeconomic inequality on head and neck cancer survival. The authors acknowledged that human papillomavirus (HPV) status was an important unmeasured variable. Their approach was to infer HPV status based on tumour site (HPV-related sites vs HPV non-related sites), using another, measured variable to attempt to represent the uncaptured latent. This study referenced the scientific literature they used as evidence to make this inference.

One study [35] (5%) built their DAG and then collected conclusive data for their variables. This study therefore did not have any unmeasured variables.

Table demonstrating which studies included, acknowledged, and made corrections for unmeasured variables				
Unmeasured variables not mentioned 32% (n=6)	Unmeasured variables acknowledged; not included in the DAG or discussed in the study 37% (n=7)	Unmeasured variables included in the DAG; not discussed in the study 11% (n=2)	Unmeasured variables included in the DAG and discussed in the study 16% (n=3)	DAG constructed before data was collected; no unmeasured variables included 5% (n=1)
Cagigas et al. [22] Duprey et al. [28] Kalkan et al. [15] Lam et al.	Asgari et al. [34] Ferrando et al. [13] Laitinen et al. [25] Qian et al.	Hoorntje et al. [31] Kerkhoffs et al. [32]	Bedir et al. [24] Pathak et al. [33] Pollmann et al. [29]	Sehrndt et al. [35]

[23] Nimmaanrat et al. [16] Tetta et al. [26]	[14] Sittivarakul et al. [30] Steele et al. [17] Wang et al. [27]			

Table 4. Table demonstrating the extent to which different studies dealt with unmeasured variables, if at all

Discussion

In this systematic review, we provide the first comprehensive critique of the methods used in developing DAGs in perioperative and surgical research. Our main finding is that, at present, DAG construction does not consistently follow the newly published ESC-DAG protocol [11] and other elements of best practice [10]. This demonstrates the importance of newly emerging guidelines to strengthen practice in this area. Our findings are congruous with other literature regarding DAG usage and construction in observational studies [37] which could potentially limit the benefits of employing this approach if readers and reviewers cannot understand the assumptions that were used to underpin subsequent causal inferences.

Although our search strategy was not designed to generate a comprehensive denominator on the total number of conducted perioperative studies it is important to highlight the small number of studies included in our final synthesis (3% of initial citations), despite its comprehensiveness (three databases) and intended sensitivity. We found this surprising given the number of studies published in recent years that emphasise how important and useful 'big data' can and might be within this field [38–40]. No study specifically mentioned the ESC-DAG protocol, including fifteen studies that were published after the ESC-DAG protocol [11] was made available. This strongly suggests that no studies in our cohort were trying to explicitly follow this guidance and could reflect either a lack of awareness, a lag in its adoption into practice, or a lack of comfort in implementing and following its steps. However, all studies demonstrated evidence of fulfilling at least one or more of the stages and steps identified in the protocol.

We can only speculate as to why so few observational studies within this field utilise DAGs, especially as relevant methodological reviews have been published in the field of anaesthesia and surgery [8,9,41]. This may represent broader trends in biomedical observational research [10]. Alternatively, researchers may be aware of DAGs; but uncomfortable using these tools when they have a limited understanding of when and how to use them. However, more education and awareness may be required before researchers can use DAGs comfortably and to their full potential, especially because, as we have documented, there are currently few examples of such best-practice frameworks being implemented or reported in the literature.

Significantly, no study appeared to undertake the 'mapping' stage whereby a directed graph is drawn for relevant, previously published studies examining the same exposure and outcome relationships. Such a step would be a transparent way to communicate the controlling of 'all known confounders' that is often mentioned in observational research. It is plausible that the authors of the final nineteen studies built several template DAGs before arriving at the one presented in the study. However, none of these template DAGs were made available and thus the author's thought processes cannot be followed. Therefore, the studies did not score anything when we were scoring them against the mapping stage of the protocol.

Fifteen of the studies (79%) carried out 50% or more of the translation stage. Translation exists to confirm which relationships in a template DAG depict a causal relationship. Our included studies demonstrated a strong variation in their discussion behind the relationships depicted in their DAG. Some studies did not qualify the causal relationships

within their DAG at all [25,26], whereas one study fulfilled all stages of translation and qualified the DAG they constructed robustly [24]. Given that a DAG is only as valid as the assumptions which underpin its construction, we would argue that even if authors are not following the ESC-DAG protocol, it is crucial that there is some discussion within a study that justifies the causal relationships depicted on a DAG. This allows readers to understand why specific relationships have been included and further understand and critique the statistical analysis within a study that the DAG guides. One strength of the DAG approach is that its robust mathematical underpinning is communicated to readers in an accessible pictorial format. However, if the rationale for its construction is not clearly communicated it is difficult for researchers to interpret results and critique findings based on the depicted relationships.

Seventeen studies (89%) undertook at least some steps of integration stage one but did not discuss this in their text. Integration one exists to combine the causal relationships identified in the translation stage into a DAG. During this stage authors build the final DAG that they will use within their study. Authors may not have discussed this stage within their text as they may have felt that they did not need to describe how they built their DAG from exposure to outcome, to the final DAG identifying all other causal relationships within their system.

Furthermore, only two studies (11%) undertook and discussed integration stage two. Integration stage two exists to combine nodes in the final DAG for either practical or substantive reasons. Like mapping, we believe studies could have undertaken integration two but may not have provided any evidence of completing them within the text. It is plausible

that authors of the studies we analysed were combining nodes within their DAG. This is a step of the integration two stage in the protocol that only two studies demonstrated any evidence of following. However, we acknowledge that more studies may have combined nodes without making this explicit in the text. An example of this could include combining many health conditions within a study as a 'co-morbidities' node on a DAG. Direct communication of these assumptions would arguably be of use to reviewers or other researchers attempting to replicate findings in other settings.

The ESC-DAG protocol does not perhaps encompass all elements of best-practice. For instance, it does not make explicit mention to the inclusion or handling of unmeasured variables, a crucial step in ensuring a DAG is causally valid [4]. Explicitly depicting unmeasured variables on a DAG helps to highlight any potential sources of unobserved confounding. It also allows authors to fully communicate their causal model, which can further lead to debate and inform future research. Some studies included their unmeasured variables on their DAG but there was wide variation between the studies analysed. We found that some studies did not acknowledge any unmeasured variables whilst others incorporating unmeasured variables into their DAG and gave detailed discussion as to how they had attempted to control for these factors. One study [35] communicated that they had built their DAG before collecting conclusive data for all the variables included. It is ideal if a DAG can be constructed using variables that have data available for them. However, omitting variables from a DAG because authors believe they will not be able to collect conclusive data for a variable goes against best practice [10]. Further to the above, it is vital that those constructing DAGs include all variables, unmeasured or measured, that may have causal relationships to the system being studied.

406

407 Another additional element of best practice we examined was to understand how time
408 was displayed or handled within published DAG studies. Perioperatively, the explicit
409 mention of the time at which a variable is measured is vital, with a low pre-operative
410 haemoglobin having a completely different interpretation and causal relationship than one
411 immediately after surgery. A further relevant example of this may include the effect of
412 arterial blood pressure (ABP) on three-day postoperative mortality. Intraoperative ABP
413 pressure may have a different causal effect on mortality than preoperative ABP. Therefore, it
414 is important researchers acknowledge how time may change their causal relationships
415 depicted. We found that twelve studies (63%) acknowledged the time point of a variable
416 within the group of nineteen studies analysed. Further to this, we thought that *Asgari et al.*
417 [34] presented a very effective way to present time within a perioperative DAG, with nodes
418 grouped according to 'pre', 'intra', and 'post' operative phases of care.

419

420 The ESC-DAG protocol gives researchers a strong framework to build DAGs. This is
421 a clear development in the field where best practice for DAG construction did not exist [37].
422 However, its relatively recent publication means that a critique of its impact on transparency
423 and reproducibility cannot yet be commented upon. Beyond the core stages of the protocol
424 that we have discussed so far, the concept of a 'directed edge index' and 'decision log' to
425 communicate assumptions around edge incorporation and direction could be a valuable tool
426 to support the open-science framework. The appendix of the ESC-DAG protocol asks
427 researchers to explicitly mention what their outcome, exposures, controls, and mediators are,
428 to document their assumptions around edge inclusion, and fully communicate their rationale.
429 Based on our literature findings we feel the use of such a reporting template (perhaps as

supplemental material) would be useful in increasing the transparency of assumptions underlying causal analyses, in a manner analogous to the 'target-trial' [42] framework that is advocated to design a causally focused observational study.

However, to our knowledge we are not aware of any surgical and perioperative journal that requires observational studies to construct a DAG if they are seeking to make a causal interpretation. The New England Journal of Medicine (NEJM) [43] explicitly state that *'causal language should not be used in observational studies where only associations can be estimated'*. However, The NEJM does not mention the DAG when discussing methods observational studies should use to explicitly discuss and estimate a causal effect. The guidance for authors publishing in the British Journal of Anaesthesia [44] and British Journal of Surgery [45] do not expect authors to have constructed DAGs. Both journals ask authors publishing observational studies to adhere to the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) [46] guidelines. Further to this, the STROBE guidelines ask authors to 'describe any efforts to address potential source of bias' in the twenty-two-item checklist provided. However, there is no mention as to which methodological approach authors should use to deal with bias and there is no mention of DAGs. This further contributes to our point that the surgical and perioperative field may not be aware of how the use of a DAG could strengthen confidence in causal inferences within observational research.

Given the benefits of DAGs in guiding analysis, their greater use in the perioperative literature could arguably improve the robustness of studies seeking to draw causal inferences. Beyond this improved accuracy, submitting information justifying the reasons behind a constructed causal model could also improve communication of assumptions from

researchers to readers. Future guidelines for observational studies from journals, requiring the reporting of a DAG and its constructions might be one way to increase uptake of the technique.

We feel our review has several strengths. Firstly, we used a broad search strategy to encompass all aspects of the surgical and perioperative literature, prospectively registered and followed our protocol, and sought to incorporate relevant aspects of best-practice beyond the newly published ESC-DAG framework. We had a strong agreement rate within our group prior to the making of a final consensus decision. This is denoted by a Cohen's kappa statistic of 0.86 [36]. We do acknowledge potential limitations. Given DAGs are normally included in the methods or supplementary section of studies, some studies that did not mention the use of a DAG within their title/abstract/keywords may have been missed. We believe the method of including any abstract that had language pertaining to causal inference or DAGs minimised the risk of missing eligible studies that had constructed a DAG, as we expected authors to draw attention to the use of this tool in the title or abstract. However, we acknowledge that a small number of studies may have still been missed if key terms were not included in the abstract or title. This should not detract from the validity of the critique of those studies that we did identify. We also acknowledge that there was a small sample size of final studies to screen against the ESC-DAG protocol and without a clear denominator as to the number of observational studies conducted in the perioperative literature in this period it is difficult to truly judge the prevalence of DAGs within the field. However, our primary focus during this study was to assess DAGs constructed in the literature against the ESC-DAG protocol, rather than assess their prevalence. A further study could address this question in greater depth. The ESC-DAG framework specifically deals with synthesising published literature into a causal model. However, expert opinion is a crucial component of such a process and studies have started to look at how qualitative techniques can be used to distil this expertise into a DAG

[47]. However, numerous studies cited the literature in their DAG and so even if a mixed methods approach was employed, the use of the 'mapping' phase would still be useful in communicating the findings of previously conducted studies. Finally, we recognise that other frameworks for DAG construction are also being published and thus the field is an emerging one. The DAG With Omitted Objects Displayed (DAGWOOD) [48] framework is another framework that can be used for generating causal models and one that reinforces best DAG practices. We did not assess our final studies against this recently published framework. However, given the quality of reporting DAG construction in our studies; we suspect that papers would still not do well when compared against this or other DAG construction frameworks. More work is required to understand what an appropriate reporting framework is for both authors and readers.

Conclusion

DAGs in the surgical and perioperative observational literature do not currently follow the ESC-DAG protocol or other areas of best practice. This has implications on transparency and reproducibility within the field but may be driven by the recency of publication or a lack of awareness. Further improvements to DAG construction and reporting within the perioperative and surgical literature will aid in reader interpretation.

References

1. Hernán MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. CHANCE [Internet]. 2019 Jan 2;32(1):42–9. Available from: https://doi.org/10.1080/09332480.2019.1579578

2. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. Am J Public Health. 2018 May;108(5):616–9.

3. Hammerton G, Munafò MR. Causal inference with observational data: the need for triangulation of evidence. Psychol Med. 2021 Mar;51(4):563–78.

4. Hernan M, Robins J. Causal Inference: What If. In: Boca Raton: Chapman & Hall/CRC. 2020.

5. Pearl J. Causal Diagrams for Empirical Research. Biometrika [Internet]. 1995 Jul 24;82(4):669–88. Available from: http://www.jstor.org/stable/2337329

6. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2:3.

7. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med [Internet]. 2018;210:2–21. Available from: https://www.sciencedirect.com/science/article/pii/S0277953617307359

8. Krishnamoorthy V, Wong DJN, Wilson M, Raghunathan K, Ohnuma T, McLean D, et al. Causal inference in perioperative medicine observational research: part 1, a graphical introduction. Br J Anaesth [Internet]. 2020 Sep 1;125(3):393–7. Available from: https://doi.org/10.1016/j.bja.2020.03.031

9. Gongola A, Bradshaw JC. Directed Acyclic Graphs in Surgical Research. Journal of Surgical Research [Internet]. 2023 Feb 1;282:285–8. Available from: http://10.0.3.248/j.jss.2022.07.017

10. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int J Epidemiol [Internet]. 2021 Apr 1;50(2):620–32. Available from: https://doi.org/10.1093/ije/dyaa213

11. Ferguson KD, McCann M, Katikireddi SV, Thomson H, Green MJ, Smith DJ, et al. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. Int J Epidemiol [Internet]. 2020;49(1):322–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med17&NEWS=N&AN=31325312

12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ [Internet]. 2021 Mar 29;372:n71. Available from: http://www.bmj.com/content/372/bmj.n71.abstract

13. Ferrando C, Librero J, Tusman G, Serpa-Neto A, Villar J, Belda FJ, et al. Intraoperative open lung condition and postoperative pulmonary complications. A secondary analysis of iPROVE and iPROVE-O2 trials. Acta Anaesthesiol Scand [Internet]. 2022;66(1):30-39. Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-02326144/full

14. Qian X, Liu C, Hong S, Xu J, Qian C, Zhu J, et al. Application of Suctioning Ureteral Access Sheath during Flexible Ureteroscopy for Renal Stones Decreases the Risk of

Postoperative Systemic Inflammatory Response Syndrome. Int J Clin Pract [Internet]. 2022;2022:9354714. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=35685551

15. Kalkan S, Cagan Efe S, Karagoz A, Zeren G, Yilmaz MF, Simsek B, et al. A New Predictor of Mortality in ST-Elevation Myocardial Infarction: The Uric Acid Albumin Ratio. Angiology [Internet]. 2022;73(5):461–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med20&NEWS=N&AN=34989646

16. Nimmaanrat S, Geater A, Plunsangkate P, Saewong L, Karnjanawanichkul O, Chanchayanon T, et al. ABO blood group is not a predictive factor for the amount of early opioid consumption in postanesthesia care unit: a prospective cohort study in 3,316 patients. BMC Anesthesiol [Internet]. 2022;22(1):48. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med20&NEWS=N&AN=35172725

17. Steele KM, Schwartz MH. Causal Effects of Motor Control on Gait Kinematics After Orthopedic Surgery in Cerebral Palsy: A Machine-Learning Approach. Front Hum Neurosci. 2022;16:846205.

18. Denzin NK, Lincoln YS. The SAGE handbook of qualitative research, 5th edition. 2017.

19. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas [Internet]. 1960 Apr 1;20(1):37–46. Available from: https://doi.org/10.1177/001316446002000104

20. Lo CKL, Mertz D, Loeb M. Newcastle-Ottawa Scale: comparing reviewers' to authors' assessments. BMC Med Res Methodol [Internet]. 2014;14(1):45. Available from: https://doi.org/10.1186/1471-2288-14-45

21. Scanagatta M, Salmerón A, Stella F. A survey on Bayesian network structure learning from data. Progress in Artificial Intelligence [Internet]. 2019;8(4):425–39. Available from: https://doi.org/10.1007/s13748-019-00194-y

22. Cagigas Fernández C, Palazuelos C, Cristobal Poch L, Gomez Ruiz M. A probabilistic model for the prediction of intra-abdominal infection after colorectal surgery. Int J Colorectal Dis. 2021 Nov;36(11):2481–8.

23. Lam S, Kumar B, Loke YK, Orme SE, Dhatariya K. Glycated haemoglobin and the risk of postoperative complications in people without diabetes: a prospective population-based study in UK Biobank. Anaesthesia [Internet]. 2022 Jun 1;77(6):659–67. Available from: https://doi.org/10.1111/anae.15684

24. Bedir A, Abera SF, Efremov L, Hassan L, Vordermark D, Medenwald Daniel; ORCID: http://orcid.org/0000-0001-8920-2853 DAOM. Socioeconomic disparities in head and neck cancer survival in Germany: a causal mediation analysis using population-based cancer registry data. J Cancer Res Clin Oncol [Internet]. 2021;147(5):1325–34. Available from: http://link.springer.de/link/service/journals/00432/index.htm

25. Laitinen MK, Kilk K, Kask G, Ehne J, Wedin R, Stevenson JD, et al. Surgical treatment of skeletal metastases in proximal tibia: a multicenter case series of 74 patients. Acta Orthop [Internet]. 2021;92(3):352–7. Available from: http://www.tandfonline.com/loi/iort20

26. Tetta C, Rocca M, Salone M, Longhi A, Ferrari C, Giugliano A, et al. Predictors of lung recurrence and disease-specific mortality after pulmonary metastasectomy for soft tissue sarcoma. Surg Oncol [Internet]. 2021;37:101532. Available from: http://www.elsevier.com/locate/suronc

27. Wang Y, Su J, Bai Z, Cai Q, Ouyang Y, Cao X. Role of locoregional surgery in treating FIGO 2009 stage IVB cervical cancer patients: A population-based study. BMJ Open [Internet]. 2021;11(8):e042364. Available from: http://bmjopen.bmj.com/content/early/by/section

28. Duprey MS, Devlin JW, Briesacher BA, Travison TG, Inouye SK, Griffith JL. Approaches to Optimize Medication Data Analysis in Clinical Cohort Studies. J Am Geriatr Soc [Internet]. 2020;68(12):2921–6. Available from: http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1532-5415

29. Pollmann CT, Dahl FA, Røtterud JHM, Gjertsen JE, Årøen A. Surgical site infection after hip fracture - mortality and risk factors: an observational cohort study of 1,709 patients. Acta Orthop. 2020 Jun;91(3):347–52.

30. Sittivarakul W, Prapakornkovit V, Jirarattanasopa P, Bhurayanontachai P, Ratanasukon M. Surgical outcomes and prognostic factors following vitrectomy in acquired immune deficiency syndrome patients with cytomegalovirus retinitis-related retinal detachment. Medicine [Internet]. 2020;99(43):e22889. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emexa&NEWS=N&AN=633308647

31. Hoorntje A, Koenraadt KLM, van Geenen RCI, Kerkhoffs GMMJ, Kuijer PPFM, van Ginneken BT, et al. Predictors of Return to Work After High Tibial Osteotomy: The

615 Importance of Being a Breadwinner. Orthop J Sports Med [Internet]. 2019;7(12).
616 Available from: http://ojs.sagepub.com/

617 32. Kerkhoffs GMMJ, Hoorntje A, Koenraadt KLM, van Geenen RCI, Kuijer PPFM, van
618 Ginneken BT, et al. Prognostic Factors for Return to Sport After High Tibial
619 Osteotomy: A Directed Acyclic Graph Approach. Am J Sports Med [Internet].
620 2019;47(8):1854–62. Available from:
621 http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed20&NEWS=N
622 &AN=628071447

623 33. Pathak IS, Jurak J, Mulla ZD, Kupesic Plavsic S. Predictors of Oophorectomy in Girls
624 Hospitalized in Texas With Ovarian Torsion. Hosp Pediatr [Internet]. 2018;8(5):274–
625 9. Available from:
626 http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med15&NEWS=N
627 &AN=29636365

628 34. Asgari MM, Warton EM, Neugebauer R, Chren MM. Predictors of patient satisfaction
629 with Mohs surgery: analysis of preoperative, intraoperative, and postoperative factors
630 in a prospective cohort. Arch Dermatol. 2011 Dec;147(12):1387–94.

631 35. Sehrndt A, Stang A, Lehmkuhl E, Regitz-Zagrosek V, Babitsch B. Untersuchung des
632 Einflusses des sozioökonomischen Status auf die gesundheitsbezogene Lebensqualität
633 bei Patienten/-innen vor und nach aortocoronarer Bypassoperation – ein
634 Anwendungsbeispiel für kausale Diagramme (DAGs) TT - Investigation of the
635 Influence. Gesundheitswesen. 2011;73(12):903–5.

636 36. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb).
637 2012;22(3):276–82.

37. Tennant PWG, Textor J, Gilthorpe MS, Ellison GTH. OP87 Dagitty and directed acyclic graphs in observational research: a critical review. J Epidemiol Community Health (1978) [Internet]. 2017 Sep 1;71(Suppl 1):A43 LP-A43. Available from: http://jech.bmj.com/content/71/Suppl_1/A43.2.abstract

38. Bellini V, Valente M, Bertorelli G, Pifferi B, Craca M, Mordonini M, et al. Machine learning in perioperative medicine: a systematic review. Journal of Anesthesia, Analgesia and Critical Care [Internet]. 2022;2(1):2. Available from: https://doi.org/10.1186/s44158-022-00033-y

39. Ackland GL, Stephens RCM. Big Data: A Cheerleader for Translational Perioperative Medicine. Vol. 122, Anesthesia and analgesia. United States; 2016. p. 1744–7.

40. Müller-Wirtz LM, Volk T. Big Data in Studying Acute Pain and Regional Anesthesia. J Clin Med. 2021 Apr;10(7).

41. Shapiro DD, Msaouel P. Causal Diagram Techniques for Urologic Oncology Research. Clin Genitourin Cancer [Internet]. 2021;19(3):271.e1-271.e7. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=32893127

42. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr;183(8):758–64.

43. NEJM Submission Guidelines [Internet]. Available from: https://www.nejm.org/author-center/new-manuscripts

44. Guide for Authors; British Journal of Anaesthesia [Internet]. Available from: https://www.elsevier.com/journals/british-journal-of-anaesthesia/0007-0912/guide-for-authors

45. Instructions to Authors; British Journal of Surgery [Internet]. Available from: https://academic.oup.com/bjs/pages/general-instructions

46. Cuschieri S. The STROBE guidelines. Saudi J Anaesth. 2019 Apr;13(Suppl 1):S31–4.

47. Rodrigues D, Kreif N, Lawrence-Jones A, Barahona M, Mayer E. Reflection on modern methods: constructing directed acyclic graphs (DAGs) with domain experts for health services research. Int J Epidemiol [Internet]. 2022 Aug 1;51(4):1339–48. Available from: https://doi.org/10.1093/ije/dyac135

48. Haber NA, Wood ME, Wieten S, Breskin A. DAG With Omitted Objects Displayed (DAGWOOD): a framework for revealing causal assumptions in DAGs. Ann Epidemiol [Internet]. 2022;68:64–71. Available from: https://www.sciencedirect.com/science/article/pii/S1047279722000035

Financial Disclosure Statement

Dr Daniel J Stubbs is funded by the Wellcome Trust (220542/Z/20/Z), Dr Katharina Kohler is funded by an NIHR Clinical Lectureship, Mr Sebastian H.M. Hickman is funded by UK Research and Innovation (EP/S022961/1) and Ms Kaya Marlen Dreesbeimdiek is funded by the Addenbrooke's Charitable Trust. No funder had a role in the design or conduct of this work.

As this work was funded, in whole or in part, by the Wellcome Trust. A CC BY or equivalent licence is applied to the Author Accepted Manuscript (AAM) arising from this submission, in accordance with the grant's open access conditions. The funders had no role in the design or conduct of the work.

Supporting Information

PRISMA Checklist. Completed PRISMA Checklist for this systematic review.

S1 Appendix. Search strategies used to identify relevant articles.

S2 Appendix. Individual scores for each study when scored against the ESC-DAG protocol [11].

S3 Appendix. Individual scores for each study when scored against the Newcastle-Ottawa Scale [20].

Systematic Search: Directed Acyclic Graphs (DAGs) in the Surgical and Perioperative Literature (30/06/2022)

OVID to search Embase

https://authproxy.bma.org.uk/process/redirects?url=https://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=2Lwk9aE9wYi4qELkyqIj8cTdFMxl3oaEuZTrUJ7aVITwkTJxMRCXn6A4nezPPPAsD

Embase <1974 to 2022 June 29>

1	exp directed acyclic graph/	2882
2	DAG.mp.	5687
3	dags.mp.	832
4	directed acyclic graph.mp.	969
5	directed a-cyclic graph.mp.	1
6	directed acyclic network.mp. or directed acyclic graph/	519
7	directed acyclical graph.mp. or directed acyclic graph/	513
8	exp surgery/	5393883
9	exp perioperative medicine/	467
10	exp anesthesiology/	20051
11	exp anesthesiology software/	61
12	exp anesthesiology diagnostic device/	33249
13	exp causal modeling/	1382
14	1 or 2 or 3 or 4 or 5 or 6 or 7 or 13	10605
15	8 or 9 or 10 or 11 or 12	5432193
16	14 and 15	563

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 5
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 5 – last paragraph of introduction
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 8 – paragraph 2 of methods
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 8 – paragraph 4 of methods
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	S1 Appendix
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 9 – paragraph 4 of methods
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 9 – paragraphs 5, 6 and 7 of methods
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 10 (table 1) and Page 7 (paragraph 9 of methods)
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	S3 Appendix and page 10 (paragraph 9 in methods)
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 9 – paragraph 5 of methods
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 9 – paragraph 2 of methods

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 9 – paragraph 5 of methods
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 10 – Table 1
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 9 – paragraph 5 of methods
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Page 9 – paragraph 5 of methods
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Page 9 – paragraph 5 of methods
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A – not looking at the results of included studies; all of the studies assessed had the information required also
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Page 9 – paragraph 5 of methods
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Page 12 – paragraph 1 of results
Study characteristics	17	Cite each included study and present its characteristics.	Table 2
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Table 2 and final paragraph of methods
Results of	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision	Table 3 +

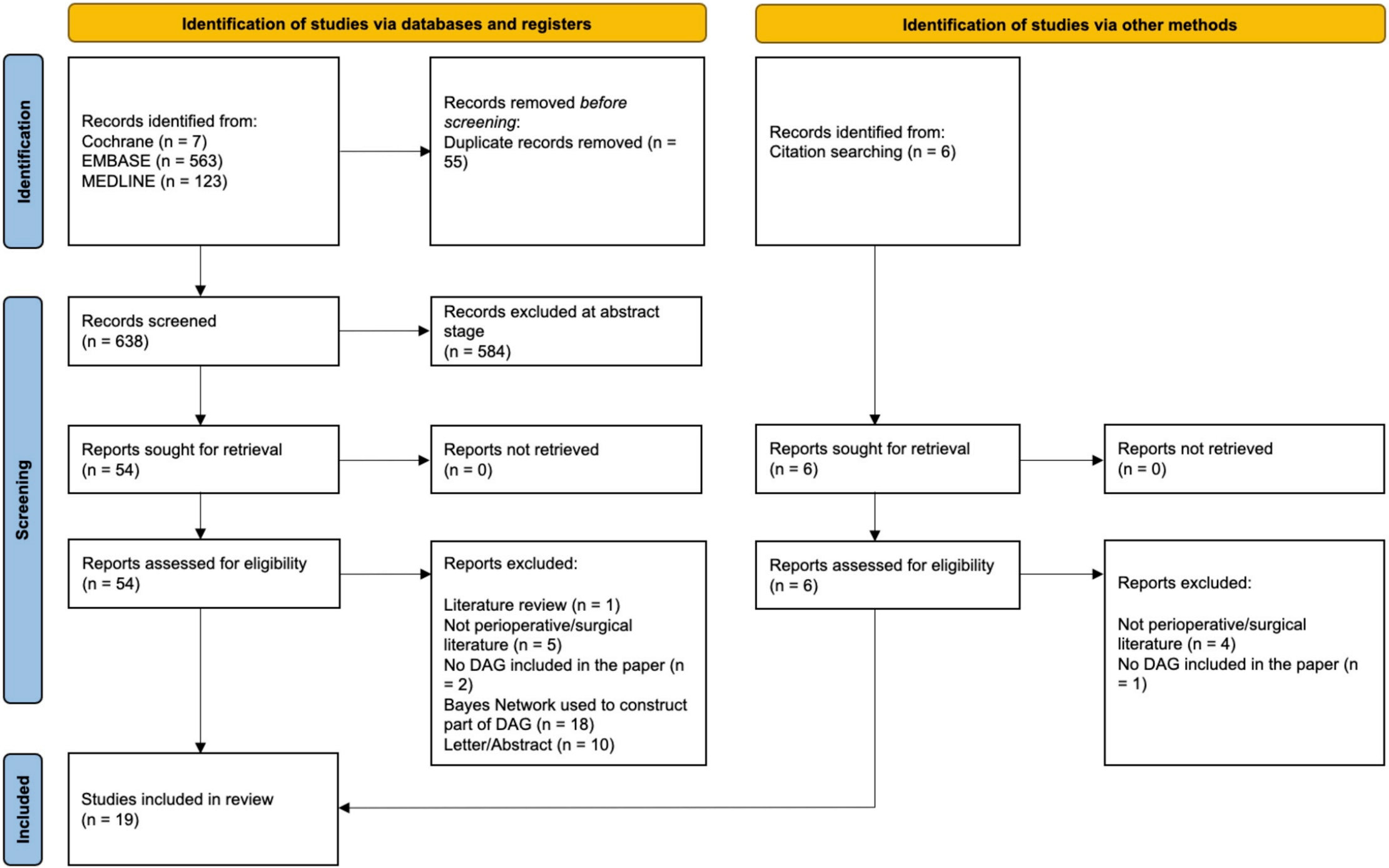
Section and Topic	Item #	Checklist item	Location where item is reported
individual studies		(e.g. confidence/credible interval), ideally using structured tables or plots.	Table 4
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Table 2 + S3 Appendix
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Table 2, 3 + 4
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A – not looking at the results of included studies
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A – not looking at the results of included studies
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Page 15 - paragraph 2 of discussion
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Page 12 – paragraph 5 of results (first paragraph in section B)
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 25 – first paragraph of discussion
	23b	Discuss any limitations of the evidence included in the review.	Page 31 – final paragraph of discussion
	23c	Discuss any limitations of the review processes used.	Page 31 – final paragraph of discussion
	23d	Discuss implications of the results for practice, policy, and future research.	Page 30 – paragraphs 9, 10 and 11 of discussion
OTHER INFORMATION			

PRISMA 2020 Checklist

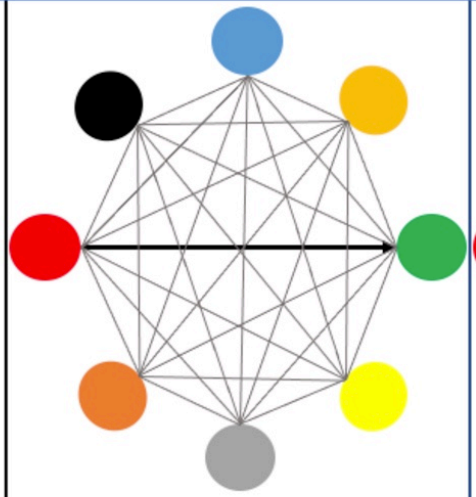
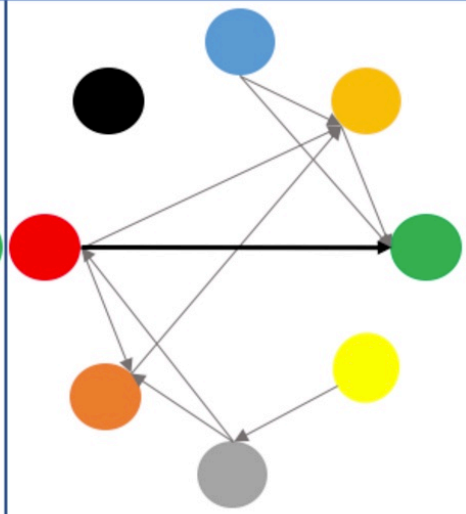
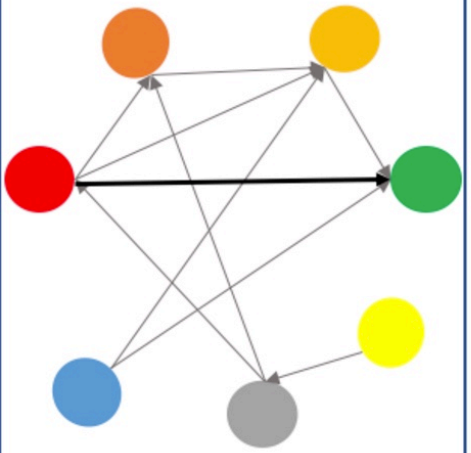
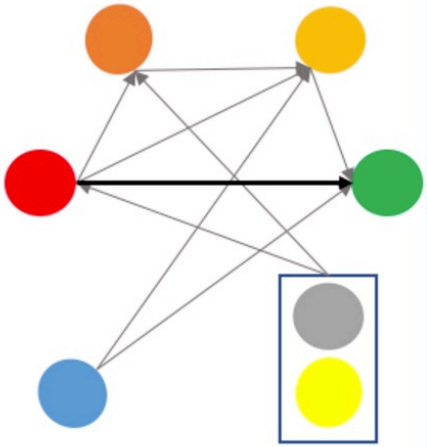
Section and Topic	Item #	Checklist item	Location where item is reported
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 8 – paragraph 1 of methods
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 9 – paragraph 5 of methods (referring to modified inductive reasoning approach taken given there is no strict protocol for this systematic review)
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Nil amendments
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Cover letter
Competing interests	26	Declare any competing interests of review authors.	Cover letter
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	S1, S2 and S3 Appendix

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

		SELECTION				COMPARABILITY	OUTCOME				
Title of Study	**Author of Study**	**Representativeness of the Exposed Cohort? (Item is assessing the representativeness of exposed individuals in the community, not the representativeness of the sample of women from some general population. For example, subjects derived from groups likely to contain middle class, better educated, health oriented women are likely to be representative of postmenopausal estrogen users while they are not representative of all women (e.g. members of a health maintenance**	**Selection of the non-exposed cohort?**	**Ascertainment of exposure?**	**Demonstration that outcome of interest was not present at start of study? (In the case of mortality studies, outcome of interest is still the presence of a disease/ incident, rather than death. That is to say that a statement of no history of disease or incident earns a star.)**	**Comparability of Cohorts on the Basis of the Design or Analysis? (A maximum of 2 stars can be allotted in this category Either exposed and non-exposed individuals must be matched in the design and/or confounders must be adjusted for in the analysis. Statements of no differences between groups or that differences were not statistically significant are not sufficient for establishing comparability. Note: If the relative risk for the exposure of interest is adjusted for the**	**Assessment of outcome? (For some outcomes (e.g. fractured hip), reference to the medical record is sufficient to satisfy the requirement for confirmation of the fracture. This would not be adequate for vertebral fracture outcomes where reference to x-rays would be required).**	**Was Follow-Up Long Enough for Outcomes to Occur? - An acceptable length of time should be decided before quality assessment begins (e.g.5 yrs. for exposure to breast implants).**	**Adequacy of Follow Up of Cohorts?**	**Total Score**	**% given 9 points (stars) available**
Approaches to Optimize Medication Data Analysis in Clinical Cohort Studies	Duprey et al.	1 - Somewhat representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for additional important factors within the study	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	8	89%
A probabilistic model for the prediction of intra-abdominal infection after colorectal surgery	Cagigas et al.	0 - No description of the derivation of the cohort	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	0 - no	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Record linkage (e.g. identified through ICD codes on database records)	0 - no	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	5	56%
Predictors of lung recurrence and disease-specific mortality after pulmonary metastasectomy for soft tissue sarcoma	Tetta et al.	0 - Selected group of users e.g. nurses volunteers	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	0 - no	1 - Complete follow up - all subjects accounted for	6	67%
Predictors of Patient Satisfaction With Mohs Surgery: Analysis of Preoperative, Intraoperative, and Postoperative Factors in a Prospective Cohort	Asgari et al.	0 - Selected group of users e.g. nurses volunteers	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Record linkage (e.g. identified through ICD codes on database records)	1 - yes	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	7	78%
Predictors of Return to Work After High Tibial Osteotomy	Hoomtje et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	8	89%
Role of locoregional surgery in treating FIGO 2009 stage IVB cervical cancer patients: a population-based study	Wang et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Complete follow up - all subjects accounted for	8	89%
Socioeconomic disparities in head and neck cancer survival in Germany: a causal mediation analysis using population-based cancer registry data	Bedir et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	8	89%
Surgical outcomes and prognostic factors following vitrectomy in acquired immune deficiency syndrome patients with cytomegalovirus retinitis-related retinal detachment	Sittivarakul et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Complete follow up - all subjects accounted for	8	89%
Surgical site infection after hip fracture – mortality and risk factors: an observational cohort study of 1,709 patients	Pollmann et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	8	89%
Surgical treatment of skeletal metastases in proximal tibia: a multi-center case series of 74 patients	Laitinen et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	0 - no	0 - Follow up rate poor, over 10% lost to follow up, and no description provided	6	67%
Glycated haemoglobin and the risk of postoperative complications in people without diabetes: a prospective population-based study in UK Biobank	Lam et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	0 - no	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Record linkage (e.g. identified through ICD codes on database records)	1 - yes	1 - Complete follow up - all subjects accounted for	7	78%
Intraoperative open lung condition and postoperative pulmonary complications. A secondary analysis of iPROVE and iPROVE-O2 trials	Ferrando et al.	0 - Selected group of users e.g. nurses volunteers	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for additional important factors within the study	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Complete follow up - all subjects accounted for	7	78%
ABO blood group is not a predictive factor for the amount of early opioid consumption in postanesthesia care unit: a prospective	Nimmaanrat et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for an important factor e.g. age, 1 - study controls for additional important factors	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	0 - Follow up rate poor, over 10% lost to follow up, and no description provided	8	89%
Investigation of the Influence of the Socioeconomic Status on the Health-Related Quality of Life in Patients Before and After Coronary Artery Bypass Grafting - An Example of the Use of Causal Diagrams (German)	Schmdt et al.	1 - Truly representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for additional important factors within the study	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to secure records (x-rays, medical records, etc)	1 - yes	0 - Follow up rate poor, over 10% lost to follow up, and no description provided	7	78%
Application of Suctioning Ureteral Access Sheath during Flexible Ureteroscopy for Renal Stones Decreases the Risk of Postoperative Systemic Inflammatory Response Syndrome	Qian et al.	1 - Somewhat representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for additional important factors within the study	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Complete follow up - all subjects accounted for	8	89%
Causal Effects of Motor Control on Gait Kinematics After Orthopedic Surgery in Cerebral Palsy: A Machine-Learning Approach	Steele et al.	0 - Selected group of users e.g. nurses volunteers	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for additional important factors within the study	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	7	78%
A New Predictor of Mortality in ST-Elevation Myocardial Infarction: The Uric Acid Albumin Ratio	Kalkan et al.	1 - Somewhat representative of the average person in the community	1 - Drawn from the same community as the exposed cohort	1 - secure record e.g. hospital records	1 - yes	1 - study controls for additional important factors within the study	1 - Independent or blind assessment stated in the paper, or confirmation of the outcome by reference to	1 - yes	1 - Subjects lost to follow up unlikely to introduce bias - small number lost (10%) or a description	8	89%

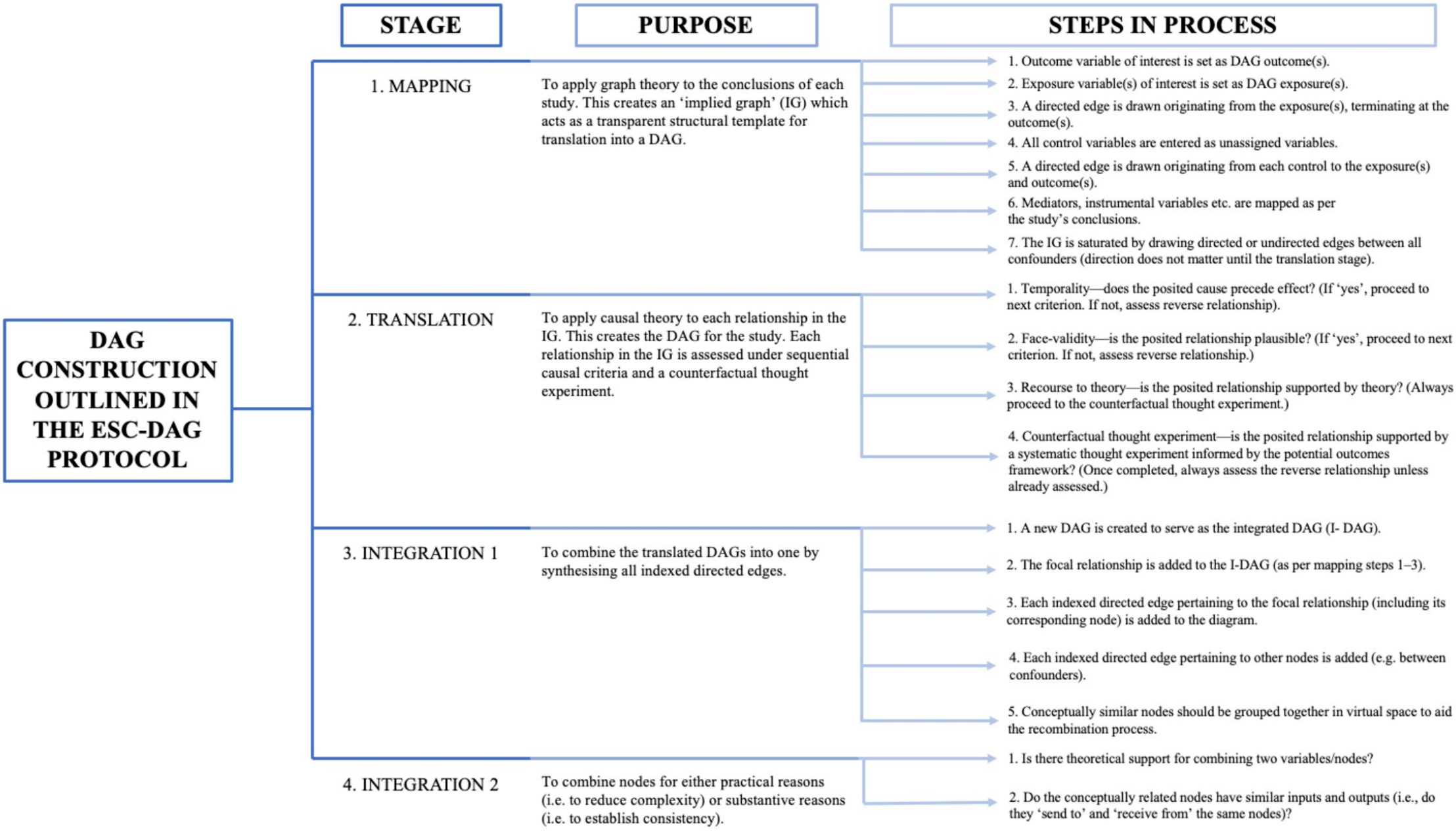

Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
Cochrane (n = 7)
EMBASE (n = 563)
MEDLINE (n = 123)
Records removed *before screening*:
Duplicate records removed (n = 55)
Records identified from:
Citation searching (n = 6)
Records screened
(n = 638)
Records excluded at abstract stage
(n = 584)
Reports sought for retrieval
(n = 54)
Reports not retrieved
(n = 0)
Reports sought for retrieval
(n = 6)
Reports not retrieved
(n = 0)
Reports assessed for eligibility
(n = 54)
Reports excluded:
Literature review (n = 1)
Not perioperative/surgical literature (n = 5)
No DAG included in the paper (n = 2)
Bayes Network used to construct part of DAG (n = 18)
Letter/Abstract (n = 10)
Reports assessed for eligibility
(n = 6)
Reports excluded:
Not perioperative/surgical literature (n = 4)
No DAG included in the paper (n = 1)
Studies included in review
(n = 19)

A hypothetical example to demonstrate the ESC-DAG construction protocol: does intraoperative blood loss cause long-term chronic kidney disease?

Stage of ESC-DAG Protocol	1. Mapping	2. Translation	3. Integration 1	4. Integration 2
Purpose of step from the ESC-DAG Protocol	To apply graph theory to the conclusions of each study. This creates an 'implied graph' (IG) which acts as a transparent structural template for translation into a DAG.	To apply causal theory to each relationship in the IG. This creates the DAG for the study. Each relationship in the IG is assessed under sequential causal criteria and a counterfactual thought experiment.	To combine the translated DAGs into one by synthesising all indexed directed edges.	To combine nodes for either practical reasons (i.e. to reduce complexity) or substantive reasons (i.e. to establish consistency).
Example DAG with captions for each stage	 The exposures, outcomes and variables of studies and expert opinions are mapped to create an implied graph. This forms a template to analyse the causal relationships of all variables included. More than one implied graph may be created at this stage from the literature and expert opinion. In this example, all variables from a hypothetical study are included within the implied graph shown.	 Each implied graph that has been created is interpreted using causal theory. Relevant causal relationships are then retained on the graph e.g. the causal relationship between age and chronic kidney disease is retained amongst others.	 If more than one implied graph has been initially mapped and translated, these graphs are integrated to form one DAG at this stage. Similar variables (nodes) can also be grouped together in space during this stage e.g. ensuring length and type of surgery are grouped together spatially.	 Variables (nodes) can be combined at this stage if they have similar inputs and outputs, or if there is theoretical support for combining two variables on the graph e.g. length and type of surgery.

Key

- Intraoperative blood loss (exposure)
- Chronic kidney disease (outcome)
- Respiratory co-morbidities
- Age
- Postoperative acute kidney injury
- Intraoperative hypotension
- Length of surgery
- Type of surgery
- Causal effect


DAG CONSTRUCTION OUTLINED IN THE ESC-DAG PROTOCOL
STAGE
PURPOSE
STEPS IN PROCESS
1. MAPPING
To apply graph theory to the conclusions of each study. This creates an 'implied graph' (IG) which acts as a transparent structural template for translation into a DAG.
1. Outcome variable of interest is set as DAG outcome(s).
2. Exposure variable(s) of interest is set as DAG exposure(s).
3. A directed edge is drawn originating from the exposure(s), terminating at the outcome(s).
4. All control variables are entered as unassigned variables.
5. A directed edge is drawn originating from each control to the exposure(s) and outcome(s).
6. Mediators, instrumental variables etc. are mapped as per the study's conclusions.
7. The IG is saturated by drawing directed or undirected edges between all confounders (direction does not matter until the translation stage).
2. TRANSLATION
To apply causal theory to each relationship in the IG. This creates the DAG for the study. Each relationship in the IG is assessed under sequential causal criteria and a counterfactual thought experiment.
1. Temporality—does the posited cause precede effect? (If 'yes', proceed to next criterion. If not, assess reverse relationship).
2. Face-validity—is the posited relationship plausible? (If 'yes', proceed to next criterion. If not, assess reverse relationship.)
3. Recourse to theory—is the posited relationship supported by theory? (Always proceed to the counterfactual thought experiment.)
4. Counterfactual thought experiment—is the posited relationship supported by a systematic thought experiment informed by the potential outcomes framework? (Once completed, always assess the reverse relationship unless already assessed.)
3. INTEGRATION 1
To combine the translated DAGs into one by synthesising all indexed directed edges.
1. A new DAG is created to serve as the integrated DAG (I- DAG).
2. The focal relationship is added to the I-DAG (as per mapping steps 1–3).
3. Each indexed directed edge pertaining to the focal relationship (including its corresponding node) is added to the diagram.
4. Each indexed directed edge pertaining to other nodes is added (e.g. between confounders).
5. Conceptually similar nodes should be grouped together in virtual space to aid the recombination process.
4. INTEGRATION 2
To combine nodes for either practical reasons (i.e. to reduce complexity) or substantive reasons (i.e. to establish consistency).
1. Is there theoretical support for combining two variables/nodes?
2. Do the conceptually related nodes have similar inputs and outputs (i.e., do they 'send to' and 'receive from' the same nodes)?

OVID to search Medline

https://authproxy.bma.org.uk/process/redirects?url=https://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=6q47RSU4OfKeiwmUNZDgfxJOknzjNqTJZDvsUwbfNKGkGUILDlhHRwh9MwzkWvdks

Ovid MEDLINE(R) ALL <1946 to June 29, 2022>

1	exp Specialties, Surgical/	214117
2	exp "Anesthesia and Analgesia"/	246945
3	exp Surgical Procedures, Operative/	3438011
4	2 or 3	3572696
5	exp Anesthesiology/	32754
6	exp Anesthesia/	201983
7	5 or 6	217892
8	exp Perioperative Medicine/	104
9	directed acyclic graph.mp.	575
10	DAG.mp.	4861
11	exp Diacylglycerol Kinase/	1289
12	exp Diglycerides/	7631
13	exp Signal Transduction/	659284
14	DAGS.mp.	762
15	directed a-cyclic graph.mp.	1
16	directed acyclic network.mp.	11
17	directed acyclical graph.mp.	2
18	causal diagram.mp.	77
19	structural causal model.mp.	21
20	exp Protein Kinase C/	45908
21	11 or 12 or 13 or 20	696092
22	(10 or 14) not 21	2663
23	9 or 15 or 16 or 17 or 18 or 19 or 22	3130
24	1 or 4 or 7 or 83729736	
25	23 and 24	123

COCHRANE

#1	MeSH descriptor: [Specialties, Surgical] explode all trees			MeSH	2036
#2	MeSH descriptor: [Anesthesia and Analgesia] explode all trees			MeSH	29018
#3	MeSH descriptor: [Surgical Procedures, Operative] explode all trees			MeSH	128873
#4	MeSH descriptor: [Anesthesiology] explode all trees			MeSH	437
#5	MeSH descriptor: [Anesthesia] explode all trees			MeSH	20608
#6	MeSH descriptor: [Perioperative Medicine] explode all trees			MeSH	0
#7	(directed acyclic graph):ti,ab,kw	S	Limits		19
#8	(DAG):ti,ab,kw	S	Limits		184
#9	(DAGs):ti,ab,kw	S	Limits		10
#10	(directed a-cyclic graph):ti,ab,kw	S	Limits		0
#11	(directed acyclic network):ti,ab,kw	S	Limits		7
#12	(directed acyclical graph):ti,ab,kw	S	Limits		1
#13	(causal diagram):ti,ab,kw	S	Limits		21
#14	(structural causal model):ti,ab,kw	S	Limits		106
#15	#1 or #2 or #3 or #4 or #5 or #6		Limits		139993
#16	#7 or #8 or #9 or #10 or #11 or #12 or #13 or #14		Limits		329
#17	#15 and #16		Limits		7