National Institute for Health and Care Excellence

Final

Abdominal aortic aneurysm: diagnosis and management

Evidence review K2: Observational evidence on the effectiveness of endovascular aneurysm repair compared with open surgical repair of unruptured abdominal aortic aneurysms

NICE guideline NG156

Methods, evidence and recommendations

March 2020

Final

This evidence review was developed by the NICE Guideline Updates Team

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-3452-2

Contents

1 Introduction

The aim of this review is to supplement Evidence review K, which assesses the comparative safety and effectiveness of endovascular aneurysm repair (EVAR) and open surgical repair (OSR) for the treatment of unruptured abdominal aortic aneurysms (AAAs).

While Evidence review K was confined to randomised evidence addressing this question, this additional review identifies, appraises and synthesises evidence from published observational studies that attempt to estimate the relative benefits and harms of EVAR and OSR while accounting for imbalances in potentially confounding factors. The particular focus of the review is to investigate the proposition that, as the RCTs recruited participants up to 20 years ago, the balance of benefits and harms has become more favourable to EVAR in more recent practice.

Details of the committee's consideration of this evidence in conjunction with the original review of randomised trials can be found in Evidence review K.

2 Methods

2.1 General

This evidence review was developed using the methods and process described in [Developing NICE guidelines: the manual (2014)] and the general methods chapter for this guideline.

2.2 Identifying the evidence

A focused search strategy was used across multiple databases to identify relevant literature; see Appendix A for details. We reviewed the references of included articles and related reviews identified in the search to find any publications that the searches had missed.

2.3 Eligibility criteria

2.3.1 Study design

Recommended adjustment methods

As observational data are always subject to selection biases, we only included studies that made an attempt to account for differences in casemix between EVAR and OSR cohorts. We considered comparative studies that used any of the methods of adjustment enumerated in NICE DSU Technical Support Document 17 (Faria et al., 2015), including:

- Regression adjustment
- Inverse probability weighting
- Regression on the propensity score
- Matching (nearest neighbour or propensity score)
- Instrumental variable methods
- Difference-in-differences designs
- Regression discontinuity design

In practice, only 3 of these designs were represented in the assembled evidence: regression on propensity score, matching by propensity score and inverse probability weighting. Each of these methods relies on the calculation of a propensity score, estimating the odds of an individual receiving 1 of the 2 treatments, given their characteristics. In regression analyses, the propensity score is used a single measure of each participant's characteristics, in an attempt to isolate the independent effect of the treatment they received when isolated from the effects of the things that led to them being chosen for that approach. In matching by propensity score, each participant who received 1 of the treatments is matched with 1 or more similar participants who received the other, in an attempt to create the kinds of balanced cohorts that would be expected if treatment assignment had been randomised. In inverse probability weighting, the propensity score is used to assign weights to each of the participants in a study, again with the aim of creating 2 cohorts that have the same average characteristics.

In addition to the recommended methods of adjusting for selection bias, simple multivariable regression analyses are commonly reported in this area. Mostly, these take the form of logistic regression models (for example, for perioperative mortality) or Cox proportional hazards models (for example, for long-term survival). Such analyses attempt to isolate the independent effect of treatment when controlling for other covariates of outcome. These approaches are generally considered insufficient to identify treatment effects in the presence of selection bias, because differences in covariate distributions between treatment and control groups may compromise the linearity on which logistic regression and Cox proportional hazards models rely (Little and Rubin, 2000; Newgard et al., 2004). However, we identified studies that use these techniques and, in stratified analyses, explored whether similar effects are estimated as with the recommended approaches.

The protocol for this review stated that studies with recommended adjustment methods only should form our primary analysis, with the inclusion of naive regressions only as a secondary analysis. However, in practice, we found that there were no systematic discrepancies between the 2 types of data so, to avoid unhelpful duplication, we present both types in single analyses throughout (although analyses are always stratified, so that results for different designs can be isolated and compared, if readers prefer).

2.3.2 Other criteria

We limited the evidence-base to studies that commenced recruitment in or after 1999. There were 2 reasons for this: firstly, the RCTs began recruiting in 1999 and, as the point of the review is to examine whether things have changed since the RCTs, it does not make sense to include evidence that predates them; secondly, 1999 is when the first EVAR devices received US FDA approval, so we had some concern that studies predating this timepoint may feature non-approved endografts (including possibly some physician-developed devices) that would not provide valid evidence as to the benefits and harms of more established practice.

Aside from study design and recruitment date, the eligibility criteria for this review were essentially unchanged from evidence review K. However, we refined our definition of long-term survival based on findings from randomised evidence, and also in the light of the original economic modelling that had been undertaken to support that review. In particular, the randomised evidence shows that, while EVAR is associated with reduced perioperative mortality, compared with OSR, long-term survival estimates for people who survive repair are either neutral or in favour of OSR. For this reason, survival data should be analysed using an approach that can account for variable hazards of at least 2 phases. Accordingly, we did not include studies that reported a single effect measure purporting to summarise overall survival, including both the perioperative and post-perioperative periods (for example, a hazard ratio from a single Cox proportional hazards regression). In a few instances, studies reported the outputs of a Cox proportional hazards model for long-term survival conditional on surviving the perioperative period, and we included those estimates.

We also omitted ‘successful exclusion of the aneurysm, aneurysm rupture, or further aneurysm growth’ as a specific outcome of interest, as we considered these factors are well captured by reintervention data.

Table 1: Eligibility criteria ('PICO' table)

Population	People undergoing surgery for a confirmed unruptured AAA Subgroups: fitness for surgery, age, sex, comorbidities (including cardiovascular disease, renal disease, COPD, obesity), ethnicity

Interventions	• Elective standard (on-instructions for use [IFU]) EVAR for infrarenal and juxtarenal AAA
	• Elective complex EVAR for infrarenal, juxtarenal and suprarenal AAA, including:
	  ○ fenestrated EVAR
	  ○ EVAR with chimneys
	  ○ EVAR with snorkels
	  ○ branched grafts
	  ○ 'CHIMPS' (CHIMneys, Periscopes, Snorkels)
	  ○ infrarenal devices used for juxtarenal AAA – that is, off-IFU use of standard devices
	• Open repair
	• Non-surgical management

Comparators	Each other

Outcomes	• Mortality/survival, including:
	  ○ Perioperative / 30-day / in-hospital mortality related to the index procedure
	  ○ Post-perioperative long-term survival
	• Peri- and post-operative complications
	• Need for reintervention
	• Quality of life
	• Resource use, including length of hospital or intensive care stay, and costs

Studies were excluded if they:

- were not in English
- were not full reports of the study (for example, published only as an abstract)
- were not peer-reviewed.

2.4 Critical appraisal

When compared with RCTs, all observational designs are at heightened risk of estimating biased treatment effects; therefore, it is particularly important to assess each study's susceptibility to different biases and critically appraise the steps taken by investigators to minimise their effects.

In choosing an instrument with which to do this, we reviewed the generic provisions of the Cochrane Collaboration's 'Risk Of Bias In Non-randomized Studies of Interventions' (ROBINS-I) tool (Sterne et al., 2016). Additionally, because it was necessary to undertake detailed appraisal of the statistical methods used to account for casemix in the included studies, we also considered the more technically focused criteria in the 'Quality of Effectiveness Estimates from Non-Randomised Studies' (QuEENS) checklist (Faria et al., 2015).

Neither instrument covered all relevant issues, and both contained multiple criteria that would never be helpful in discriminating the reliability of studies included in this particular review. For example, the ROBINS-I tool asks 'Is there potential for confounding of the effect of intervention in this study?' to which the answer will always be 'yes', and the QuEENS checklist asks 'Are potential instrumental variables excluded from the set of conditioning variables?' to which we would always have answered 'no', given that it is not clear that any such variables exist, in this case.

Therefore, we developed a bespoke instrument that incorporated important elements of the 2 published checklists, and also added consideration of areas that are not explicitly covered by either. Because the instrument was not designed to be reusable in other contexts, we also took the opportunity to specify criteria that are explicitly focused on our review question (for example, we ask ‘Does the study control appropriately for AAA characteristics?’, rather than ROBINS-I’s ‘Did the authors use an appropriate analysis method that controlled for all the important confounding domains?’ and so on). The instrument is shown in Appendix D; the table also highlights the domain(s) of the ROBINS-I and QuEENS checklists that each criterion reflects.

2.5 Outcomes

We adopted a single outcome for perioperative mortality, comprising data reported as 30-day mortality or in-hospital deaths. In the few cases where studies reported both outcomes, we extracted whichever was the higher.

Long-term survival corresponds to the definition of post-perioperative survival used in the original economic model – that is, survival conditional on surviving 30 postoperative days. A small number of studies report this outcome from their own analyses; however, in most cases, it was necessary for us to calculate the relevant effect ourselves. This was possible for any study that published Kaplan–Meier time-to-event curves for casemix-adjusted cohorts. We used digitising software (Engauge Digitizer v10.10) to extract data from the graphs, and then used the method of Guyot et al. (2012) to reconstruct approximate patient-level data and estimate hazard ratios (HRs) for the dataset. We checked the accuracy of this method by (a) comparing our results across the length of the survival function with HRs published by authors – in each case, we found that our estimated HR and its 95% confidence interval matched the published values extremely closely (in most cases, it was correct to within 2 decimal places); and (b) overlaying Kaplan–Meier curves from the reconstructed data on the published graphs – again, there was excellent agreement between the 2. Having reconstructed patient-level data, we removed cases that died or were lost to follow-up in the first 30 days and estimated post-perioperative HRs from the remaining data. We reviewed ‘log–log’ plots of the post-perioperative cumulative hazard functions, to assess the appropriateness of summarising treatment effect using a single HR (that is, assuming proportional hazards). We noted that, in most cases, the lines were broadly parallel, suggesting limited evidence of non-proportional hazards. Where any anomalies were present, they tended to be in the early part of the functions, suggesting that the excess mortality associated with OSR may not, in some datasets, be fully realised by the 31st postoperative day. However, any deviations resolved quickly as follow-up time extended, so it appeared reasonable to assume approximately proportional hazards, and we calculated HRs using Cox models with a single explanatory variable for treatment assignment.

2.6 Evidence categorisation and synthesis

General methods for evidence synthesis are as set out in the methods chapter for this guideline.

As we had for Evidence review K, we wanted to subdivide the decision problem according to AAA anatomy – infrarenal and complex (with the latter representing juxta-, para- and supra-renal AAAs, as well as type IV thoracoabdominal aneurysms). However, in contrast to the RCTs, which all recruited people with infrarenal AAAs only, several studies in the observational dataset do not have clear eligibility criteria and many explicitly include all AAAs regardless of anatomy. Therefore, we have subdivided our analyses as follows:

- ‘Exclusively or predominantly infrarenal AAAs’, in which we present stratified analyses, comprising

  - ‘Infrarenal AAAs’ – studies that are clearly limited to infrarenal anatomy
  - ‘All AAAs’ – studies that do not distinguish AAA anatomy (including some that label their results as ‘infrarenal’ but do not take adequate steps to limit their dataset to such cases)

  We consider that these categories are likely to provide broadly comparable results, as any studies that do not distinguish between AAA anatomy are likely to include a preponderance of infrarenal cases. However, because this is a potential source of bias, we present results in stratified analyses, so estimated effects in cases we can confidently call infrarenal can be seen and compared with the less well defined group.

- ‘Complex AAAs’. We were able to subdivide this group further into
  - Studies comparing fenestrated EVAR grafts only with OSR performed in an analogous population (although how investigators identified the latter group is a possible source of bias, as reflected in our critical appraisal).
  - Studies that identify a population with complex AAAs and compare all relevant endovascular approaches with OSR.

We noted that several datasources form the basis of more than 1 study. Examples include the US Medicare database, the American College of Surgeons’ National Surgical Quality Improvement Program and the National Inpatient Sample. To prevent double-counting of participants, we only entered 1 study per datasource into any given synthesis (except in the case of studies with non-overlapping recruitment periods). Where we had multiple studies to choose from, we selected the study to include according to the following hierarchy:

- Where naïve multivariable regressions are included alongside more robust methods of adjustment, we prefer any study that uses recommended methods.
- We prefer studies that distinguish clearly between AAA anatomy – that is, we prefer those we can categorise as ‘infrarenal’ over those in the ‘all AAA’ category that are likely to comprise a preponderance of infrarenal AAAs, but provide no detail about the types of AAAs reflected in their data.
- We aim to accrue the largest sample size possible. In most cases, this will be achieved by selecting the study with the largest number of participants. However, it might be better to include 2 smaller studies with non-overlapping recruitment periods if, between them, they represent a larger sample than is available in any other single study.

Wherever we have excluded datapoints from meta-analyses on this basis, we have nonetheless shown the data in the relevant forest plot, to maximise transparency and to facilitate comparison with included data; however, the excluded datapoints do not contribute to pooled totals.

The potential for double-counting between – as well as within – datasources remains. For example, it is possible that cases in single-centre studies also had data submitted to regional and/or national surgical registries, and the same people may also appear in studies based on routine administrative data. It is not possible to ascertain the extent of any residual duplication from the aggregate-level data available to us.

2.7 Meta-analysis and meta-regression

We present results in stratified meta-analyses that distinguish both between extent of AAA and method of adjustment (recommended techniques *versus* naive multivariable regression).

We chose fixed-effect and random-effects models on a priori grounds, rather than on inspection of statistical heterogeneity. For RCTs, we follow the decision-rule adopted by the Cochrane review of randomised evidence, in this area (Paravastu et al., 2014) – that is, where there is evidence of statistical heterogeneity (I^2>50%), a random-effects model is used; otherwise, a fixed-effect analysis is preferred. However, our a priori expectation of observational data is that – owing to varied selection biases, varied approaches to casemix adjustment and varied outcome definitions – they are likely to be heterogeneous. Moreover, whereas it is natural to place proportional weight on larger randomised studies, this is not a desirable property of syntheses of observational data, as the size of a study has no bearing on its ability to estimate accurate effects – there is every danger that, in the fixed-effect paradigm, a single, large study could swamp a meta-analysis containing smaller studies that may be less biased. For these reasons, we used random-effects (der Simonian and Laird) models for all syntheses of observational data. We test for differences between study designs with *z*-tests.

To test the hypothesis that between-treatment effects have changed over time, we employ meta-regression with midpoint of recruitment as a continuous covariate of relative effects (mixed-effects model with REML estimator of between-study variance).

3 Results

3.1 Identification of evidence

From an initial database of 2,964 references, we could confidently exclude 2,696 on the basis of title and abstract, leaving 268 articles to be reviewed in full. Of these, we excluded 217 (for a list with reasons, see Appendix C), leaving 51 that meet the eligibility criteria for this review. We identified 3 additional studies by consulting the reference lists of included publications, leading to a final total of 54 included studies.

3.2 Description and critical appraisal of included studies

3.2.1 Exclusively or predominantly infrarenal AAAs

We included 42 studies that explore AAAs that are likely to be infrarenal. Their characteristics are summarised in Table 2. Between them, these studies reflect cohorts recruited from 1999 to 2017. There are 18 studies that report exclusively infrarenal AAAs and 24 that either explicitly include all AAAs regardless of anatomy or adopt methods that would have led to the inclusion of some complex AAAs among a mainly infrarenal cohort. Two of the studies report UK practice; however, the majority (31/42) originate from the USA, with 3 from Canada, 1 from Japan, and the remainder from mainland Europe.

Before adjustment for duplication of participants, these studies comprise a total of 914,062 participants (532,816 undergoing EVAR and 381,246 undergoing OSR). Once we deduplicate, to ensure that there are no studies with overlapping recruitment periods from the same datasource, we are left with a maximum possible sample size of 417,032 participants (238,873 EVAR and 178,159 OSR).

In studies that report unadjusted baseline characteristics, the EVAR cohort is older and less likely to be female than the OSR group. In the relatively few studies that report baseline AAA diameter, cases undergoing OSR tend to have somewhat larger AAAs than those receiving EVAR.

On evaluation according to our bespoke appraisal instrument (see Appendix D), we identified meaningful threats to the internal validity of almost all studies. Among studies adopting recommended methods of casemix adjustment, we considered only 1 study at low risk of bias (unfortunately, this study only provided data for 1 safety outcome – incidence of acute kidney injury); 2 were judged at moderate risk of bias and the remaining 11 had a high risk of bias. Common issues include a failure to account for AAA anatomy among adjustment variables, limited consideration of missing data, and a failure to examine the overlap (or ‘common support’) of propensity-matched cohorts, a step that is critical to assess the validity of those methods (see Faria et al., 2015). All-bar-1 of the studies using matched cohorts trimmed 25–50% of the unmatched population in the matching process (for Sugimoto et al., 2017, the figure was 57%). None of these studies reports details of the process by which trimming was accomplished; however, it is evident that – as would be expected – the process resulted in a concentration on the area of overlap between the cohorts. For example, matched EVAR candidates are younger than unmatched ones, so disproportionately many older cases must have been discarded, and matched OSR candidates are older than unmatched ones, so disproportionately many younger cases must have been trimmed.

Similarly, of the 28 included studies relying on naïve multivariable regression, all-bar-2 are at high risk of bias. The quality of regression analyses is generally poor: only 2 studies take any account of AAA anatomy (Locham et al., 2017, and Locham et al., 2018b), and only 2 explore interactions between treatment assignment and any other covariates (Locham et al., 2018b, and Wald et al., 2006). The majority of studies attempt to consider more covariates than is plausible, given the number of events available in their datasets; an extreme example is Gupta et al. (2012), which reports a multivariable logistic regression based on 5 deaths with 4 covariates (that had been selected from a larger pool of unreported size by stepwise elimination).

We had fewer concerns about the external validity of the evidence: 5 studies were judged partially applicable to our whole target population, as they focus on restricted cohorts defined by age and/or comorbidity.

Table 2: Characteristics of included studies – exclusively or predominantly infrarenal AAAs

Study	Recruitment period	Setting(s); datasource(s)	Baseline characteristics					Risk of bias[a]							Applic-ability
			Cohort	N	Mean age	Sex (% male)	Mean diameter	Selection	Confounding	Data collection	Analysis	Matching	Regression	OVERALL	
Infrarenal AAA															
Inverse probability weighting															
Liang et al. (2018)	2003–2014	USA; national surgical registry (VQI)	EVAR[b] OSR[b]	1,928 713	62.0[c] 61.0[c]	88.0% 85.3%	54.0mm[c] 55.0mm[c]	L	L	M	H	M	N	**H**	Partial[d]
Propensity-score matching															
Huang et al. (2015)	2000–2011	USA; single centre (Mayo)	EVAR OSR	558 558	74.0 72.0	85.8%	57.0mm 59.0mm	L	L	M	M	H	N	**H**	Direct
Sugimoto et al. (2017)	2007–2014	Japan; single centre (Nagoya)	EVAR OSR	157 157	75.0[c] 74.0[c]	86.6% 86.0%	53.4mm 53.4mm	M	M	M	M	H	N	**H**	Direct
Zabrocki et al. (2018)	2007–2011	Germany; single centre (Bremen)	EVAR OSR	91 91	74.0 72.0	91.2% 86.8%	63.0mm 62.0mm	L	L	L	L	M	N	**L**	Direct
Regression on propensity score															
Feringa et al. (2007)	2002-2006	Netherlands Single centre (Rotterdam)	EVAR OSR	49 126	65%>70 52%>70	85.7% 83.3%	NR	L	M	L	M	H	H	H	Direct
Naive multivariable regression															
Behrendt et al. (2017)	2008–2015	Germany; national routine database (DAK-G)	EVAR OSR	3,493 1,457	74.0[c] 71.0[c]	85.4% 82.8%	NR	L	M	H	M	N	H	**H**	Direct
Bush et al. (2006)	2001–2003	USA; national surgical registry (VA NSQIP)	EVAR OSR	717 1,187	72.9 71.8	99.4% 99.2%	NR	M	H	M	M	N	H	**H**	Direct
Bush et al. (2007)	2001–2004	USA; national surgical registry (VA NSQIP)	EVAR OSR	788 1,580	71.6 70.2	99.6% 99.1%	NR	M	M	M	M	N	H	**H**	Partial[e]
Chadi et al. (2012)	2000–2010	Canada; single centre (London, ON)	EVAR OSR	875 1,067	75.0 71.0	87.5% 82.2%	NR	L	M	L	H	N	M	**H**	Direct
Choke et al. (2012)	2000–2010	UK; single centre (Leicester)	EVAR OSR	419 391	72.3[c]	88.9%	NR	M	M	L	M	N	H	**H**	Direct
Elkouri et al. (2004)	1999–2001	USA; single centre (Mayo)	EVAR OSR	94 261	77.0 73.0	90.4% 87.7%	57.0mm 57.0mm	M	M	L	H	N	M	**H**	Direct

Study	Recruitment period	Setting(s); datasource(s)	Baseline characteristics					Risk of bias[a]							Applic-ability
			Cohort	N	Mean age	Sex (% male)	Mean diameter	Selection	Confounding	Data collection	Analysis	Matching	Regression	OVERALL	
Gupta et al. (2012)	2007–2009	USA; national surgical registry (ACS NSQIP)	EVAR OSR	369 282	56.0[c] 56.0[c]	90.8% 80.5%	NR	L	M	L	M	N	H	**H**	Partial[f]
Hicks et al. (2015)	2003–2012	USA; single centre (Johns Hopkins)	EVAR OSR	214 83	74.3 69.2	80.8% 75.9%	NR	L	M	L	M	N	H	**H**	Direct
Hua et al. (2005)	2000–2003	USA; national surgical registry (NSQIP-PS)	EVAR OSR	460 582	74.0 71.2	84.6% 79.6%	NR	L	M	M	H	N	H	**H**	Direct
Locham et al. (2017)	2011–2014	USA; national surgical registry (ACS NSQIP)	EVAR OSR	3,869 360	78.4 76.8	79.4% 68.9%	58.0mm 63.0mm	L	M	L	M	N	H	**H**	Partial[g]
Malas et al. (2014)	2005–2011	USA; national surgical registry (ACS NSQIP)	EVAR OSR	15,807 5,308	74.2 71.1	82.0% 74.4%	NR	L	M	M	M	N	M	**M**	Direct
Nguyen et al. (2013)	2005–2010	USA; national surgical registry (ACS NSQIP)	EVAR OSR	1,256 3,886	77.5 74.4	78.5% 67.4%	NR	L	M	M	H	N	H	**H**	Partial[h]
All AAA															
Inverse probability weighting															
Salata et al. (2019)	2003–2016	Canada; regional routine databases (multiple)	EVAR OSR	4,010 4,010	73.10 72.80	82.5% 81.6%	NR	M	M	M	M	M	N	M	Direct
Propensity-score matching															
Egorova et al. (2011)	2000–2006	USA; national routine database (Medicare)	EVAR OSR	42,320 42,320	76.44 74.76	78.5% 78.5%	NR	M	M	H	H	M	N	**H**	Direct
Giles et al. (2011)	2001–2004	USA; national routine database (Medicare)	EVAR OSR	22,826 22,826	76.3 76.2	80.3% 80.6%	NR	L	M	H	H	M	N	**H**	Direct
Johnson et al. (2006)	2001–2003	USA; national surgical registry (VA NSQIP)	EVAR[b] OSR[b]	670 670	71.6 70.2	99.6% 99.1%	NR	L	M	M	M	H	N	**H**	Direct
Mark et al. (2013)	2000–2006	USA; regional routine database (California Statewide)	EVAR OSR	4,483 4,483	51% ≥75 48% ≥75	85.0% 84.0%	NR	L	H	H	H	M	N	**H**	Direct
Schermerhorn et al. (2015)	2001–2008	USA; national routine database (Medicare)	EVAR OSR	39,966 39,966	75.7 75.5	77.7% 77.6%	NR	L	M	M	M	M	N	**M**	Direct
Symonides et al. (2018)	2011–2016	Poland; national routine database (NHF)	EVAR OSR	2,336 2,336	68.7 68.5	85.3% 84.8%	NR	M	M	M	H	H	N	**H**	Direct

Study	Recruitment period	Setting(s); datasource(s)	Baseline characteristics					Risk of bias[a]							Applic-ability
			Cohort	N	Mean age	Sex (% male)	Mean diameter	Selection	Confounding	Data collection	Analysis	Matching	Regression	OVERALL	
Regression on propensity score															
Jackson et al. (2012)	2003–2007	USA; national routine database (Medicare)	EVAR[b,i] OSR[b,i]	3,264 639	76.4 75.2	79.9% 70.8%	NR	M	M	H	M	M	H	**H**	Direct
Jetty et al. (2010)	2002–2007	Canada; regional routine database (CIHI-DAD Ontario)	EVAR OSR	888 5,573	76.0[c] 72.0[c]	86.2% 80.3%	NR	L	M	H	H	H	H	**H**	Direct
Naive multivariable regression															
Casey et al. (2013)	2003-2008	USA; regional database (California SID)	EVAR OSR	9,356 6,380	75.0 72.1	84.7% 75.6%	NR	M	M	H	H	N	M	**H**	Direct
Chang et al. (2015)	2001-2009	USA; regional routine database (California Statewide)	EVAR OSR	12,239 11,431	75.1 72.3	84.4% 77.5%	NR	L	M	H	H	N	M	**H**	Direct
Davenport et al. (2013)	2005–2009	USA; national surgical registry (ACS NSQIP)	EVAR OSR	8,502 3,967	73.2	80.2%	NR	M	H	M	H	N	H	**H**	Direct
de la Motte et al. (2013)	2007–2010	Denmark; national surgical registry (DVR)	EVAR OSR	520 1,137	74.0 70.5	90.0% 80.0%	NR	M	M	M	H	N	M	**H**	Direct
Eslami et al. (2017)	2005–2011	USA; national surgical registry (ACS NSQIP)	EVAR OSR	14,276 4,641	62.9% >70	80.5%	NR	M	M	H	M	N	H	**H**	Direct
Karthikesalingam et al. (2016)	2005–2010	USA; national routine database (NIS)	EVAR OSR	126,211 69,902	73.0[c]	76.1%	NR	L	M	M	H	N	M	**H**	Direct
		UK; national surgical registry (HES)	EVAR OSR	7,937 13,335	74.0[c]	86.6%	NR	L	M	M	H	N	M	**H**	Direct
Lee et al. (2004)	2001–2001	USA; national routine database (NIS)	EVAR OSR	2,565 4,607	73.4 71.9	84.4% 78.1%	NR	M	M	M	H	N	M	**H**	Direct
Lo et al. (2013)	2003–2011	USA; regional surgical database (VSGNE)	EVAR OSR	2,159 1,867	72.7	77.2%	NR	M	M	M	M	N	H	**H**	Direct
Locham et al. (2018b)	2003–2017	USA; national surgical registry (VQI)	EVAR OSR	26,723 6,359	73.5 69.5	81.4% 73.8%	55.6mm 59.5mm	M	M	L	M	N	M	**M**	Direct
Quintana et al. (2019)	2002–2012	Spain; national routine database (CMBDAH)	EVAR OSR	6,809 9,928	71.40	96.7%	NR	M	M	H	M	N	M	**H**	Direct

Study	Recruitment period	Setting(s); datasource(s)	Baseline characteristics					Risk of bias[a]							Applic-ability
			Cohort	N	Mean age	Sex (% male)	Mean diameter	Selection	Confounding	Data collection	Analysis	Matching	Regression	OVERALL	
Raval et al. (2012)	2005–2008	USA; national surgical registry (ACS NSQIP)	EVAR OSR	5,587 2,349	25.6% ≥80	81.1%	NR	L	M	L	M	N	H	**H**	Direct
Schwarze et al. (2009)	2001–2006	USA; national routine database (NIS)	EVAR OSR	90,925 75,222	48.5% ≥75 35.8% ≥75	NR	NR	M	M	M	H	N	H	**H**	Direct
Tarbunou et al. (2019)	2008–2014	USA Cerner Health Facts	EVAR OSR	1,486 992	69.42	71.7%	NR	M	M	H	M	N	H	**H**	Direct
Wald et al. (2006)	2002–2002	USA; national routine database (NIS)	EVAR OSR	2,651 3,865	73.5 71.6	85.3% 77.0%	NR	L	M	H	M	N	L	**H**	Direct
Williams et al. (2013)	2005–2008	USA; national routine database (NIS)	EVAR OSR	62,728 24,253	73.7 71.0	82.6% 75.6%	NR	L	M	M	H	N	M	**H**	Direct

[a] *for details of appraisal instrument, please see Appendix D*
[b] *before matching / weighting*
[c] *median*
[d] *people aged <65 only*
[e] *high-risk (age ≥60 years and ASA classification 3 or 4)*
[f] *age <60*
[g] *age ≥70*
[h] *patients with pre-existing chronic renal insufficiency*
[i] *baseline characteristics include participants with ruptured AAA (NB results reported in sufficient detail to exclude)*

ACS NSQIP – American College of Surgeons National Surgical Quality Improvement Program; CIHI-DAD – Canadian Institute for Health Information – Discharge Abstract Database; CMBDAH – Conjunto Mínimo Básico de Datos al Alta Hospitalaria; DAK-G – DAK-Gesundheit; DVR – Danish Vascular Registry; H – high risk of bias; HES – Hospital Episode Statistics; L – low risk of bias; M – moderate risk of bias; N – not applicable; NHF – National Health Fund; NIS – National Inpatient Sample; NR – not reported; NSQIP-PS – National Surgical Quality Improvement Program – Private Sector; SID – State Inpatient Database; VA NSQIP – Veterans Affairs National Surgical Quality Improvement Program; VQI – Vascular Quality Initiative; VSGNE – Vascular Study Group of New England

3.2.2 Complex AAAs

We included 12 studies; 6 use a recommended method to adjust for factors that may confound treatment effect and a further 6 studies use naive multivariable regression alone. Their characteristics are summarised in Table 3. Eight focus explicitly on fenestrated or fenestrated/branched EVAR; the other 4 look at all complex AAA. Between them, they reflect cohorts recruited from 2001 to 2016. Four studies report experience from European centres, while 9 include participants from the USA (1 has both). All 8 of the USA-only publications were derived from the same datasource, the American College of Surgeons' National Surgical Quality Improvement Program.

These studies comprise a total of 11,321 participants (3,974 undergoing EVAR and 7,347 undergoing OSR). Once we deduplicate, to ensure that there are no studies with overlapping recruitment periods from the same datasource, we are left with a maximum sample size of 4,363 participants (980 EVAR and 3,383 OSR).

Two of the studies are at moderate risk of bias and the remaining 10 have a high risk of bias. As for infrarenal AAAs, common issues include a failure to account for AAA anatomy among adjustment variables, limited consideration of missing data, and a failure to examine the overlap (or 'common support') of propensity-matched cohorts. In the studies using matched cohorts, 25–50% of the unmatched populations were trimmed in the matching process. None of these studies reports details of the process by which trimming was accomplished.

The regression methods are suboptimal in all 6 cases: each study attempts to include too many covariates, relative to the number of events observed, and no interactions with the treatment effect are considered.

Ten of the 12 studies are directly applicable to our decision problem; 1 study only includes octogenarians and another is only partially applicable because it includes some supradiaphragmatic thoracoabdominal aortic aneurysms, which are beyond the scope of this guideline (with a greater proportion of these among the EVAR cohort).

Table 3: Characteristics of included studies – complex AAAs

Study	Recruitment period	Setting(s); datasource(s)	Baseline characteristics					Risk of bias[a]							Applic-ability
			Cohort	N	Mean age	Sex (% male)	Mean diameter	Selection	Confounding	Data collection	Analysis	Matching	Regression	OVERALL	
Complex AAA															
Inverse probability weighting															
Varkevisser et al. (2018)	2012–2016	USA; national surgical registry (ACS NSQIP)	fEVAR OSR	220 181	75.0[b] 72.0[b]	82.3% 76.2%	56.0mm[b] 58.0mm[b]	M	L	M	M	L	N	**M**	Direct
Propensity-score matching															
Fiorucci et al. (2019)	2006–2015	3 centres: Germany & Italy	fEVAR OSR	41 102	73.0 71.0	95.1% 94.1%	NR NR	M	M	M	H	H	N	**H**	Direct
Michel et al. (2015)	2010–2012	France; registry (WINDOW; f/b EVAR) & national database (PMSI; OSR)	f/b EVAR OSR	268 1,678	71.6 69.2	93.3% 91.7%	NR NR	M	L	H	M	H	N	**H**	Partial[c]
Orr et al. (2017)	2012–2015	USA; national surgical registry (ACS NSQIP)	cEVAR OSR	263 263	42% >75 41% >75	74.0% 76.0%	61.0mm 63.0mm	L	M	M	H	M	N	**H**	Direct
Raux et al. (2014)	2001–2012	2 centres: France (Créteil; EVAR) & USA (Massachusetts; OSR)	fEVAR OSR	42 147	73.0 73.0	88.0% 82.0%	NR NR	M	L	L	M	M	N	**M**	Direct
Tinelli et al. (2018)	2010–2016	2 centres: France (Lille; EVAR) & Italy (Rome; OSR)	fEVAR OSR	102 102	71.8 71.7	95.1% 92.2%	59.8mm 60.6mm	M	L	M	H	M	N	**H**	Direct
Naive multivariable regression															
de Guerre et al. (2019)	2011–2017	USA; national surgical registry (ACS NSQIP)	cEVAR OSR	1,260 1,010	73–75[b,d]	78.6% 69.3%	5.5–5.6mm[b,d] 5.7–6.0mm[b,d]	M	L	M	M	N	H	**H**	Direct
Gupta et al. (2017)	2008–2013	USA; national surgical registry (ACS NSQIP)	fEVAR OSR	535 1,207	75.0[b] 72.0[b]	82.0% 72.0%	NR NR	L	M	M	M	N	H	**H**	Direct
Locham et al. (2018a)	2006–2015	USA; national surgical registry (ACS NSQIP)	fEVAR OSR	242 306	83.0[b] 82.0[b]	81.7% 64.1%	NR NR	L	M	L	M	N	H	**H**	Partial[e]
Locham et al. (2019)	2012–2016	USA; national surgical registry (ACS NSQIP)	cEVAR OSR	326 865	74–75[b,f] 72[b]	76.7% 71.3%	5.8mm[b] 5.9mm[b]	L	M	M	M	N	H	**H**	Direct

Study	Recruitment period	Setting(s); datasource(s)	Baseline characteristics					Risk of bias[a]							Applic-ability
			Cohort	N	Mean age	Sex (% male)	Mean diameter	Selection	Confounding	Data collection	Analysis	Matching	Regression	OVERALL	
Tsilimparis et al. (2013)	2005–2010	USA; national surgical registry (ACS NSQIP)	f/b EVAR OSR	264 1,091	74.0 71.0	82.2% 71.5%	NR NR	M	M	M	H	N	H	**H**	Direct
Ultee et al. (2017)	2011–2013	USA; national surgical registry (ACS NSQIP)	cEVAR OSR	411 395	74.9[b] 72.2[b]	77.6% 66.8%	59.0mm 62.0mm	L	M	M	H	N	H	**H**	Direct

[a] *for details of appraisal instrument, please see Appendix D*
[b] *median*
[c] *includes c6% supradiaphragmatic thoracoabdominal aortic aneurysms (outside scope of review)*
[d] *range of medians for men and women*
[e] *age 80–89*
[f] *range of medians for fEVAR and ChEVAR*

ACS NSQIP – American College of Surgeons National Surgical Quality Improvement Program; cEVAR – complex EVAR; ChEVAR – chimney EVAR; fEVAR – fenestrated EVAR; f/b EVAR – fenestrated/branched EVAR; H – high risk of bias; L – low risk of bias; M – moderate risk of bias; N – not applicable; NR – not reported; PMSI – Programme de médicalisation des systèmes d'information (national hospital discharge database)

3.3 Evidence synthesis – exclusively or predominantly infrarenal

3.3.1 Perioperative mortality

Figure 1 shows a random-effects meta-analysis of casemix-adjusted observational evidence reporting perioperative mortality with EVAR and OSR. It suggests that people undergoing OSR have approximately 3 times higher odds of death than those receiving EVAR.

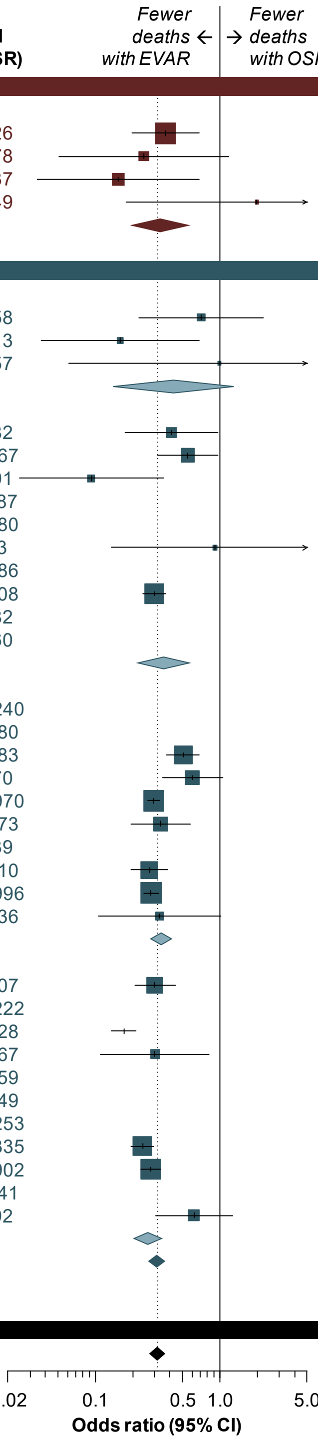
The results of this analysis are conspicuously similar to those found in RCTs: the data are strongly consistent with a null hypothesis of no difference between RCTs and observational studies (p=0.847). If we restrict the dataset to the 8 casemix-adjusted observational studies that explicitly limit their datasets to infrarenal cases, the pooled OR is 0.41 (0.27 to 0.62) – marginally less favourable to RCTs for EVAR.

Figure 2 plots the same data as Figure 1, but also illustrates the period over which each study recruited its participants. It is commonly asserted that the balance of benefits between EVAR and OSR is likely to have shifted in EVAR's favour in the time since the RCTs were undertaken, owing to improvements in devices and/or operator technique. If this were true in the domain of perioperative mortality, we would expect to see a secular trend with relative effects becoming increasingly favourable for EVAR, and the most recent evidence would show distinctly greater benefit than was observed in the RCTs. Neither of these things appears true in Figure 2: there is no appearance of a trend (as confirmed in formal meta-regression, which finds the data to be consistent with a null hypothesis of no change over time; p=0.318), and the pooled estimate from the RCTs appears to be a valid estimate of effect over the entire period covered.

There is no evidence of publication bias in funnel plots (not shown) or on formal hypothesis testing (Egger's p=0.240).

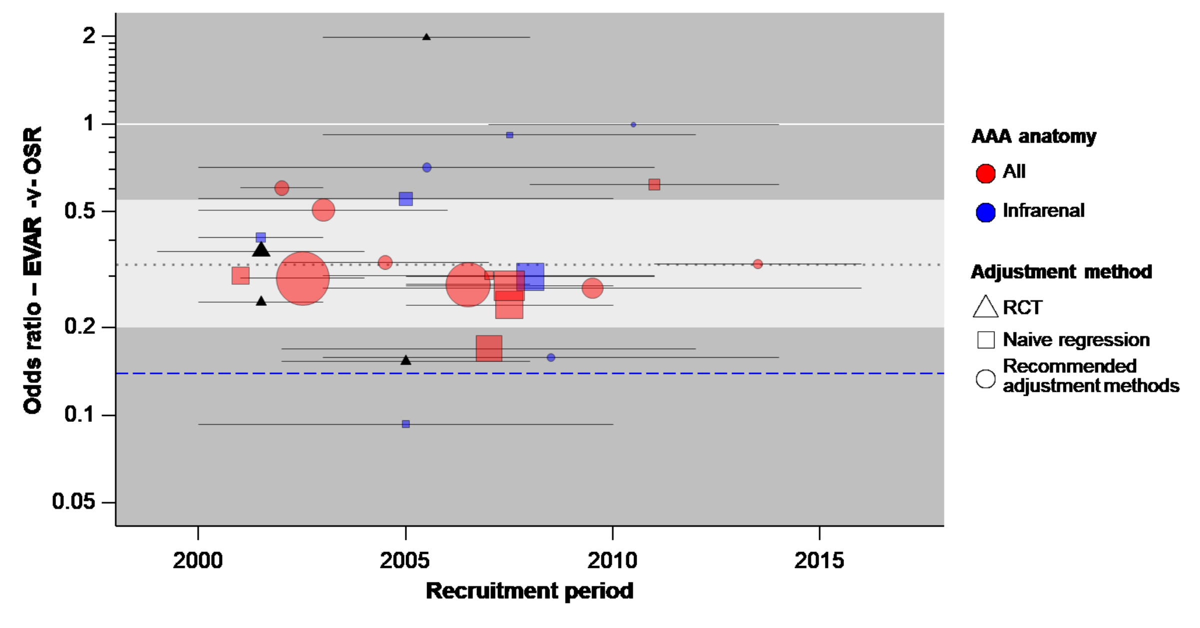
In addition to this between-study evidence, Schermerhorn et al. (2015) provide an analysis of within-study time-trends. This shows that, for people undergoing repair in the period 2001–2008, perioperative mortality declined with both EVAR (p=0.001) and OSR (p=0.013), with no evidence that the year-on-year change was different between the 2 approaches (interaction p=0.129).

Study or subgroup	Recruitment period	N (EVAR)	N (OSR)	Odds ratio (95%CI)	Weight (stratified)	Weight (pooled)
RANDOMISED CONTROLLED TRIALS						
Infrarenal						
EVAR−1	1999–2004	626	626	0.37 (0.20, 0.69)	62.2%	3.3%
DREAM	2000–2003	173	178	0.25 (0.05, 1.17)	13.7%	0.7%
OVER	2002–2008	444	437	0.15 (0.03, 0.68)	22.4%	0.8%
ACE	2003–2008	150	149	1.99 (0.18, 22.2)	1.7%	0.3%
All randomised (FE model) I^2 = 11%, τ^2 = 0.060, p = 0.337				**0.33 (0.20, 0.55)**	**100.0%**	**5.2%**
Test for overall effect: z = −4.21 (p < 0.001)						
CASEMIX−ADJUSTED OBSERVATIONAL DATA						
Infrarenal -- Recommended adjustment methods						
Huang et al. (2015)	2000–2011	558	558	0.71 (0.22, 2.26)	1.4%	1.3%
Liang et al. (2018)	2003–2014	1928	713	0.16 (0.04, 0.68)	0.9%	0.8%
Sugimoto et al. (2017)	2007–2014	157	157	1.00 (0.06, 16.1)	0.3%	0.2%
Subtotal I^2 = 31%, τ^2 = 0.304, p = 0.236				**0.43 (0.14, 1.29)**	**2.5%**	**2.4%**
Infrarenal -- Naive multivariable regression						
Hua et al. (2005)	2000–2003	460	582	0.41 (0.17, 0.97)	2.2%	2.1%
Chadi et al. (2012)	2000–2010	875	1067	0.56 (0.32, 0.96)	4.2%	3.9%
Choke et al. (2012)	2000–2010	419	391	0.09 (0.02, 0.35)	1.1%	1.0%
Bush et al. (2006)	2001–2003	717	1187	0.59 (0.36, 0.98)	[a]	[a]
Bush et al. (2007)	2001–2004	788	1580	0.65 (0.42, 1.02)	[a]	[a]
Hicks et al. (2015)	2003–2012	214	83	0.92 (0.14, 6.27)	0.5%	0.5%
Nguyen et al. (2013)	2005–2010	1256	3886	0.27 (0.19, 0.38)	[b]	[b]
Malas et al. (2014)	2005–2011	15807	5308	0.30 (0.24, 0.37)	8.7%	8.3%
Gupta et al. (2012)	2007–2009	369	282	1.61 (0.15, 17.1)	[b]	[b]
Locham et al. (2017)	2011–2014	3869	360	0.20 (0.12, 0.35)	[b]	[b]
Subtotal I^2 = 55%, τ^2 = 0.137, p = 0.062				**0.36 (0.22, 0.57)**	**16.7%**	**15.8%**
All -- Recommended adjustment methods						
Egorova et al. (2011) [Men]	2000–2006	33240	33240	0.37 (0.34, 0.41)	[c]	[c]
Egorova et al. (2011) [Women]	2000–2006	9080	9080	0.49 (0.42, 0.56)	[c]	[c]
Mark et al. (2013)	2000–2006	4483	4483	0.51 (0.38, 0.69)	7.3%	6.9%
Johnson et al. (2006)	2001–2003	670	670	0.61 (0.35, 1.05)	4.2%	3.9%
Schermerhorn et al. (2015)	2001–2004	22883	22970	0.30 (0.26, 0.33)	10.0%	9.6%
Jetty et al. (2010)	2002–2007	888	5573	0.34 (0.20, 0.58)	4.3%	4.0%
Jackson et al. (2012)	2003–2007	3264	639	0.21 (0.10, 0.45)	[c]	[c]
Salata et al. (2019)	2003–2016	4010	4010	0.27 (0.20, 0.38)	6.8%	6.4%
Schermerhorn et al. (2015)	2005–2008	17083	16996	0.28 (0.25, 0.32)	9.7%	9.3%
Symonides et al. (2018)	2011–2016	2336	2336	0.33 (0.11, 1.03)	1.4%	1.3%
Subtotal I^2 = 68%, τ^2 = 0.034, p = 0.004				**0.34 (0.28, 0.41)**	**43.6%**	**41.4%**
All -- Naive multivariable regression						
Lee et al. (2004)	2001–2001	2565	4607	0.30 (0.21, 0.44)	6.2%	5.8%
Schwarze et al. (2009)	2001–2006	90925	75222	0.23 (0.19, 0.28)	[d]	[d]
Quintana et al. (2019)	2002–2012	6809	9928	0.17 (0.14, 0.21)	8.5%	8.1%
Lo et al. (2013)	2003–2011	2159	1867	0.30 (0.11, 0.83)	1.7%	1.6%
Locham et al. (2018b)	2003–2017	26723	6359	0.26 (0.18, 0.38)	[e]	[e]
Raval et al. (2012)	2005–2008	5587	2349	0.40 (0.28, 0.59)	[b]	[b]
Williams et al. (2013)	2005–2008	62728	24253	0.19 (0.17, 0.21)	[d]	[d]
Karthikesalingam et al. (2016) [UK]	2005–2010	7937	13335	0.24 (0.19, 0.30)	8.7%	8.3%
Karthikesalingam et al. (2016) [USA]	2005–2010	126211	69902	0.28 (0.23, 0.34)	9.1%	8.6%
Eslami et al. (2017)	2005–2011	14276	4641	0.37 (0.29, 0.47)	[b]	[b]
Tarbunou et al. (2019)	2008–2014	1486	992	0.62 (0.30, 1.27)	3.0%	2.8%
Subtotal I^2 = 75%, τ^2 = 0.066, p = 0.001				**0.26 (0.20, 0.34)**	**37.1%**	**35.2%**
All observational (RE model) I^2 = 69%, τ^2 = 0.051, p < 0.001				**0.31 (0.27, 0.36)**	**100.0%**	**94.8%**
Test for overall effect: z = −15.72 (p < 0.001)						
Test for between-substratum differences: χ^2 = 2.89, df = 3 (p = 0.409)						
BOTH DESIGNS						
Pooled (RE model) I^2 = 64%, τ^2 = 0.050, p < 0.001				**0.31 (0.27, 0.36)**		**100.0%**
Test for overall effect: z = −16.24 (p < 0.001)						
Test for between-design differences: z = −0.193 (p = 0.847)						



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) VA NSQIP – Johnson et al. (2006) preferred, because of superior method of analysis
(b) ACS NSQIP – Malas et al. (2014) preferred, as largest publication
(c) Medicare – Schermerhorn et al. (2015) preferred, as largest, most recent publication
(d) NIS – Karthikesalingam et al. (2016) preferred, as largest publication; Lee et al. (2004) can also be included, as its recruitment period does not overlap
(e) VQI – Liang et al. (2018) preferred, because of superior method of analysis

Figure 1: Perioperative mortality – meta-analysis of casemix-adjusted observational data, with comparison with RCTs



Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval. Dashed line shows odds ratio implied by unadjusted data from 2017 NVR.

Meta-regression (observational only) – OR per year: 0.971 (0.916 to 1.029); p=0.318
Meta-regression (RCTs and observational) – OR per year: 0.972 (0.921 to 1.026); p=0.301

Figure 2: Perioperative mortality – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.2 Duration of procedure

Only 1 observational study reports duration of procedure. Figure 3 summarises the evidence.

The data are not consistent with between-design homogeneity (p<0.001). It is difficult to ascribe a single reason to this apparent discrepancy: Sugimoto et al.'s study (2017) is more recent that the RCTs, but it also took place in a very different setting (Japan).

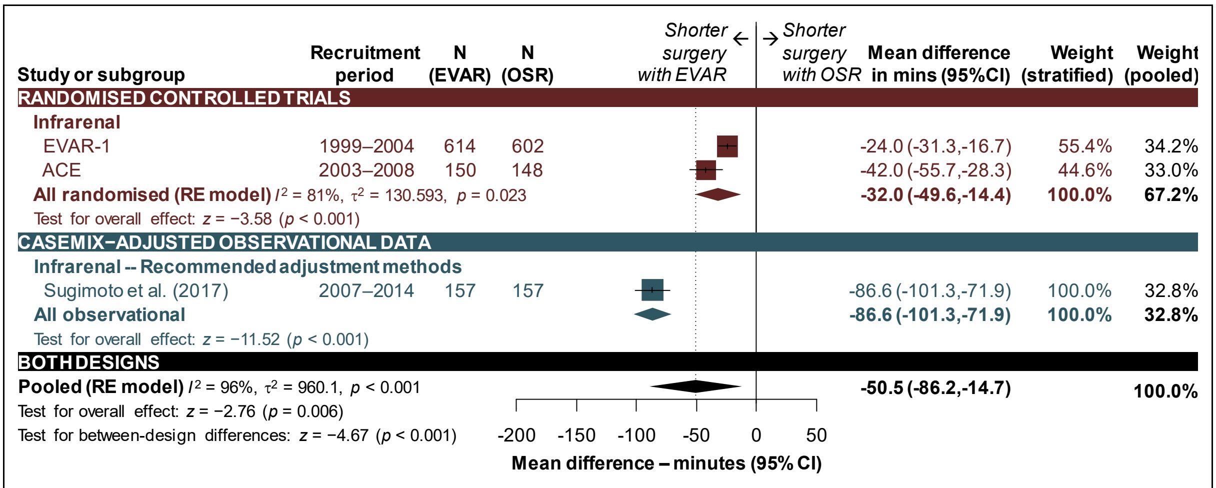


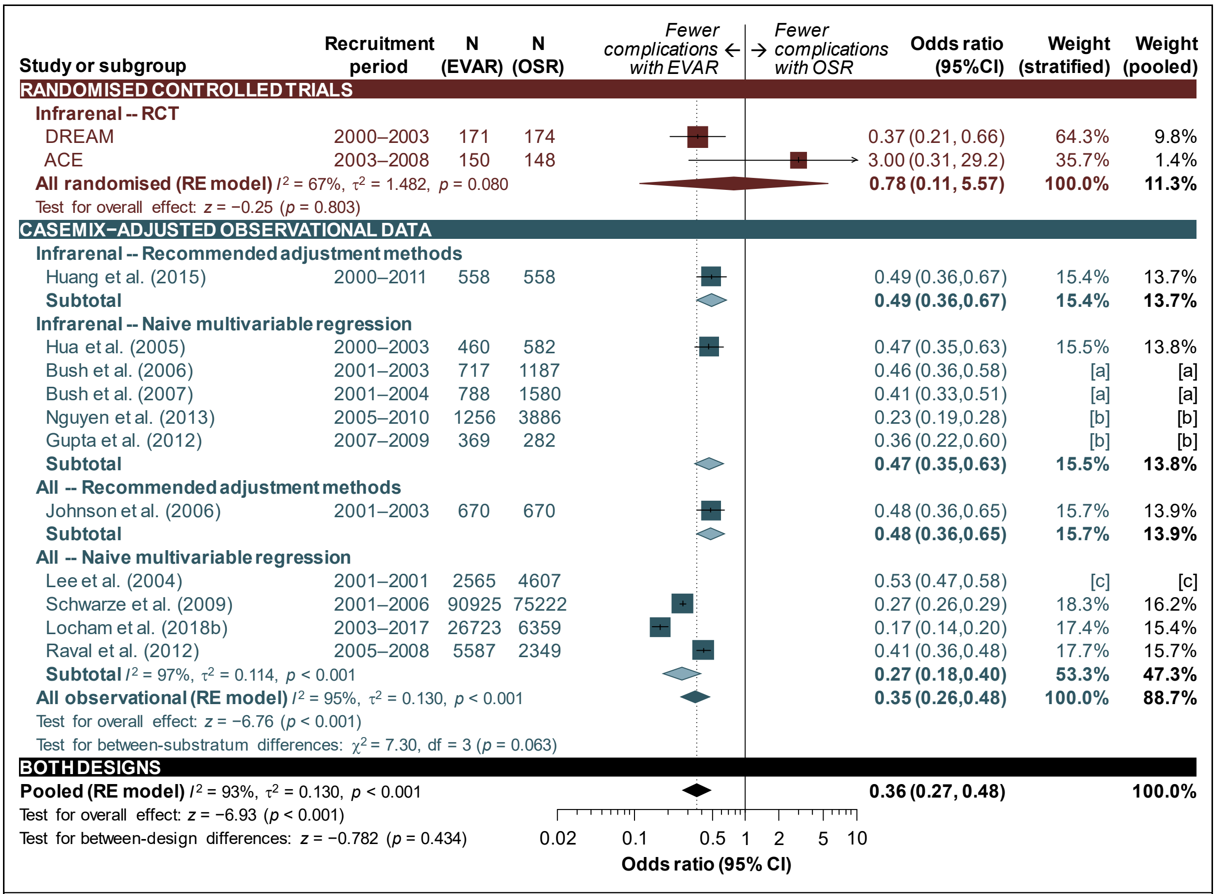
Figure 3: Duration of procedure – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

3.3.3 Perioperative complications

3.3.3.1 Perioperative complications – all

Figure 4 presents relative differences in proportions of participants experiencing at least 1 perioperative complication (according to authors' definitions).

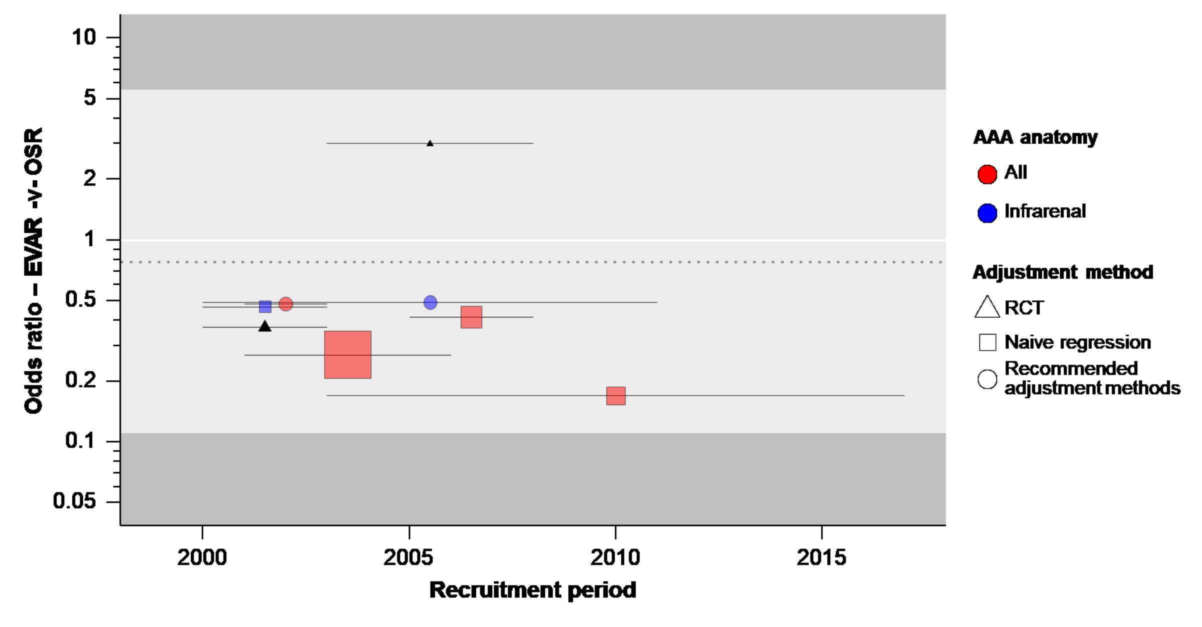
Statistically, there is evidence of heterogeneity, especially in the substratum relating to naive multivariable regressions in studies including all types of AAA. However, this is almost entirely confined to numerical variation between large studies that all agree that EVAR is associated with substantially fewer perioperative complications. The observational data appear consistent with those reported in RCTs.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) VA NSQIP – Johnson et al. (2006) preferred, because of superior method of analysis
(b) ACS NSQIP – Raval et al. (2014) preferred, as largest publication
(c) NIS – Schwarze et al. (2009) preferred, as largest publication

Figure 4: Perioperative complications (all) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

Figure 5 examines evidence for an interaction between treatment effect and recruitment period. There is no suggestion of a secular trend.



Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

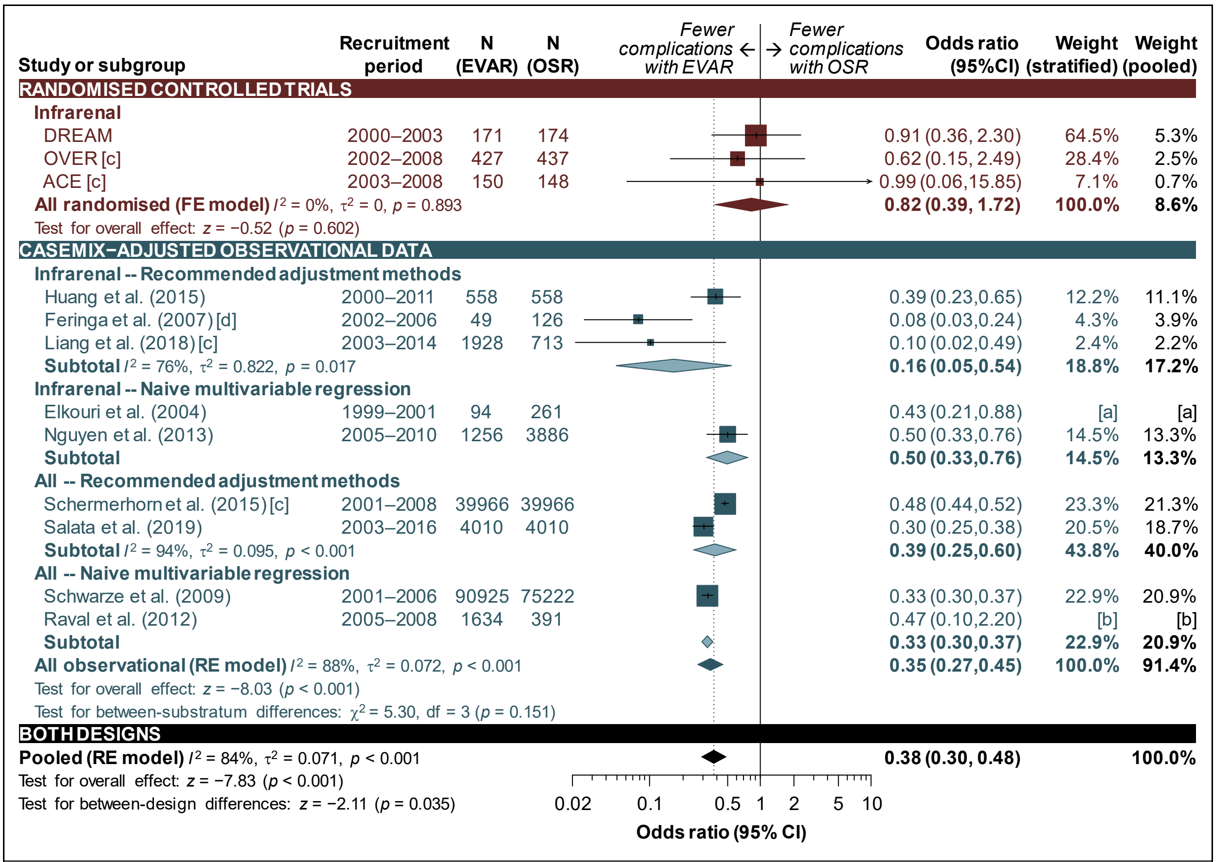
Meta-regression (observational only) – OR per year: 0.910 (0.825 to 1.004); p=0.0598
Meta-regression (RCTs and observational) – OR per year: 0.921 (0.843 to 1.007); p=0.0716

Figure 5: Perioperative complications (all) – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.3.2 Perioperative complications (cardiovascular)

Figure 6 summarises data relating to perioperative cardiovascular complications. Again, the quantitative heterogeneity in observational results reflects varying findings regarding the magnitude, rather than the direction, of effects: all included studies agree that EVAR is associated with at least half the odds of cardiovascular complication seen with OSR. The presence of statistically detectable heterogeneity is unsurprising, in a dataset including 2 very large observational studies with tight confidence intervals. It would be less likely in a synthesis of randomised trials, where trials with greater precision can be expected to converge on a 'true' mean; we can have no such expectation here, because increasing sample size will reduce the variance, but not the accuracy, of observational estimates that are subject to varying degrees of selection bias.

The casemix-adjusted observational studies estimate a benefit for EVAR that is meaningfully different from the RCT results, in which the data are consistent with no difference.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) Mayo Clinic – Huang et al. (2015) preferred, because of superior method of analysis
(b) ACS NSQIP – Nguyen et al. (2013) preferred, as largest publication
Other notes:
(c) Outcome is myocardial infarction
(d) Outcome is myocardial injury (defined as ST-segment changes and/or troponin T release)

Figure 6: Perioperative complications (cardiovascular) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

There is no evidence of any time-trend at a study level (Figure 7), though it can be seen that relatively little of the evidence comes from more recent years.

Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

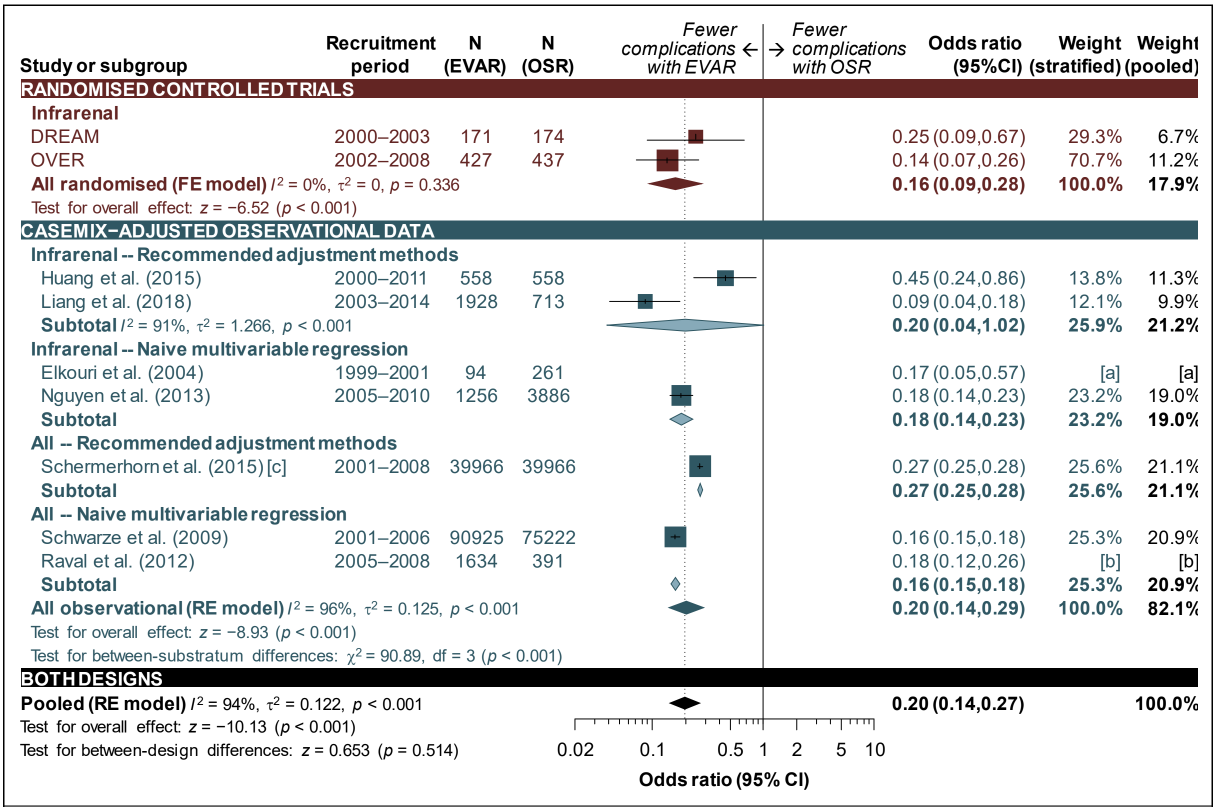
Meta-regression (observational only) – OR per year: 1.000 (0.846 to 1.182); p=0.996
Meta-regression (RCTs and observational) – OR per year: 0.946 (0.832 to 1.076); p=0.401

Figure 7: Perioperative complications (cardiovascular) – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.3.3 Perioperative complications (respiratory)

Results for perioperative respiratory complications (Figure 8) are closely comparable to those for cardiovascular events. All included studies agree that EVAR is associated with at least half the odds of respiratory complication seen with OSR, with a pooled estimate of around one-fifth the odds. This time, the estimate is effectively identical to the RCTs.

As before, we do not consider the presence of statistically detectable heterogeneity among the observational studies to be an important finding, in this context.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) Mayo Clinic – Huang et al. (2015) preferred, because of superior method of analysis
(b) ACS NSQIP – Nguyen et al. (2013) preferred, as largest publication
Other notes:
(c) Outcome is pneumonia

Figure 8: Perioperative complications (respiratory) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

There is no evidence of any time-trend at a study level (Figure 9); again, the evidence is heavily concentrated in an era that does not extend much beyond the RCTs.

Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

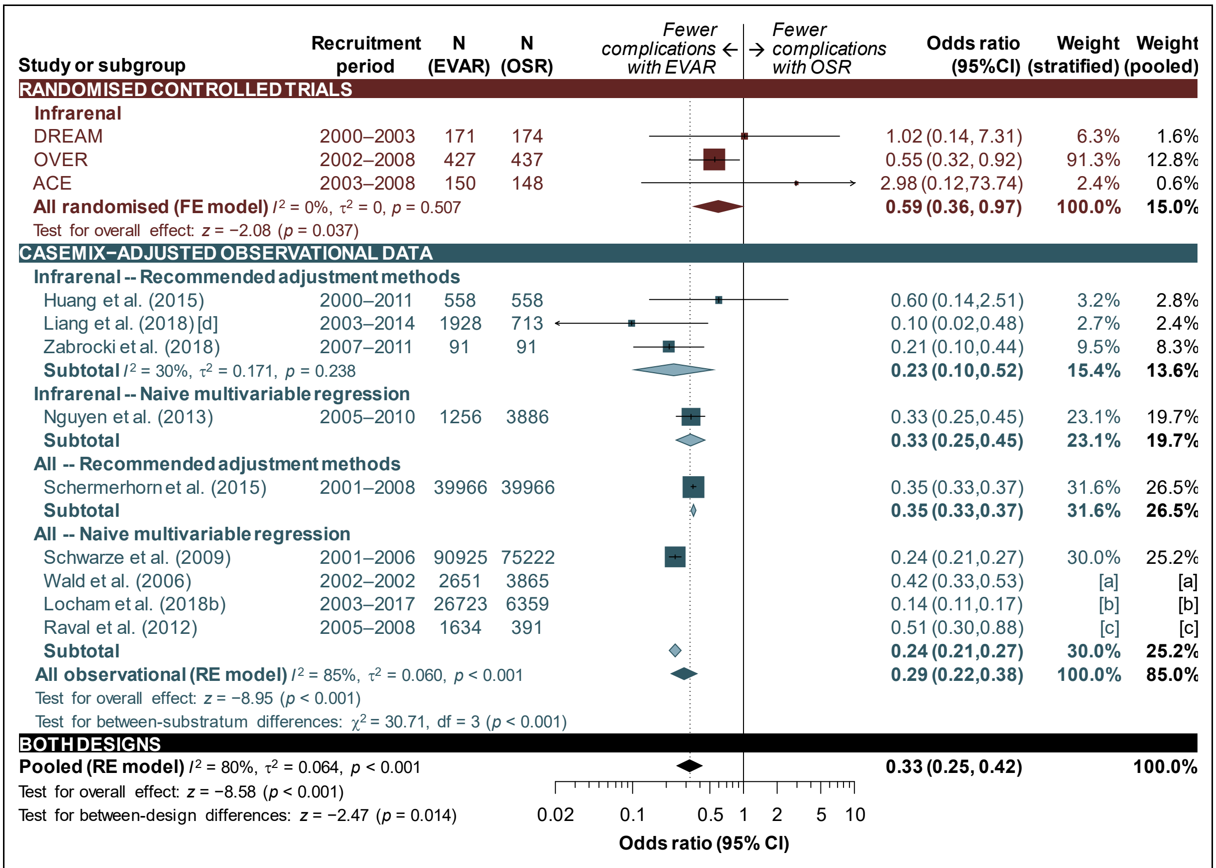
Meta-regression (observational only) – OR per year: 0.888 (0.678 to 1.164); p=0.390
Meta-regression (RCTs and observational) – OR per year: 0.918 (0.765 to 1.103); p=0.363

Figure 9: Perioperative complications (respiratory) – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.3.4 Perioperative complications (renal)

Findings for renal complications (Figure 10) are consistent with the picture seen for other perioperative morbidity. As before, the apparent heterogeneity of effect in the observational evidence is almost entirely driven by 2 very large studies – Schwarze et al. (2009) and Schermerhorn et al. (2015) – that have qualitatively similar findings.

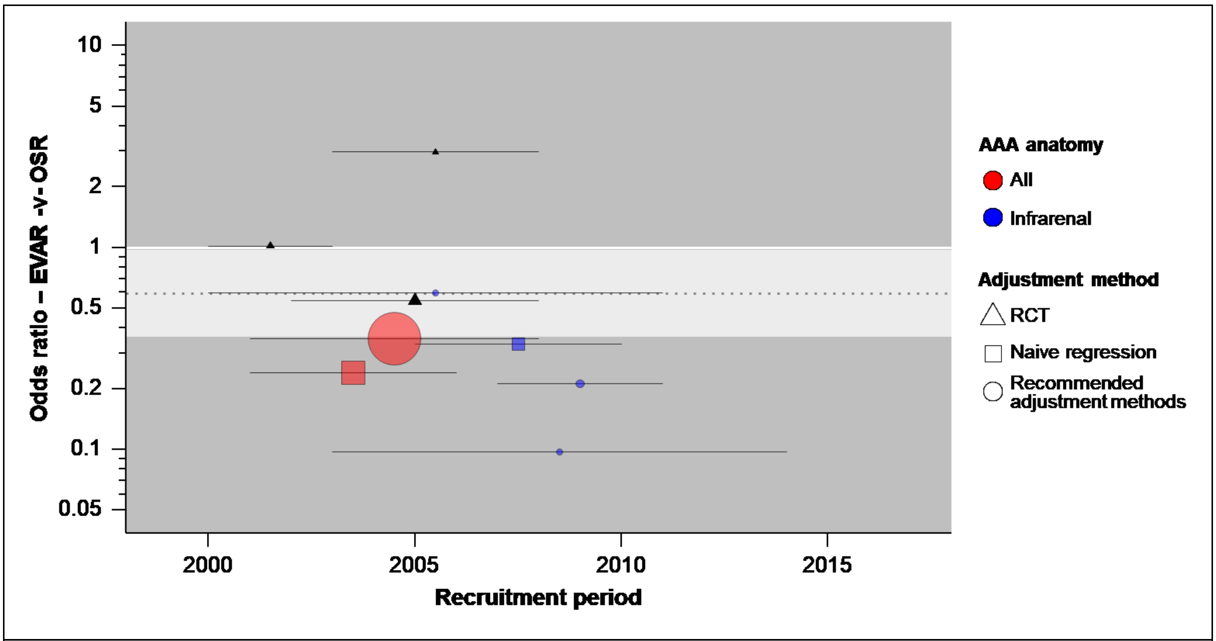
There is evidence that RCTs find a meaningfully smaller benefit for EVAR compared with OSR; however, on this occasion, the pooled summary estimates at least agree that EVAR is superior, at a 95% confidence level, in this domain.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) NIS – Schwarze et al. (2009) preferred, as largest publication
(b) VQI – Liang et al. (2018) preferred, because of superior method of analysis
(c) ACS NSQIP – Nguyen et al. (2013) preferred, as largest publication
Other notes:
(d) Outcome is requirement for haemodialysis

Figure 10: Perioperative complications (renal) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

There is no evidence of any time-trend at a study level (Figure 11).



Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

Meta-regression (observational only) – OR per year: 0.977 (0.846 to 1.128); p=0.752
Meta-regression (RCTs and observational) – OR per year: 0.921 (0.778 to 1.089); p=0.336

Figure 11: Perioperative complications (renal) – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.4 Length of critical care stay

Only 2 observational studies report length of stay in critical care. Figure 12 summarises the evidence, showing somewhere between 1.5 and 2 days' benefit for EVAR. This agrees with randomised evidence; both designs together produce a pooled effect estimate of −1.72 (−1.93 to –1.51), with no evidence of between-design heterogeneity (p=0.155).

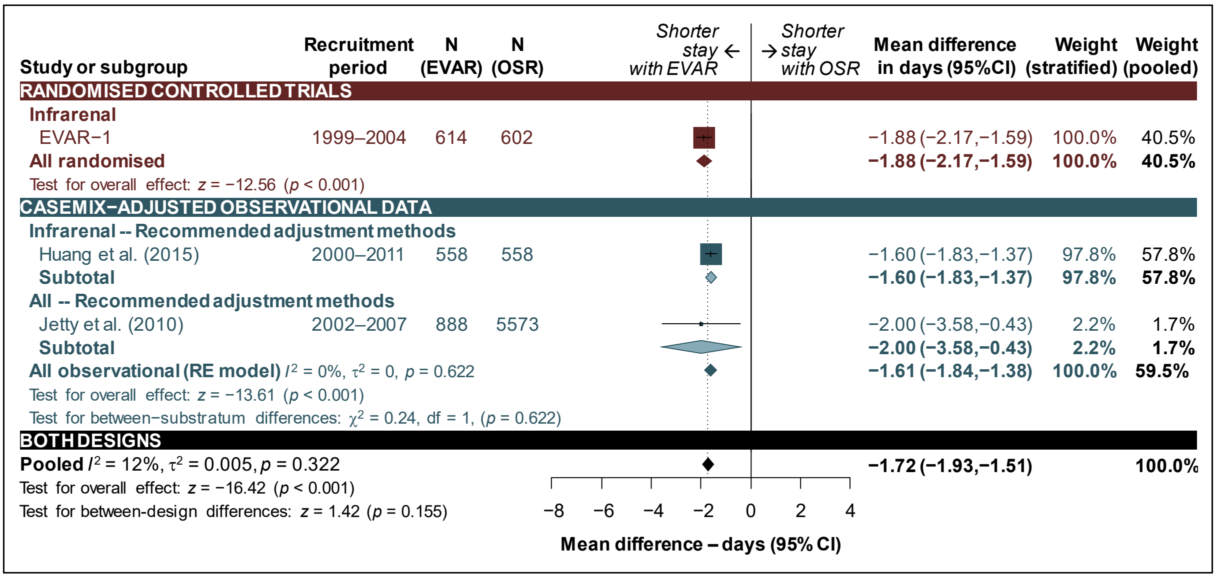


Figure 12: Length of critical care stay – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

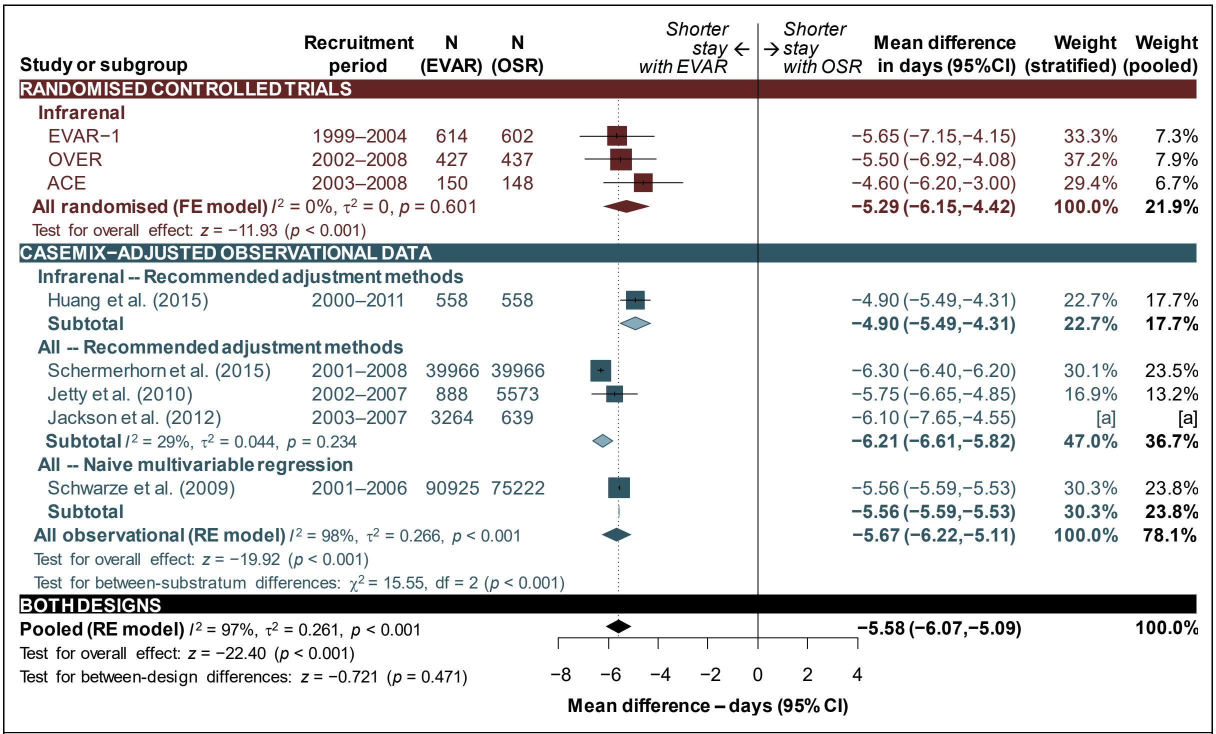
3.3.5 Length of hospital stay

Figure 13 shows a random-effects meta-analysis of casemix-adjusted observational evidence on total length of hospital stay with EVAR and OSR. It shows a clear benefit for EVAR, of the order of 5–6 fewer days' hospitalisation.

There is, on the face of it, clear evidence of heterogeneity between the observational estimates. However, this finding predominantly reflects statistical uncertainty about the precise extent to which length of stay is shorter with EVAR, rather than meaningful uncertainty about whether a large effect exists.

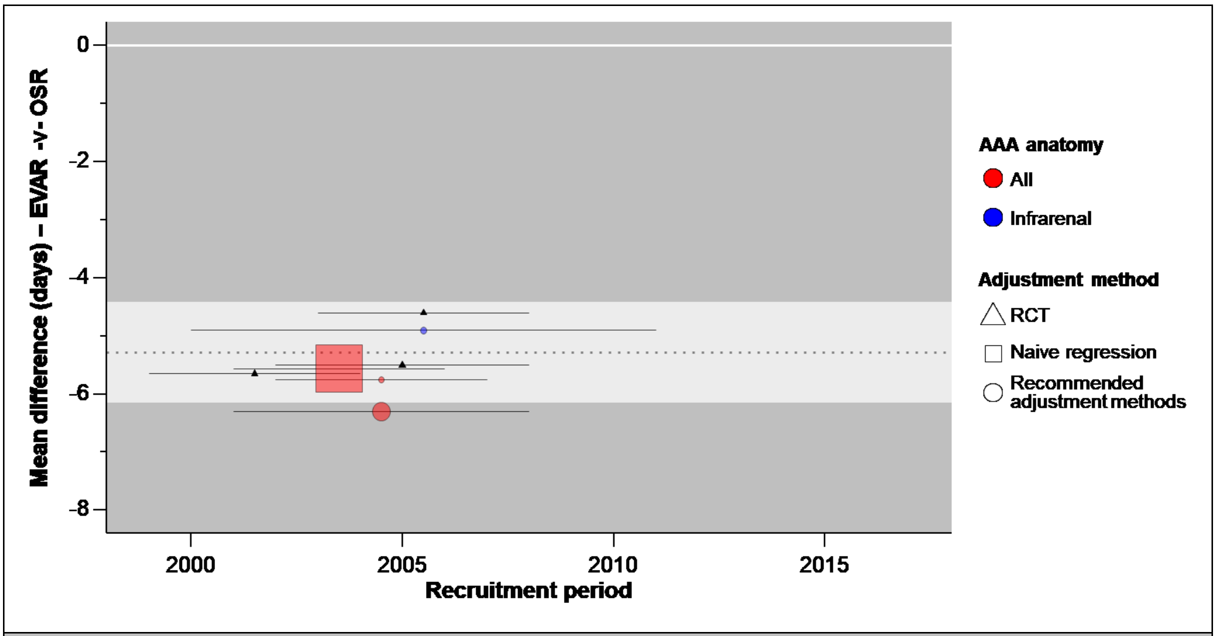
There is no evidence of between-design differences in effect (p=0.471).

Plotting the effect estimates against recruitment period (Figure 14) makes it clear that all the observational evidence is approximately contemporaneous with the RCTs, and we do not have any more recent data.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) Medicare – Schermerhorn et al. (2015) preferred, as largest, most recent publication

Figure 13: Length of hospital stay – meta-analysis of casemix-adjusted observational data, with comparison with RCTs



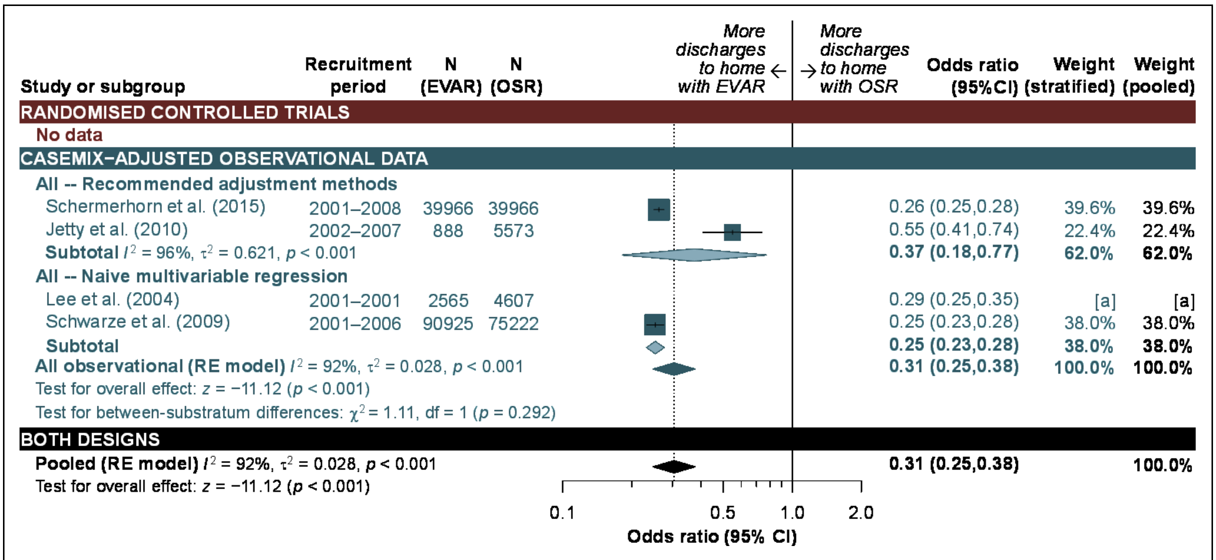
Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

Meta-regression (observational only) – mean difference (days) per year: 0.291 (-0.655 to 1.237); p=0.547
Meta-regression (RCTs and observational) – mean difference (days) per year: 0.177 (-0.284 to 0.638); p=0.452

Figure 14: Length of hospital stay – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.6 Discharge to location other than home

We identified 3 non-duplicated, casemix-adjusted observational studies that report data on whither patients were discharged at the end of their operative admission. In particular, they allow us to estimate the relative odds of participants being discharged to somewhere other than their home. This outcome is not available in any of the RCTs.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) NIS – Schwarze et al. (2009) preferred, as largest, most recent publication

Figure 15: Discharge to location other than home – meta-analysis of casemix-adjusted observational data

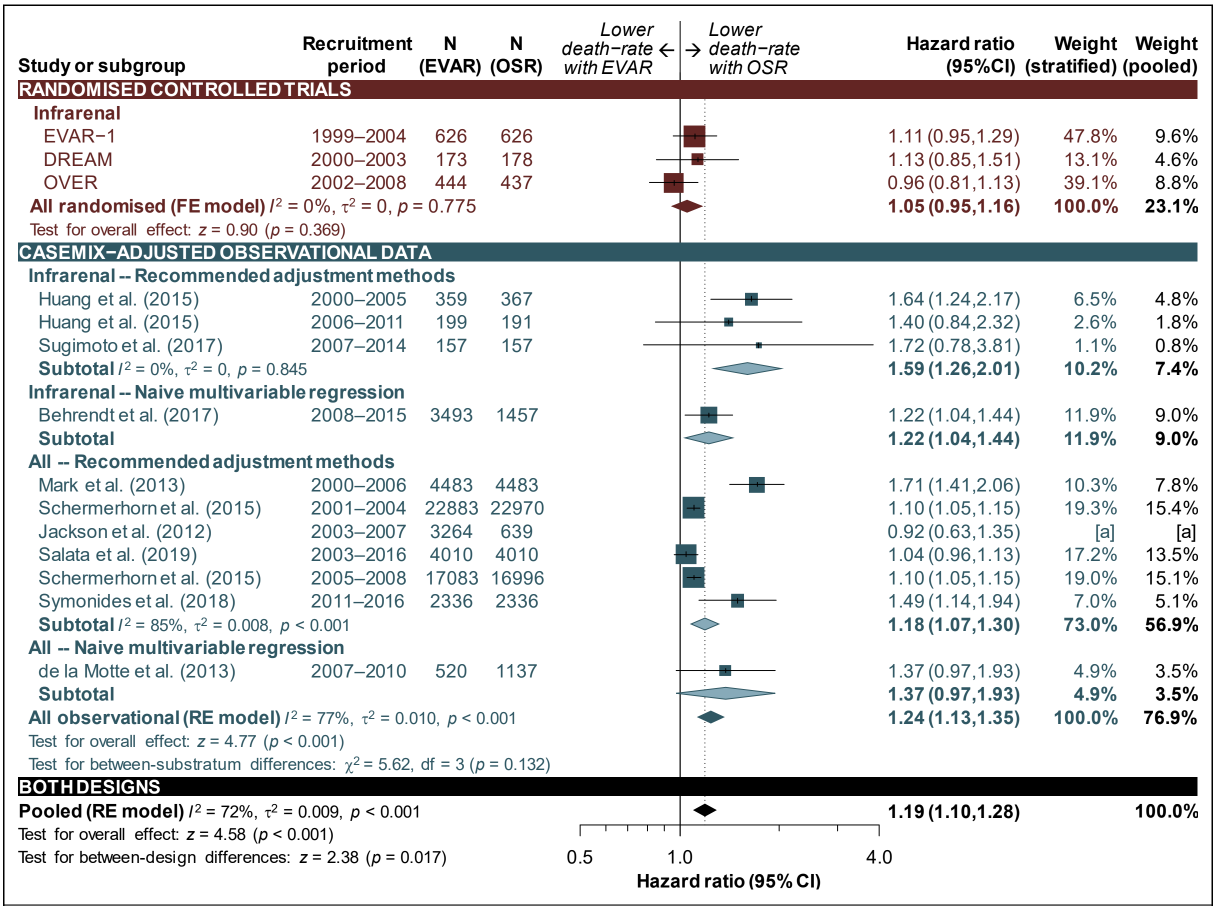
As depicted in Figure 15, the studies agree that EVAR is associated with a substantially lower chance of discharge to somewhere other than home, although there is obvious heterogeneity between the estimates.

3.3.7 Post-perioperative survival

As depicted in Figure 16, a meta-analysis of casemix-adjusted observational studies shows that EVAR is associated with a 24% (13% to 35%) increase in the hazard of post-perioperative death compared with OSR. The observational studies reflect follow-up of between 4 and 12 years, whereas the RCTs are 12–15 years.

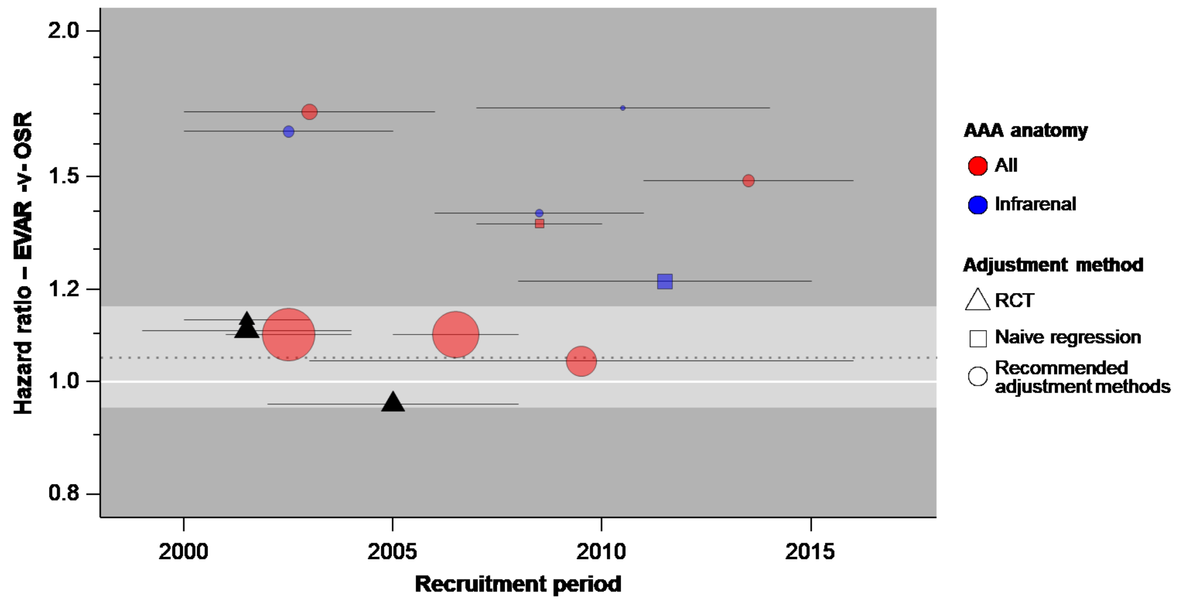
The RCTs estimate that EVAR raises this hazard to a lesser degree; at a 95% confidence level, we would reject a null hypothesis of no difference between RCTs and observational studies (p=0.017). A random-effects meta-analysis combining the design types estimates a pooled HR of 1.19 (1.10 to 1.28).

Figure 17 shows that there is no apparent trend in relative effect over time (p=0.544).



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) Medicare – Schermerhorn et al. (2015) preferred, as largest, most recent publication

Figure 16: Post-perioperative survival (long-term survival conditional on surviving the perioperative period) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs



Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

Meta-regression (observational only) – HR per year: 0.994 (0.959 to 1.030); p=0.735
Meta-regression (RCTs and observational) – HR per year: 1.004 (0.976 to 1.033); p=0.776

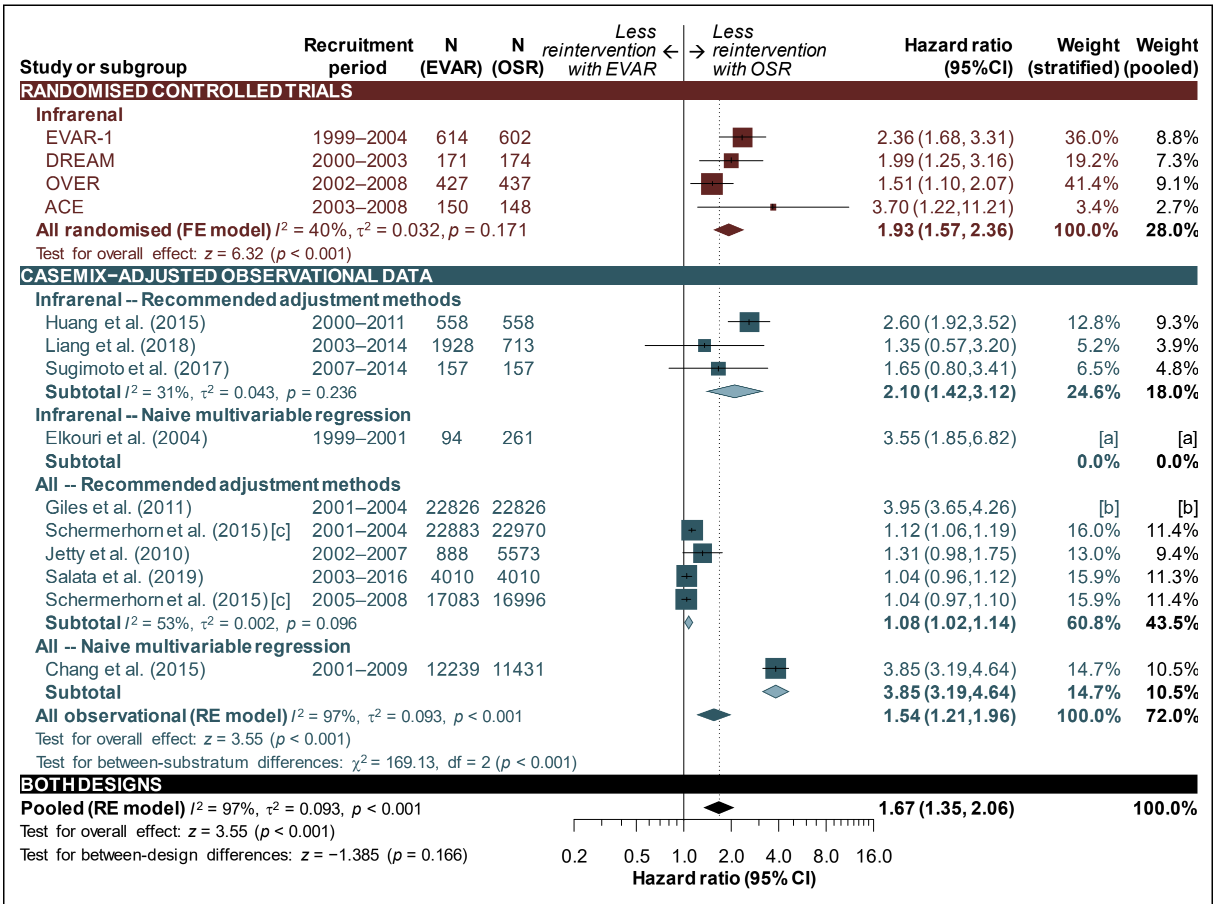
Figure 17: Perioperative mortality – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.8 Reinterventions

Individual studies adopt heterogeneous criteria to define reinterventions – some include only 'AAA-related' events; some include any surgical procedure; some do not report what definition they used. For this reason, we analysed the data in 3 groups: overall (using whatever definition the authors adopted), vascular only (e.g. graft revisions, embolisations, thrombectomies) and non-vascular only (predominantly laparotomy-related procedures, e.g., incisional hernias and bowel resections).

3.3.8.1 Reinterventions (all or unspecified)

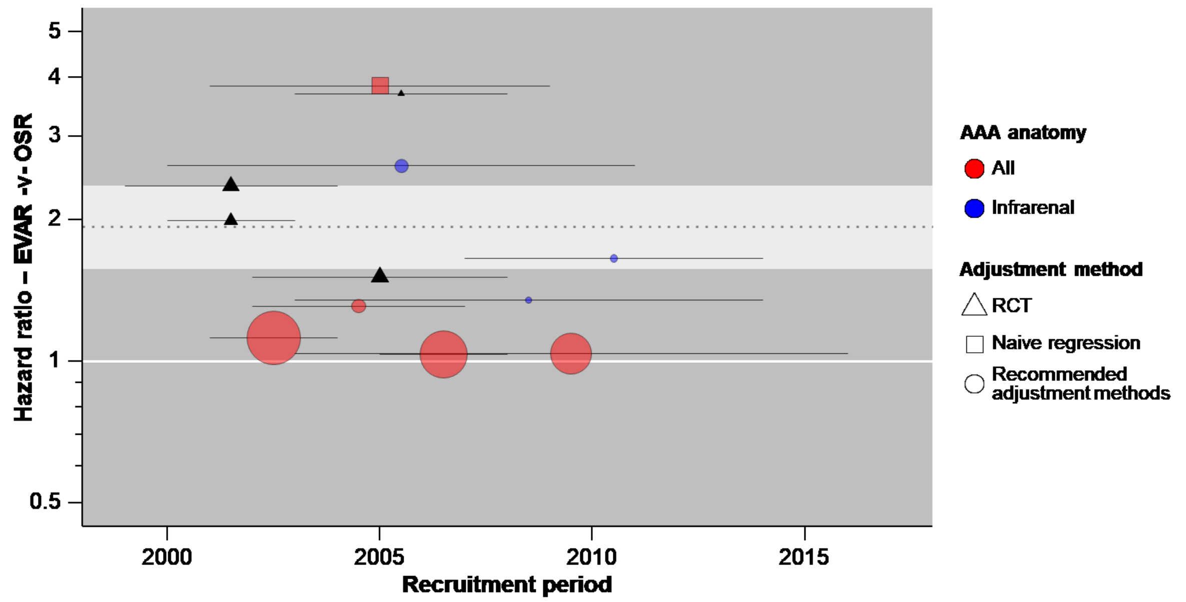
Given the clinical heterogeneity in what constitutes reintervention, it is unsurprising that there is conspicuous statistical heterogeneity between estimates (Figure 18). Nevertheless, there is a consistent finding that EVAR is associated with higher rates of reinterventions. The average effect is closely comparable to that observed in RCTs.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) Mayo clinic – Huang et al. (2015) preferred, because of superior analysis method
(b) Medicare – Schermerhorn et al. (2015) preferred, as largest, most recent publication
Other notes:
(c) Outcome is rupture or reintervention

Figure 18: Reinterventions (all or unspecified) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

There is no evidence of a between-study time-trend in meta-regression (p=0.387; see Figure 19). However, 1 included publication provides evidence of a within-study trend. Schermerhorn et al. (2015) show that, over the 8 years covered by their recruitment period (2001–2008), there was a year-on-year decrease in the number of reinterventions within 2 years of repair for people undergoing EVAR (p<0.001), but no such phenomenon in their OSR cohort (p=0.650). A test for interaction between reintervention trend and repair approach confirms the difference, suggesting the observed data are inconsistent with the null hypothesis that event-rates changed at a similar rate (p=0.001).



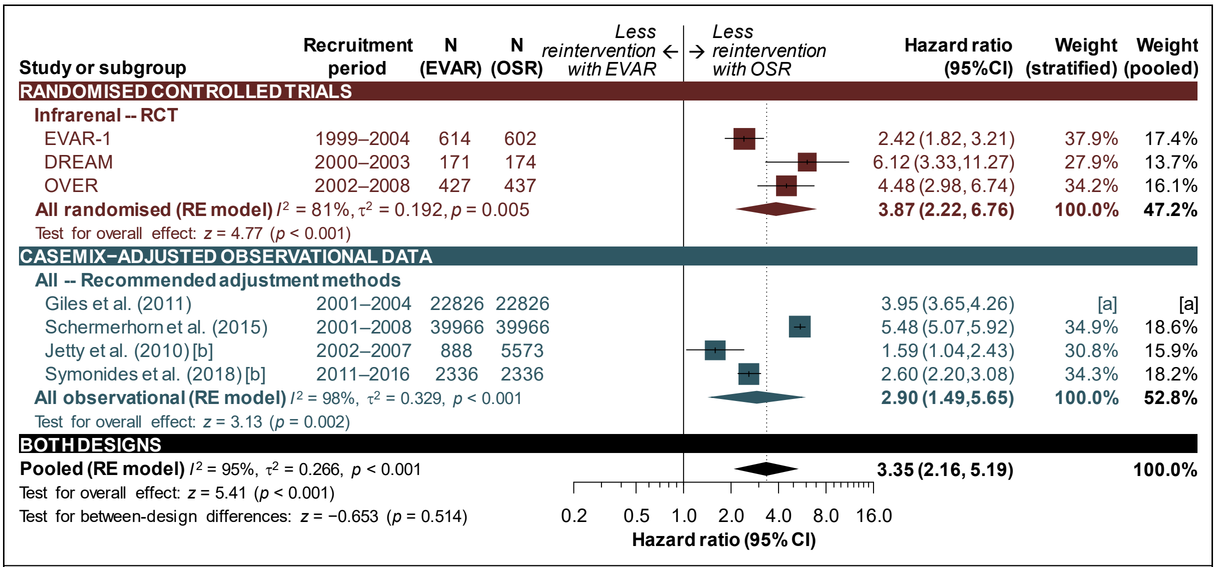
Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

Meta-regression (observational only) – HR per year: 0.970 (0.832 to 1.129); p=0.691
Meta-regression (RCTs and observational) – HR per year: 0.954 (0.863 to 1.055); p=0.362

Figure 19: Reinterventions (all or unspecified) – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

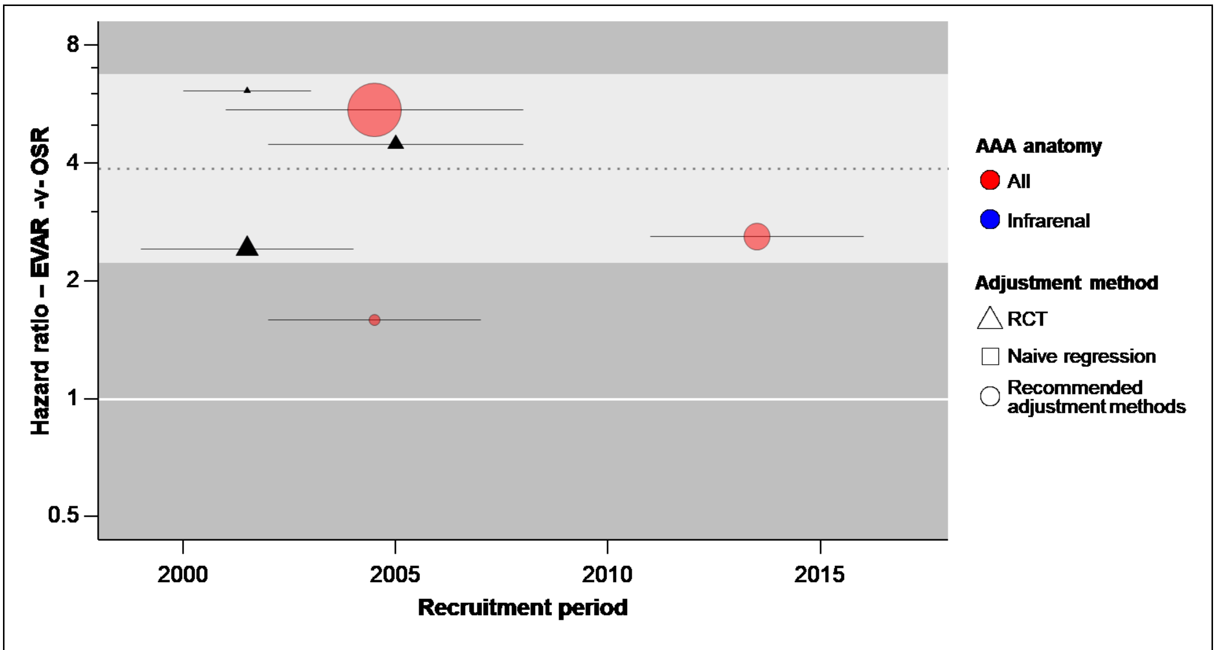
3.3.8.2 Reinterventions (vascular only)

Findings are similar in the subgroup of reinterventions that relate to vascular procedures: the included studies estimate a benefit for OSR that is heterogeneous in magnitude but consistent in direction, with a pooled effect that is closely comparable to that observed in RCTs (Figure 20). There is no evidence of a secular trend at study level (Figure 21). However, one included study (Schermerhorn et al., 2015) explores the evidence for a within-study time-trend in 2-year reintervention rates analysed by year of recruitment. These analyses show that the year-on-year reduction in all EVAR reinterventions (see 3.3.8.1, above) is primarily driven by a decrease in vascular procedures (p<0.001) which, in turn, is driven by a decrease in minor procedures, primarily coil embolization (p<0.001). In contrast, there is no evidence of a decrease in major vascular reinterventions (p=0.424). The authors summarise these data by saying the trend 'probably represents a more conservative attitude toward the management of type 2 (side branch) endoleak'. There are no trends in the rate of vascular reinterventions following OSR over the same period (p=0.112), and a test for interaction confirms the difference between the 2 approaches, in this regard (p=0.002).



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) Medicare – Schermerhorn et al. (2015) preferred, as largest, most recent publication
Other notes:
(b) Outcome reported is vascular readmissions

Figure 20: Reinterventions (vascular only) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs



Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of RCTs and light band shows 95% confidence interval.

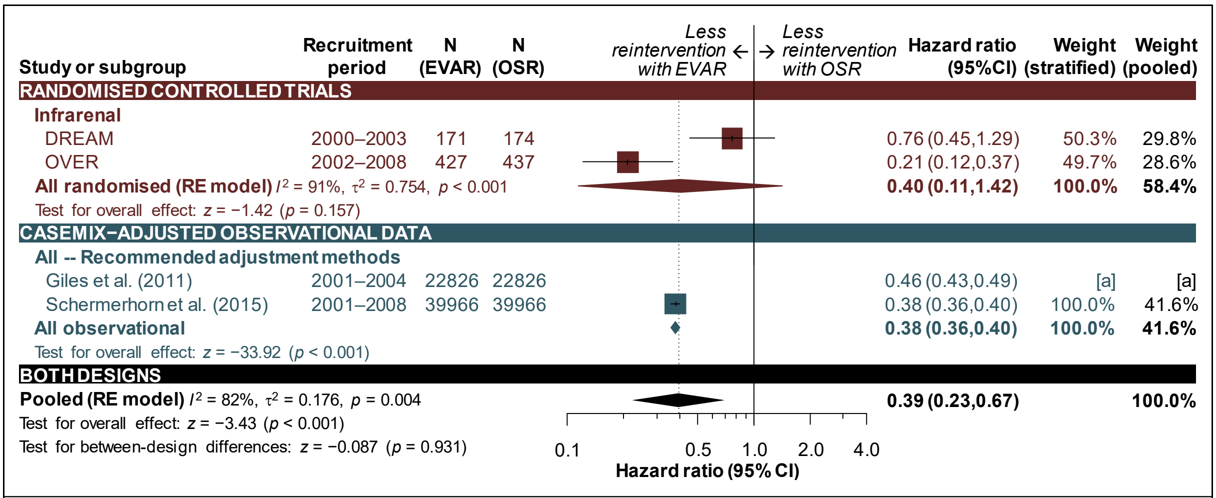
Meta-regression (observational only) – HR per year: 0.984 (0.781 to 1.240); p=0.892
Meta-regression (RCTs and observational) – HR per year: 0.973 (0.871 to 1.086); p=0.620

Figure 21: Reinterventions (vascular only) – relationship between estimated treatment effects and time of recruitment, with comparison with RCTs

3.3.8.3 Reinterventions (non-vascular only)

The 1 included nonrandomised study shows that OSR is associated with a hazard of non-vascular reintervention that is approximately 2.5 times higher than that for EVAR.

Once again, we cannot explore time-trends between studies, but Schermerhorn et al.'s (2015) publication provides details of within-study trends; in this instance, it appears there are none (p=0.550 for EVAR; p=0.845 for OSR; p=0.533 for interaction).



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) Medicare – Schermerhorn et al. (2015) preferred, as largest, most recent publication

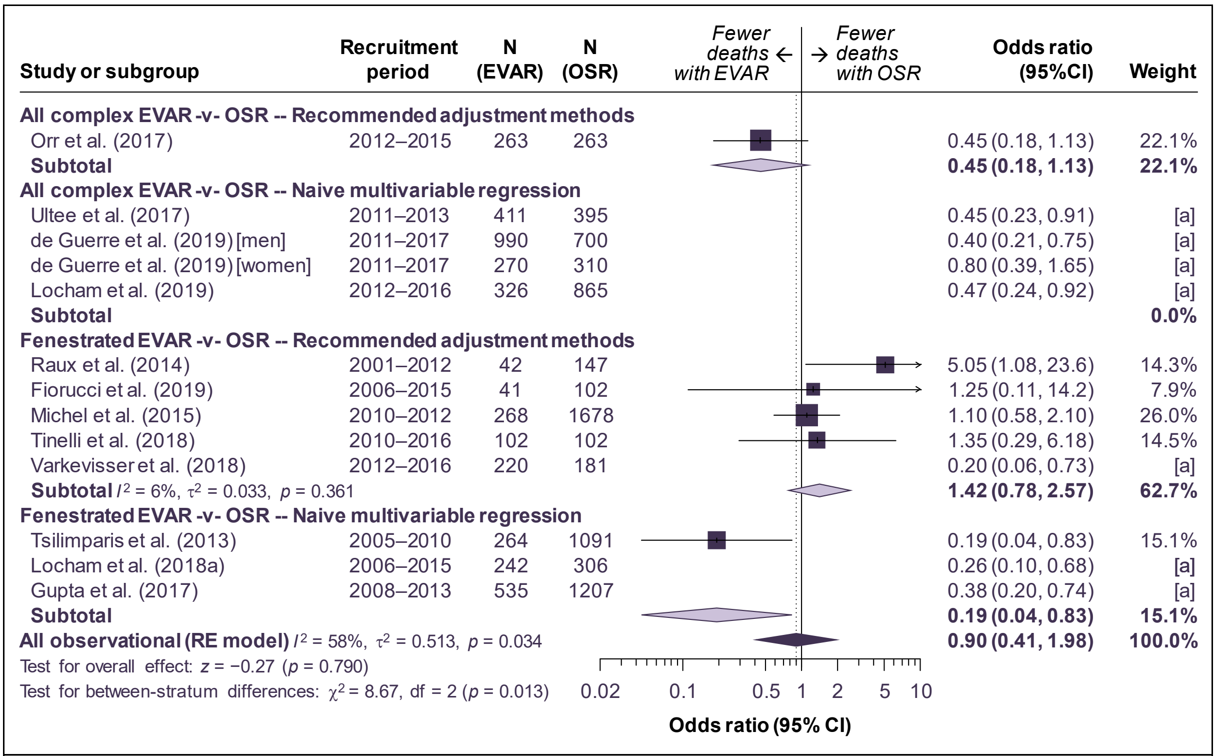
Figure 22: Reinterventions (non-vascular only) – meta-analysis of casemix-adjusted observational data, with comparison with RCTs

3.4 Evidence synthesis – complex AAAs

3.4.1 Perioperative mortality

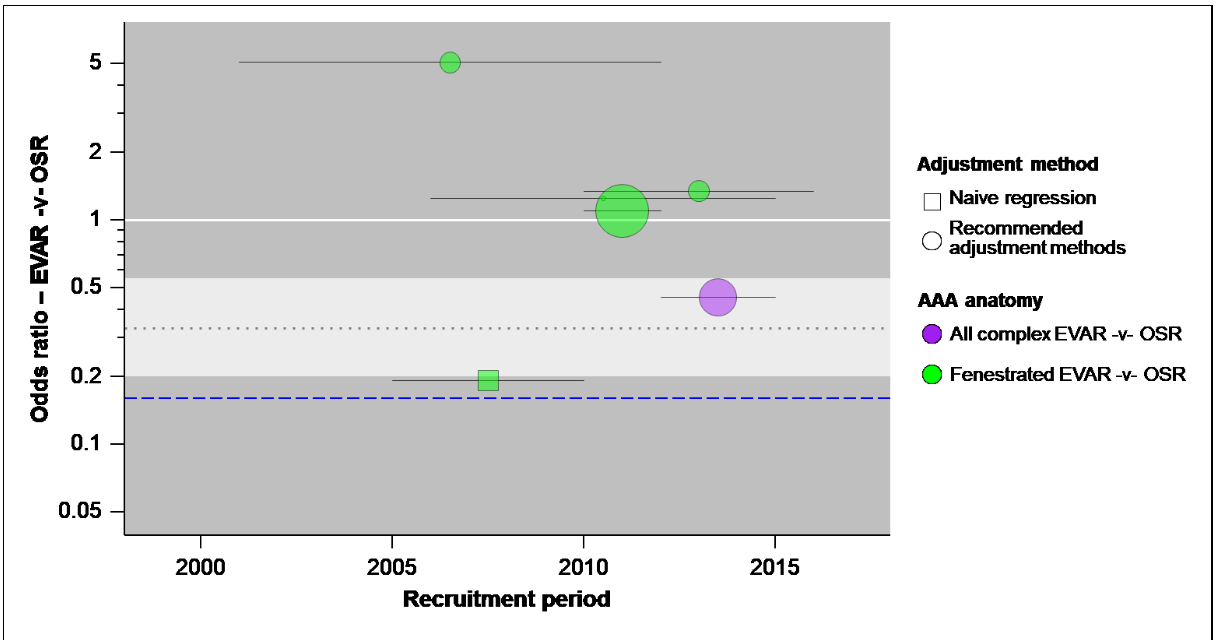
Of the 5 studies comparing perioperative mortality with EVAR and OSR using a recommended method to adjust for confounders, none estimates a benefit for EVAR, in marked contrast to the findings in the analogous analysis for infrarenal AAAs (see 3.3.1). The results of this meta-analysis (Figure 23) show that the data are consistent with no difference between surgical approaches.

Figure 24 shows that there is no evidence of a time-trend in effects across these studies.



Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) ACS NSQIP – Orr et al. (2017) preferred as largest sample with recommended method of analysis; Tsilimparis et al. (2013) can also be included as its recruitment period does not overlap with Orr et al. (2017)

Figure 23: Perioperative mortality – meta-analysis of casemix-adjusted observational data



Area of points is inversely proportional to variance of estimate. Horizontal bars indicate each study's period of recruitment. Dotted line shows mean estimate from meta-analysis of infrarenal RCTs and light band shows 95% confidence interval. Dashed line shows odds ratio implied by unadjusted data from 2017 NVR.

Meta-regression – OR per year: 0.920 (0.642 to 1.317); p=0.647

Figure 24: Perioperative mortality – relationship between estimated treatment effects and time of recruitment

3.4.2 Duration of procedure

Two casemix-adjusted studies report duration of procedure. In an echo of the infrarenal analysis (see 3.3.2), the 2 studies share the conclusion that EVAR procedures are significantly shorter than analogous OSRs, but they are at odds with each other as regards the magnitude of benefit. As shown in Figure 25, Orr et al. (2017) report a difference of just over 1 hour, whereas Tinelli et al.'s (2018) estimate is 2.5 hours. It may be relevant to note that the fEVAR and OSR cohorts in the latter study underwent their repairs at different hospitals in different countries. This makes it likely that the difference in theatre time reflects factors that go beyond the requirements imposed by the repair itself. While the study adjusted for factors related to the patients and their aneurysms, it cannot adjust for any structural and/or cultural modifiers that are exclusively associated with 1 form of repair or the other.

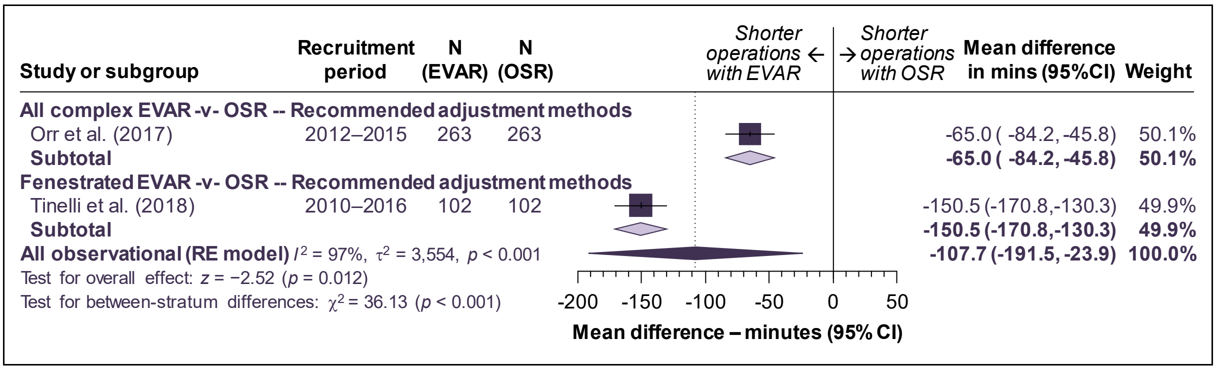


Figure 25: Duration of procedure – meta-analysis of casemix-adjusted observational data

3.4.3 Perioperative complications

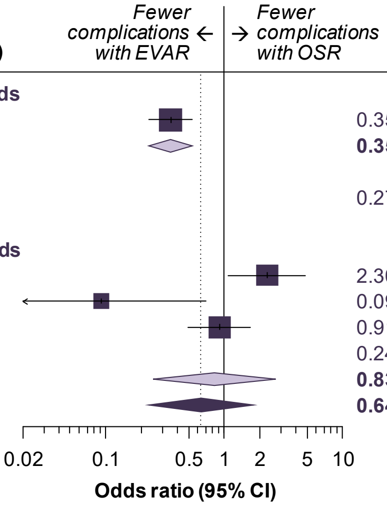
The same 4 studies (Fiorucci et al., 2019; Orr et al., 2017, Raux et al., 2014, Tinelli et al., 2018) provide unique data for all 4 datasets relating to perioperative complications. The pattern is similar in 3 of the 4 syntheses – all complications (Figure 26), respiratory events (Figure 28) and renal morbidity (Figure 29). There is conspicuous heterogeneity in these analyses, with Orr et al. and Fiorucci et al. estimating a substantial advantage for EVAR, while Raux et al. and Tinelli et al. find no such difference – indeed, Raux et al.'s overall estimate suggests that fEVAR is associated with more perioperative complications than OSR. This mirrors these studies' results for perioperative mortality (see 3.4.1).

In the case of perioperative cardiovascular complications, in contrast, there is homogeneous evidence of fewer events with EVAR than OSR across all 4 studies.

3.4.3.1 Perioperative complications – all

Study or subgroup	Recruitment period	N (EVAR)	N (OSR)	Odds ratio (95%CI)	Weight
All complex EVAR -v- OSR -- Recommended adjustment methods					
Orr et al. (2017)	2012–2015	263	263	0.35 (0.23,0.54)	30.2%
Subtotal				**0.35 (0.23,0.54)**	**30.2%**
All complex EVAR -v- OSR -- Naive multivariable regression					
Ultee et al. (2017)	2011–2013	411	395	0.27 (0.18,0.40)	[a]
Subtotal					**0.0%**
Fenestrated EVAR -v- OSR -- Recommended adjustment methods					
Raux et al. (2014)	2001–2012	42	147	2.30 (1.09,4.85)	27.2%
Fiorucci et al. (2019)	2006–2015	41	102	0.09 (0.01,0.70)	14.0%
Tinelli et al. (2018)	2010–2016	102	102	0.91 (0.50,1.66)	28.6%
Varkevisser et al. (2018)	2012–2016	220	181	0.24 (0.13,0.43)	[a]
Subtotal $I^2 = 80\%$, $\tau^2 = 0.810$, $p = 0.007$				**0.83 (0.25,2.75)**	**69.8%**
All observational (RE model) $I^2 = 87\%$, $\tau^2 = 0.848$, $p < 0.001$				**0.64 (0.23,1.76)**	**100.0%**

Test for overall effect: $z = -0.87$ ($p = 0.386$)
Test for between-stratum differences: $\chi^2 = 1.75$, df = 1 ($p = 0.186$)



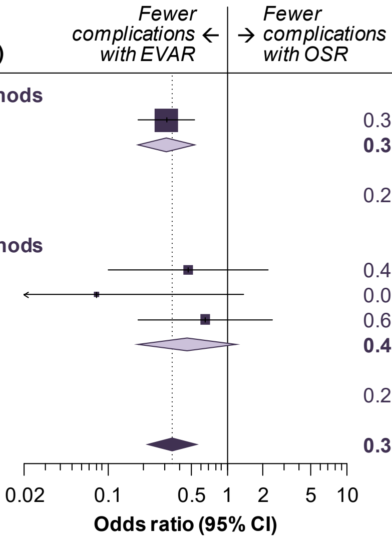
Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) ACS NSQIP – Orr et al. (2017) preferred as largest sample with recommended method of analysis

Figure 26: Perioperative complications (all) – meta-analysis of casemix-adjusted observational data

3.4.3.2 Perioperative complications (cardiovascular)

Study or subgroup	Recruitment period	N (EVAR)	N (OSR)	Odds ratio (95%CI)	Weight
All complex EVAR -v- OSR -- Recommended adjustment methods					
Orr et al. (2017) [b]	2012–2015	263	263	0.31 (0.18,0.54)	74.8%
Subtotal				**0.31 (0.18,0.54)**	**74.8%**
All complex EVAR -v- OSR -- Naive multivariable regression					
Locham et al. (2019) [c]	2012–2016	326	865	0.22 (0.12,0.40)	[a]
Subtotal					**0.0%**
Fenestrated EVAR -v- OSR -- Recommended adjustment methods					
Raux et al. (2014)	2001–2012	42	147	0.47 (0.10,2.20)	9.3%
Fiorucci et al. (2019)	2006–2015	41	102	0.08 (0.00,1.38)	2.7%
Tinelli et al. (2018)	2010–2016	102	102	0.65 (0.18,2.39)	13.2%
Subtotal $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.420$				**0.46 (0.18,1.18)**	**25.2%**
Fenestrated EVAR -v- OSR -- Naive multivariable regression					
Gupta et al. (2017)	2008–2013	535	1207	0.29 (0.15,0.56)	[a]
Subtotal					**0.0%**
All observational (RE model) $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.525$				**0.34 (0.21,0.55)**	**100.0%**

Test for overall effect: $z = -4.45$ ($p < 0.001$)
Test for between-stratum differences: $\chi^2 = 0.50$, df = 1 ($p = 0.479$)

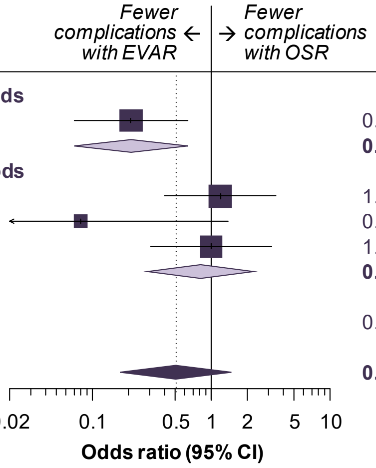


Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) ACS NSQIP – Orr et al. (2017) preferred as largest sample with recommended method of analysis
Other notes:
(b) Outcome is cardiac or respiratory failure
(c) Outcome is cardiopulmonary failure

Figure 27: Perioperative complications (cardiovascular) – meta-analysis of casemix-adjusted observational data

3.4.3.3 Perioperative complications (respiratory)

Study or subgroup	Recruitment period	N (EVAR)	N (OSR)	Odds ratio (95%CI)	Weight
All complex EVAR -v- OSR -- Recommended adjustment methods					
Orr et al. (2017) [b]	2012–2015	263	263	0.21 (0.07,0.63)	30.0%
Subtotal				**0.21 (0.07,0.63)**	**30.0%**
Fenestrated EVAR -v- OSR -- Recommended adjustment methods					
Raux et al. (2014)	2001–2012	42	147	1.19 (0.41,3.48)	30.4%
Fiorucci et al. (2019)	2006–2015	41	102	0.08 (0.00,1.38)	10.8%
Tinelli et al. (2018)	2010–2016	102	102	1.00 (0.31,3.21)	28.8%
Subtotal I^2 = 35%, τ^2 = 0.288, p = 0.215				**0.81 (0.29,2.26)**	**70.0%**
Fenestrated EVAR -v- OSR -- Naive multivariable regression					
Gupta et al. (2017)	2008–2013	535	1207	0.11 (0.07,0.19)	[a]
Subtotal					**0.0%**
All observational (RE model) I^2 = 61%, τ^2 = 0.690, p = 0.052				**0.50 (0.17,1.47)**	**100.0%**
Test for overall effect: $z = -1.25$ (p = 0.211)					
Test for between-stratum differences: χ^2 = 3.13, df = 1 (p = 0.077)					

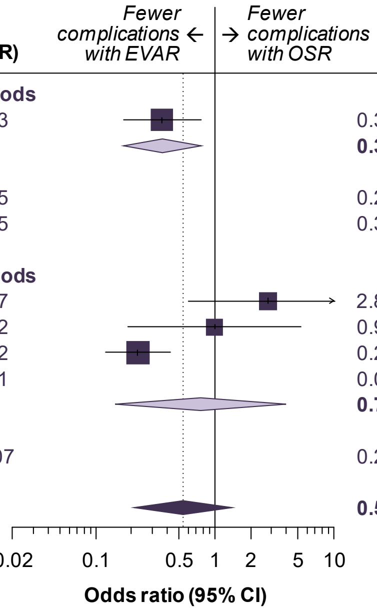


Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) ACS NSQIP – Orr et al. (2017) preferred as largest sample with recommended method of analysis
Other notes:
(b) Outcome is pneumonia

Figure 28: Perioperative complications (respiratory) – meta-analysis of casemix-adjusted observational data

3.4.3.4 Perioperative complications (renal)

Study or subgroup	Recruitment period	N (EVAR)	N (OSR)	Odds ratio (95%CI)	Weight
All complex EVAR -v- OSR -- Recommended adjustment methods					
Orr et al. (2017)	2012–2015	263	263	0.36 (0.17, 0.76)	30.8%
Subtotal				**0.36 (0.17, 0.76)**	**30.8%**
All complex EVAR -v- OSR -- Naive multivariable regression					
Ultee et al. (2017)	2011–2013	411	395	0.21 (0.10, 0.46)	[a]
Locham et al. (2019)	2012–2016	326	865	0.35 (0.18, 0.68)	[a]
Subtotal					**0.0%**
Fenestrated EVAR -v- OSR -- Recommended adjustment methods					
Raux et al. (2014)	2001–2012	42	147	2.80 (0.60, 13.0)	19.0%
Fiorucci et al. (2019)	2006–2015	41	102	0.99 (0.19, 5.35)	17.3%
Tinelli et al. (2018)	2010–2016	102	102	0.23 (0.12, 0.42)	32.8%
Varkevisser et al. (2018)	2012–2016	220	181	0.08 (0.02, 0.28)	[a]
Subtotal I^2 = 81%, τ^2 = 1.705, p = 0.006				**0.76 (0.14, 4.00)**	**69.2%**
Fenestrated EVAR -v- OSR -- Naive multivariable regression					
Gupta et al. (2017) [b]	2008–2013	535	1207	0.26 (0.13, 0.52)	[a]
Subtotal					**0.0%**
All observational (RE model) I^2 = 71%, τ^2 = 0.610, p = 0.015				**0.54 (0.21, 1.40)**	**100.0%**
Test for overall effect: $z = -1.26$ (p = 0.208)					
Test for between-stratum differences: χ^2 = 0.64, df = 1 (p = 0.422)					

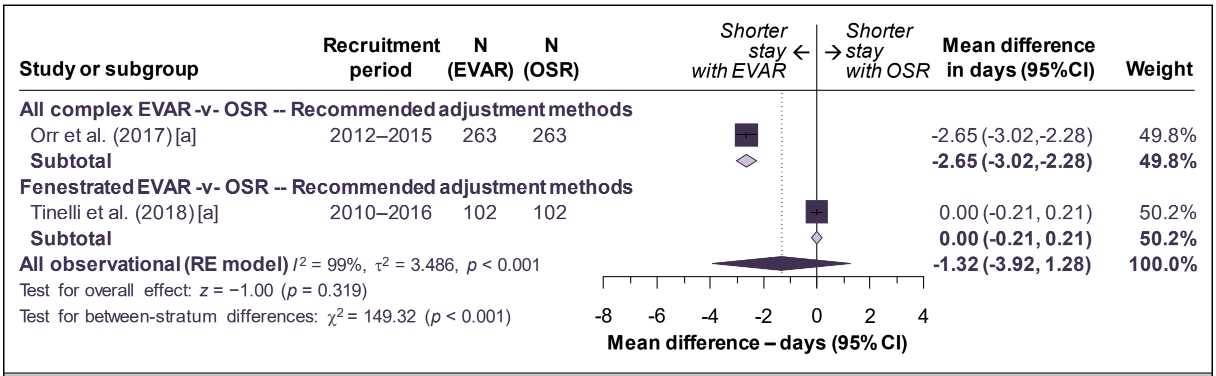


Datapoints excluded from meta-analysis to avoid double-counting patients from same datasource:
(a) ACS NSQIP – Orr et al. (2017) preferred as largest sample with recommended method of analysis
Other notes:
(b) Outcome is acute renal failure needing dialysis

Figure 29: Perioperative complications (renal) – meta-analysis of casemix-adjusted observational data

3.4.4 Length of critical care stay

No studies report mean duration of critical care. However, 2 studies report medians and interquartile ranges from propensity-score-matched cohorts. In the absence of other data, we used published methods to estimate mean and variance from these quantiles (Wan et al., 2014; Luo et al., 2018). Results are shown in Figure 30. There is very clear disagreement between the 2 datapoints: Orr et al. (2017) find that people undergoing EVAR require over 2.5 fewer days' critical care than people who have had OSR. Tinelli et al. (2018), on the other hand, report no difference between the 2 groups (in fact, it appears that the mean duration of stay is likely to be somewhat longer in the EVAR group, as the maximum observed value in that group was 286 days, compared with 11 days for EVAR; even if the very high observation in the EVAR group represents a single, extreme outlying estimate, the mean expectation of critical care time would be substantially affected in a sample of 102 participants).

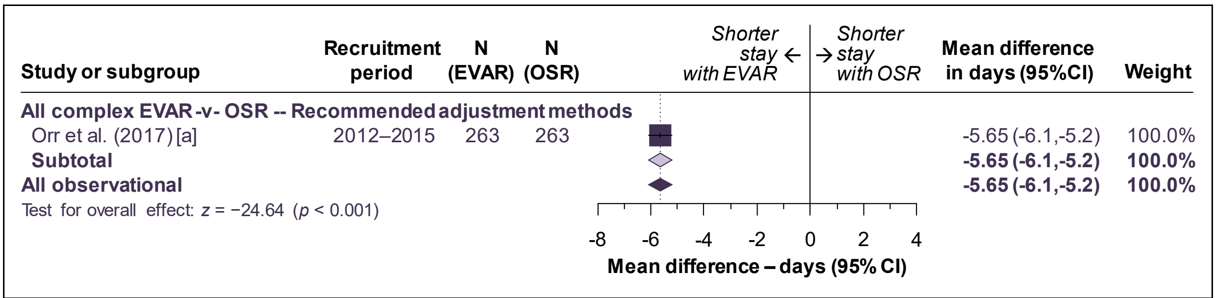


(a) Mean difference and its variance approximated from arm-level medians and IQRs using the methods of Wan et al. (2014) and Luo et al. (2018)

Figure 30: Length of critical care stay – meta-analysis of casemix-adjusted observational data

3.4.5 Length of hospital stay

The 1 casemix-adjusted observational study that reports length of hospital stay (again, as median and IQR, from which we have approximated mean and SD) has findings that are extremely similar to those in the infrarenal dataset (see Figure 31 and compare with 3.3.5). Both datasets identify a substantial benefit of the order of around 5.5 fewer days' hospitalisation for people receiving EVAR (the infrarenal RCTs reach a closely comparable conclusion, too).



(a) Mean difference and its variance approximated from arm-level medians and IQRs using the methods of Wan et al. (2014) and Luo et al. (2018)

Figure 31: Length of hospital stay – meta-analysis of casemix-adjusted observational data

3.4.6 Discharge to location other than home

There is only 1 estimate of discharge probability available for complex EVAR – see Figure 32. The estimate that EVAR is associated with around a fourfold reduction in the odds of discharge to somewhere other than home is closely comparable to the evidence found in infrarenal cases (see 3.3.6).

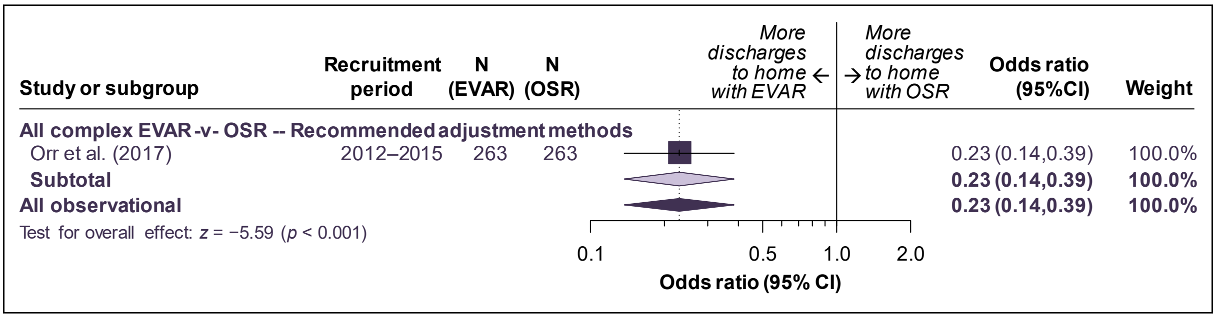


Figure 32: Discharge to location other than home – meta-analysis of casemix-adjusted observational data

3.4.7 Post-perioperative survival

Two studies report long-term survival following complex EVAR or OSR. They show that, for people who survive the perioperative period, fEVAR is associated with approximately double the hazard of mortality.

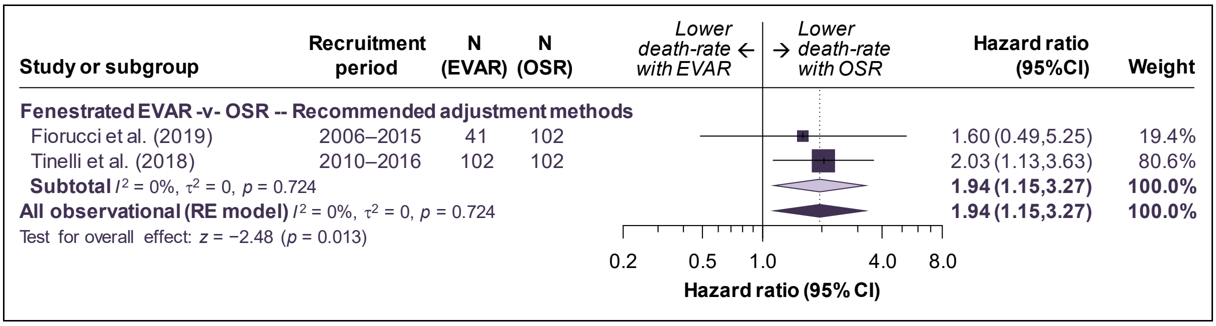


Figure 33: Post-perioperative survival (long-term survival conditional on surviving the perioperative period) – meta-analysis of casemix-adjusted observational data

3.4.8 Reinterventions

3.4.8.1 Reinterventions (all or unspecified)

One study (Tinelli et al., 2018) reports time to reintervention with fEVAR and OSR. The data suggest that people experience 3 times the rate of reintervention with fEVAR as with OSR. The authors report 2 hypothesis tests for their dataset, one of which is adjudged significant and one of which is not; our analysis suggests that the data are consistent with no difference at a 95% confidence level.

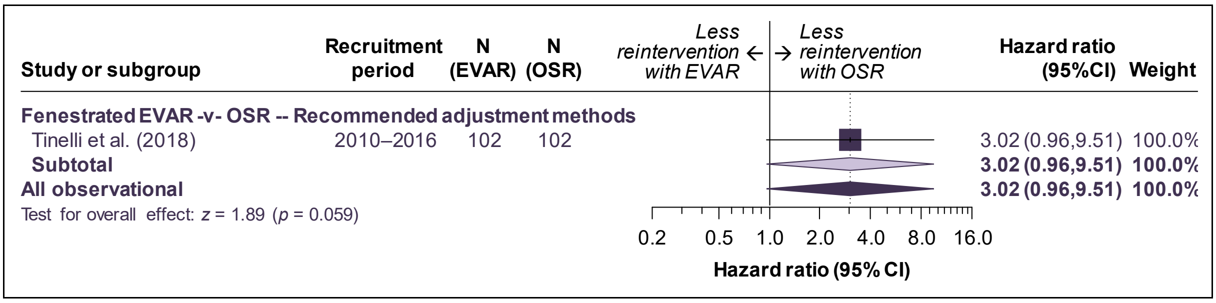


Figure 34: Reinterventions (all or unspecified) – meta-analysis of casemix-adjusted observational data

4 Evidence statements

4.1 Exclusively or predominantly infrarenal AAA

Very low-quality evidence from casemix-adjusted observational studies shows that:

- EVAR is superior to OSR in the following domains:
  - fewer perioperative deaths (21 studies comprising 386,466 participants)
  - shorter procedure time (1 study comprising 314 participants)
  - fewer perioperative complications (6 studies comprising 210,663 participants), including
    - fewer perioperative cardiovascular complications (7 studies comprising 263,173 participants)
    - fewer perioperative respiratory complications (5 studies comprising 254,978 participants)
    - fewer perioperative renal complications (6 studies comprising 255,160 participants)
  - shorter duration of critical care (2 studies comprising 7,577 participants)
  - shorter duration of hospitalisation (4 studies comprising 253,656 participants)
  - a greater chance of discharge to home – as opposed to another hospital or residential care (3 studies comprising 252,540 participants)
  - fewer non-vascular reinterventions (1 study with 79,932 participants; low-quality evidence)
- OSR is superior to EVAR in the following domains:
  - a lower hazard of death in people who survive the perioperative period (10 studies comprising 109,627 participants)
  - fewer reinterventions (8 studies comprising 122,154 participants), including
    - fewer vascular reinterventions (3 studies comprising 91,065 participants)
- There is no evidence that the relative effectiveness of EVAR and OSR has changed over time in any of these domains
- There is no evidence of systematic differences in direction or magnitude of effect between casemix-adjusted observational data and randomised trials

4.2 Complex AAA

Very low-quality evidence from casemix-adjusted observational studies shows that:

- Complex EVAR is superior to complex OSR in the following domains:
  - shorter procedure time (2 studies comprising 730 participants)
  - fewer perioperative cardiovascular complications (4 studies comprising 1,062 participants)
  - shorter duration of hospitalisation (1 study comprising 526 participants)
  - a greater chance of discharge to home – as opposed to another hospital or residential care (1 study comprising 526 participants)
- Complex OSR is superior to complex EVAR in the following domains:
  - a lower hazard of death in people who survive the perioperative period (2 studies comprising 347 participants)
- Complex EVAR and complex OSR cannot be differentiated in the following domains:
  - perioperative mortality (6 studies comprising 4,363 participants)
  - perioperative complications (4 studies comprising 1,062 participants), including

    - perioperative respiratory complications (4 studies comprising 1,062 participants)
    - perioperative renal complications (4 studies comprising 1,062 participants)
  - duration of critical care (2 studies comprising 730 participants)
  - reintervention rate (1 study comprising 204 participants)

5 References

Faria, R., Hernandez Alava, M., Manca, A., Wailoo, A.J. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. Available from http://www.nicedsu.org.uk

Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC medical research methodology. 2012 Dec;12(1):9.

Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annual review of public health. 2000 May;21(1):121-45.

Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Statistical methods in medical research. 2018 Jun;27(6):1785-805.

Newgard CD, Hedges JR, Arthur M, Mullins RJ. Advanced statistics: the propensity score—a method for estimating treatment effect in observational research. Academic Emergency Medicine. 2004 Sep;11(9):953-61.

Sedgwick P. Meta-analyses: what is heterogeneity? BMJ. 2015 Mar 16;350:h1435.

Sterne JAC, Higgins JPT, Elbers RG, Reeves BC and the development group for ROBINS-I. Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance, updated 12 October 2016. Available from http://www.riskofbias.info [accessed September 2018]

Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC medical research methodology. 2014 Dec;14(1):135.

Appendix A Literature search strategies

A.1 Clinical search literature search strategy

A.1.1 Databases

Bibliographic databases searched for the review question:

- EMBASE (Ovid)
- MEDLINE (Ovid)
- MEDLINE Epub Ahead of Print (Ovid)
- MEDLINE In-Process (Ovid)

A.1.2 Search strategy

The strategy used in Medline is detailed below; the same search (with appropriate adjustments) was used in the other databases listed above.

Medline strategy, searched 6 August 2018
Database: Ovid MEDLINE(R) <1946 to 3 August 2018
Search Strategy:

Strategy used:

Database: Ovid MEDLINE(R) <1946 to August 02, 2018>
Search Strategy:
----
1 Aortic Aneurysm, Abdominal/ (17305)
2 (aneurysm* adj4 (abdom* or thoracoabdom* or thoraco-abdom* or aort* or spontan* or juxtarenal* or juxta-renal* or juxta renal* or paraerenal* or para-renal* or para renal* or suprarenal* or supra renal* or supra-renal* or short neck* or short-neck* or shortneck* or visceral aortic segment*)).tw. (34306)
3 (AAA or cAAA).tw. (10648)
4 or/1-3 (41805)
5 Vascular Surgical Procedures/ (29038)
6 exp Endovascular Procedures/ (107137)
7 ((endovasc* or intravasc* or vascul*) adj4 (technique* or procedur* or surg*)).tw. (31184)
8 (endovasc* adj4 aneurysm* adj4 repair*).tw. (4143)
9 (endovasc* adj4 aneurysm* adj4 manag*).tw. (213)
10 (EVAR or EVRAR or fEVAR or f-EVAR or bEVAR or b-EVAR or BREVAR or BR-EVAR or CHEVAR or CH-EVAR or Co-EVAR or CoEVAR or Co-FEVAR or CoFEVAR).tw. (2883)
11 (endoprosthe* or endograft*).tw. (8684)
12 or/5-11 (161356)
13 Elective Surgical Procedures/ (11856)
14 (elect* adj4 (surg* or procedure*)).tw. (34432)
15 ((open or adjuvant) adj4 (repair* or surgical* or surger* or operat* or procedure*)).tw. (57457)
16 or/13-15 (97048)
17 4 and 12 and 16 (3769)
18 animals/ not humans/ (4448061)
19 17 not 18 (3760)
20 limit 19 to english language (3484)
21 Observational Studies as Topic/ (3159)
22 Observational Study/ (50595)

Medline strategy, searched 6 August 2018 Database: Ovid MEDLINE(R) <1946 to 3 August 2018 Search Strategy:	
23	Epidemiologic Studies/ (7730)
24	exp Case-Control Studies/ (931375)
25	exp Cohort Studies/ (1765000)
26	Cross-Sectional Studies/ (270993)
27	Controlled Before-After Studies/ (342)
28	Historically Controlled Study/ (140)
29	Interrupted Time Series Analysis/ (454)
30	Comparative Study.pt. (1804738)
31	case control$.tw. (99018)
32	case series.tw. (49229)
33	cohort*.tw. (398997)
34	(follow up adj (study or studies)).tw. (41985)
35	(observational adj (study or studies)).tw. (68131)
36	longitudinal.tw. (175572)
37	prospective.tw. (435773)
38	retrospective.tw. (369287)
39	cross sectional.tw. (232708)
40	Registry/ (75494)
41	regist*.tw. (277182)
42	or/21-41 (4211074)
43	20 and 4fcopy2 (2212)

Appendix B Study selection

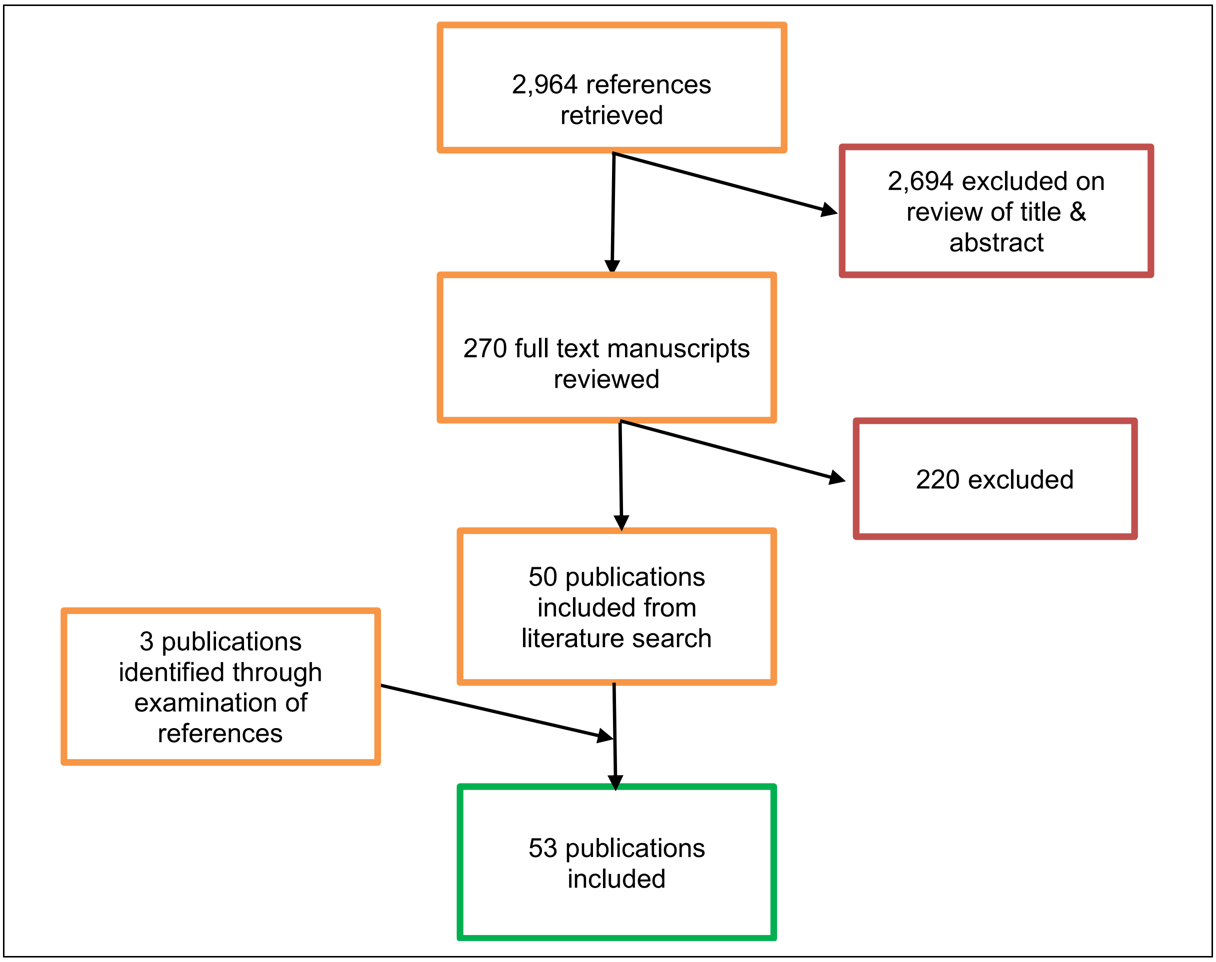


Figure 35: PRISMA flowchart depicting evidence identification

Appendix C Excluded studies

No.	Study	Reason for exclusion
1	Aarts F, van Sterkenburg , S , and Blankensteijn J D (2005) Endovascular aneurysm repair versus open aneurysm repair: comparison of treatment outcome and procedure-related reintervention rate. Annals of vascular surgery 19(5), 699-704	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
2	Adas, Z. A., Nypaver, T. J., Shepard, A. D. et al. (2018) Survival after abdominal aortic aneurysm repair is affected by socioeconomic status. Journal of Vascular Surgery	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
3	Aho, Pekka, Vikatmaa, Leena, Niemi-Murola, Leila et al. (2018) Simulation training streamlines the real-life performance in endovascular repair of ruptured abdominal aortic aneurysms. Journal of vascular surgery	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
4	Akbulut, M., Aksoy, E., Kara, I. et al. (2018) Quality of Life After Open Surgical versus Endovascular Repair of Abdominal Aortic Aneurysms. Brazilian journal of cardiovascular surgery 33(3): 265-270	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
5	Aljabri B, Al Wahaibi, K , Abner D, Mackenzie K S, Corriveau M M, Obrand D I, Meshefedjian G, and Steinmetz O K (2006) Patient-reported quality of life after abdominal aortic aneurysm surgery: A prospective comparison of endovascular and open repair. Journal of Vascular Surgery 44(6), 1182	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
6	Al-Jubouri Mustafa, Comerota Anthony J, Thakur Subhash, Aziz Faisal, Wanjiku Steven, Paolini David, Pigott John P, and Lurie Fedor (2013) Reintervention after EVAR and open surgical repair of AAA: a 15-year experience. Annals of surgery 258(4), 652-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
7	Alsac Jean-Marc, Houbballah Rabih, Francis Fady, Paraskevas Nikolaos, Coppin Thierry, Cerceau Olivier, Castier Yves, and Leseche Guy (2008) Impact of the introduction of endovascular aneurysm repair in high-risk patients on our practice of elective treatment of infrarenal abdominal aortic aneurysms. Annals of vascular surgery 22(6), 829-33	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
8	Altaf N, Abisi S, Yong Y, Saunders J H, Braithwaite B D, and MacSweeney S T (2013) Mid-term results of endovascular aortic aneurysm repair in the young. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 46(3), 315-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
9	Altieri, Maria S., Yang, Jie, Jones, Tyler et al. (2018) Incidence of Ventral Hernia Repair after Open Abdominal Aortic Aneurysm and Open Aortofemoral or Aortoiliac Bypass Surgery: An Analysis of 17,594 Patients in the State of New York. The American surgeon 84(8): 1388-1393	Study reports outcomes of OSR only – no comparison with EVAR

No.	Study	Reason for exclusion
10	Antonello Michele, Menegolo Mirko, Piazza Michele, Bonfante Luciana, Grego Franco, and Frigatti Paolo (2013) Outcomes of endovascular aneurysm repair on renal function compared with open repair. Journal of vascular surgery 58(4), 886-93	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
11	Antonello, M., Squizzato, F., Bassini, S. et al. (2019) Open repair versus endovascular treatment of complex aortoiliac lesions in low risk patients. Journal of Vascular Surgery 05: 05	Aortoiliac aneurysms
12	Antoniou, G. A., Rojoa, D., Antoniou, S. A. et al. (2019) Effect of Low Skeletal Muscle Mass on Post-operative Survival of Patients With Abdominal Aortic Aneurysm: A Prognostic Factor Review and Meta-Analysis of Time-to-Event Data. European Journal of Vascular & Endovascular Surgery 13: 13	Meta-analysis concerned with other risk factors that does not include a comparison of EVAR and OSR
13	Arko Frank R, Hill Bradley B, Olcott Cornelius, Harris E John, Jr, Fogarty Thomas J, and Zarins Christopher K (2002) Endovascular repair reduces early and late morbidity compared to open surgery for abdominal aortic aneurysm. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 9(6), 711-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
14	Arko Frank R, Hill Bradley B, Reeves Terrence R, Olcott Cornelius, Harris E John, Fogarty Thomas J, and Zarins Christopher K (2003) Early and late functional outcome assessments following endovascular and open aneurysm repair. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 10(1), 2-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
15	Arko Frank R, Lee W Anthony, Hill Bradley B, Olcott Cornelius th, Dalman Ronald L, Harris E John, Jr, Cipriano Paul, Fogarty Thomas J, and Zarins Christopher K (2002) Aneurysm-related death: primary endpoint analysis for comparison of open and endovascular repair. Journal of vascular surgery 36(2), 297-304	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
16	Barbey, S. M., Scali, S. T., Kubilis, P. et al. (2019) Interaction between frailty and sex on mortality after elective abdominal aortic aneurysm repair. Journal of Vascular Surgery.	Results presented separately for EVAR and OSR with no multivariable model
17	Barilla D, Sobocinski J, Stilo F, Maurel B, Spinelli F, and Haulon S (2014) Juxtarenal aortic aneurysm with hostile neck anatomy: midterm results of minilaparotomy versus f-EVAR. International angiology : a journal of the International Union of Angiology 33(5), 466-73	Study compares EVAR with minilaparotomy (which is not within the scope of this review)
18	Bath, Jonathan, Smith, Jamie B., Kruse, Robin L. et al. (2018) Cohort study of risk factors for 30-day readmission after abdominal aortic aneurysm repair. VASA. Zeitschrift fur Gefasskrankheiten: 1-11	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
19	Beck Adam W, Goodney Philip P, Nolan Brian W, Likosky Donald S, Eldrup-Jorgensen Jens, Cronenwett Jack L, Vascular Study Group of Northern New, and England (2009) Predicting 1-year mortality after elective abdominal aortic aneurysm repair. Journal of vascular surgery 49(4), 838-4	Study assesses risk factors. No comparisons were made between EVAR and OSR.
20	Becquemin J P, Bourriez A, D'Audiffret A, Zubilewicz T, Kobeiter H, Allaire E, Melliere D, and Desgranges P (2000) Mid-term results of endovascular versus open repair for abdominal aortic aneurysm in patients anatomically suitable for endovascular repair. European Journal of Vascular and Endovascular Surgery 19(6), 656-661	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
21	Beffa Lucas R, Petroski Gregory F, Kruse Robin L, and Vogel Todd R (2015) Functional status of nursing home residents before and after abdominal aortic aneurysm repair. Journal of vascular nursing : official publication of the Society for Peripheral Vascular Nursing 33(3), 106-11	No protocol-specified outcomes were reported.
22	Berchiolli, Raffaella, Tomei, Francesca, Marconi, Michele et al. (2019) Hand-assisted laparoscopic surgery versus endovascular repair in abdominal aortic aneurysm treatment. Journal of vascular surgery	Laparoscopic surgery is beyond scope
23	Bergqvist, D., Mani, K., Troeng, T. et al. (2018) Treatment of aortic aneurysms registered in Swedvasc: Development reflected in a national vascular registry with an almost 100% coverage. Gefasschirurgie : Zeitschrift fur vaskulare und endovaskulare Chirurgie : Organ der Deutschen und der Osterreichischen Gesellschaft fur Gefasschirurgie unter Mitarbeit der Schweizerischen Gesellschaft fur Gefasschirurgie 23(5): 340-345	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
24	Biancari F, Catania A, and D'Andrea V (2011) Elective endovascular vs. open repair for abdominal aortic aneurysm in patients aged 80 years and older: systematic review and meta-analysis. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 42(5), 571-6	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
25	Birch S E, Stary D R, and Scott A R (2000) Cost of endovascular versus open surgical repair of abdominal aortic aneurysms. The Australian and New Zealand journal of surgery 70(9), 660-6	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
26	Bosch J L, Beinfeld M T, Halpern E F, Lester J S, and Gazelle G S (2001) Endovascular versus open surgical elective repair of infrarenal abdominal aortic aneurysm: predictors of patient discharge destination. Radiology 220(3), 576-80	Study includes people who received EVAR before 1999.
27	Bosch Johanna L, Kaufman John A, Beinfeld Molly T, Adriaensen Miraude E. A. P. M, Brewster David C, and Gazelle G Scott (2002) Abdominal aortic aneurysms: cost-effectiveness of elective	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
	endovascular and open surgical repair. Radiology 225(2), 337-44	
28	Bostock, Ian C., Zarkowsky, Devin S., Hicks, Caitlin W. et al. (2018) Outcomes and Risk Factors Associated with Prolonged Intubation after EVAR. Annals of vascular surgery 50: 167-172	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
29	Boult Magg, Babidge Wendy, Anderson John, Denton Michael, Fitridge Robert, Harris John, Lawrence-Brown Michael, May James, Myerstt Kenneth, and Maddern Guy (2002) Australian audit for the endoluminal repair of abdominal aortic aneurysm: the first 12 months. ANZ journal of surgery 72(3), 190-5	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
30	Brazzelli, Miriam, Hernandez, Rodolfo, Sharma, Pawana et al. (2018) Contrast-enhanced ultrasound and/or colour duplex ultrasound for surveillance after endovascular abdominal aortic aneurysm repair: a systematic review and economic evaluation. Health technology assessment (Winchester, England) 22(72): 1-220	Not a comparison of EVAR and OSR
31	Brewster D C, Geller S C, Kaufman J A, Cambria R P, Gertler J P, LaMuraglia G M, Atamian S, and Abbott W M (1998) Initial experience with endovascular aneurysm repair: comparison of early results with outcome of conventional open repair. Journal of vascular surgery 27(6), 992-5	Insufficient details were provided to ascertain study methodology.
32	British Society for Endovascular, Therapy , the Global Collaborators on Advanced Stent-Graft Techniques f, and Registry (2012) Early results of fenestrated endovascular repair of juxtarenal aortic aneurysms in the United Kingdom. Circulation 125(22), 2707-15	This is a single arm study (case series) assessing of patients who underwent fenestrated EVAR.
33	Bruen Kevin J, Feezor Robert J, Daniels Michael J, Beck Adam W, and Lee W Anthony (2011) Endovascular chimney technique versus open repair of juxtarenal and suprarenal aneurysms. Journal of vascular surgery 53(4), 895-5	Undocumented, apparently subjective matching procedure
34	Budtz-Lilly J, Venermo M, Debus S, Behrendt C A, Altreuther M, Beiles B, Szeberin Z, Eldrup N, Danielsson G, Thomson I, Wigger P, Bjorck M, Loftus I, and Mani K (2017) Editor's Choice - Assessment of International Outcomes of Intact Abdominal Aortic Aneurysm Repair over 9 Years. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 54(1), 13-20	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
35	Burgers L T, Vahl A C, Severens J L, Wiersema A M, Cuypers P W. M, Verhagen H J. M, and Redekop W K (2016) Cost-effectiveness of Elective Endovascular Aneurysm Repair Versus Open Surgical Repair of Abdominal Aortic Aneurysms. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 52(1), 29-40	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
36	Canavati Rana, Millen Alistair, Brennan John, Fisher Robert K, McWilliams Richard G, Naik Jagjeeth B, and Vallabhaneni Srinivasa R (2013) Comparison of fenestrated endovascular and open repair of abdominal aortic aneurysms not suitable for standard endovascular repair. Journal of vascular surgery 57(2), 362-7	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
37	Cao Piergiorgio, Verzini Fabio, Parlani Gianbattista, Romano Lydia, De Rango , Paola , Pagliuca Valentino, and Iacono Gustavo (2004) Clinical effect of abdominal aortic aneurysm endografting: 7-year concurrent comparison with open repair. Journal of vascular surgery 40(5), 841-8	Study includes people who received EVAR before 1999.
38	Capoccia L, Marino M, Gazzetti M, Biello A, Sbarigia E, and Speziale F (2011) Octogenarians submitted to elective infrarenal abdominal aortic aneurysm repair: Can they currently be considered "high-risk" for open repair?. Italian Journal of Vascular and Endovascular Surgery 18(2), 57-63	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
39	Capoccia Laura, and Riambau Vicente (2015) Endovascular repair versus open repair for inflammatory abdominal aortic aneurysms. The Cochrane database of systematic reviews (4), CD010313	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
40	Chahwan Santiago, Comerota Anthony J, Pigott John P, Scheuermann Barry W, Burrow Julia, and Wojnarowski Dennis (2007) Elective treatment of abdominal aortic aneurysm with endovascular or open repair: the first decade. Journal of vascular surgery 45(2), 258-262	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
41	Chambers D, Epstein D, Walker S, Fayter D, Paton F, Wright K, Michaels J, Thomas S, Sculpher M, and Woolacott N (2009) Endovascular stents for abdominal aortic aneurysms: a systematic review and economic model. Health technology assessment (Winchester, and England) 13(48), 1-iii	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
42	Chan Y C, Morales J P, Gulamhuseinwala N, Sabharwal T, Carmichael M, Thomas S, Carrell T W. G, Reidy J F, and Taylor P R (2007) Large infra-renal abdominal aortic aneurysms: endovascular vs. open repair--single centre experience. International journal of clinical practice 61(3), 373-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
43	Chandra Venita, Trang Karen, Virgin-Downey Whitt, Dalman Ronald L, and Mell Matthew W (2018) Long-term outcomes after repair of symptomatic abdominal aortic aneurysms. Journal of vascular surgery.	Study compares outcomes of symptomatic and asymptomatic elective AAA repair.
44	Charbonneau, P., Hongku, K., Herman, C. R. et al. (2019) Long-term survival after endovascular and open repair in patients with anatomy outside instructions for use criteria for endovascular aneurysm repair. Journal of Vascular Surgery 21: 21	No protocol-specified outcomes in adjusted analyses: only Cox regression over entire follow-up (including perioperative period)

No.	Study	Reason for exclusion
45	Chinsakchai K, Phetpoonpipat W, Ruangsetakit C, Wongwanit C, Mutirangura P, Sermsathanasawadi N, Hongku K, and Hahtapornsawan S (2017) Outcomes of asymptomatic abdominal aortic aneurysm as compared between open aortic repair and endovascular aneurysm repair. Journal of the Medical Association of Thailand 100(3 Supplement 2), S162-S169	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
46	Chinsakchai, K., Prapassaro, T., Salisatkorn, W. et al. (2018) Outcomes of Open Repair, Fenestrated Stent Grafting, and Chimney Grafting in Juxtarenal Abdominal Aortic Aneurysm: Is It Time for a Randomized Trial?. Annals of Vascular Surgery	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
47	Chisci Emiliano, Kristmundsson Thorarinn, de Donato , Gianmarco , Resch Timothy, Setacci Francesco, Sonesson Bjorn, Setacci Carlo, and Malina Martin (2009) The AAA with a challenging neck: outcome of open versus endovascular repair with standard and fenestrated stent-grafts. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 16(2), 137-46	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
48	Choi K, Han Y, Ko G Y, Cho Y P, and Kwon T W (2018) Early and Late Outcomes of Endovascular Aortic Aneurysm Repair versus Open Surgical Repair of an Abdominal Aortic Aneurysm: A Single-Center Study. Annals of Vascular Surgery 51, 187-191	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
49	Cochennec Frederic, Marzelle Jean, Allaire Eric, Desgranges Pascal, and Becquemin Jean-Pierre (2010) Open vs endovascular repair of abdominal aortic aneurysm involving the iliac bifurcation. Journal of vascular surgery 51(6), 1360-6	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
50	Cohnert T U, Oelert F, Wahlers T, Gohrbandt B, Chavan A, Farber A, Galanski M, and Haverich A (2000) Matched-pair analysis of conventional versus endoluminal AAA treatment outcomes during the initial phase of an aortic endografting program. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 7(2), 94-100	Insufficient details were provided to ascertain study methodology.
51	Coscas Raphael, Dennery Marc, Javerliat Isabelle, Di Centa , Isabelle , Cudennec Tristan, Teillet Laurent, Goeau-Brissonniere Olivier, and Coggia Marc (2014) Laparoscopy versus EVAR for the treatment of abdominal aortic aneurysms in the octogenarian. Annals of vascular surgery 28(7), 1634-41	Outcome measure assessed was the principal criterion of judgment: the composite rate of mortality and severe systemic complication. This is out of scope of the guideline.
52	Curci John A, Fillinger Mark F, Naslund Thomas C, Rubin Brian G, Excluder Bifurcated Endoprosthesis, and Investigators (2007) Clinical trial results of a modified gore excluder endograft: comparison with open repair and original device design. Annals of vascular surgery 21(3), 328-38	Multivariate analysis does not provide any information about comparative efficacy of EVAR vs. OSR
53	Dakour Aridi, Hanaa N, Locham Satinderjit, Nejim Besma, Ghajar Nasr S, Alshaikh Husain, and Malas Mahmoud B (2018) Indications, risk factors,	This study specifically looks at "failure to rescue", which is defined as: "30-day mortality rate in patients with at least

No.	Study	Reason for exclusion
	and outcomes of 30-day readmission after infrarenal abdominal aneurysm repair. Journal of vascular surgery 67(3), 747-758.e7	one complication". This is out of scope of this review.
54	Dakour-Aridi, Hanaa, Nejim, Besma, Locham, Satinderjit et al. (2019) Anemia and postoperative outcomes after open and endovascular repair of intact abdominal aortic aneurysms. Journal of vascular surgery	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
55	Davis, F. M., Jerzal, E., Albright, J. et al. (2019) Variation in the elective management of small abdominal aortic aneurysms and physician practice patterns. Journal of Vascular Surgery 02: 02	Propensity matching for small and large AAA only
56	de Bruin , J L, Groenwold R H. H, Baas A F, Brownrigg J R, Prinssen M, Grobbee D E, Blankensteijn J D, and Group Dream Study (2016) Quality of life from a randomized trial of open and endovascular repair for abdominal aortic aneurysm. The British journal of surgery 103(8), 995-1002	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
57	de Bruin , Jorg Lucas, Karthikesalingam Alan, Holt Peter J, Prinssen Monique, Thompson Matt M, Blankensteijn Jan D, Dutch Randomised Endovascular Aneurysm Management Study, and Group (2016) Predicting reinterventions after open and endovascular aneurysm repair using the St George's Vascular Institute score. Journal of vascular surgery 63(6), 1428-1433.e1	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
58	de Jesus-Silva, Seleno Glauber, de Oliveira, Victor Rodrigues, de Moraes-Silva, Melissa Andreia et al. (2018) Risk factors and short and medium-term survival after open and endovascular repair of abdominal aortic aneurysms. Jornal vascular brasileiro 17(3): 201-207	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
59	De Martino , Randall R, Brooke Benjamin S, Robinson William, Schanzer Andres, Indes Jeffrey E, Wallaert Jessica B, Nolan Brian W, Cronenwett Jack L, and Goodney Philip P (2013) Designation as "unfit for open repair" is associated with poor outcomes after endovascular aortic aneurysm repair. Circulation. Cardiovascular quality and outcomes 6(5), 575-81	No comparisons were made between EVAR and OSR.
60	de Virgilio , C , Bui H, Donayre C, Ephraim L, Lewis R J, Elbassir M, Stabile B E, and White R (1999) Endovascular vs open abdominal aortic aneurysm repair: a comparison of cardiac morbidity and mortality. Archives of surgery (Chicago, and Ill. : 1960) 134(9), 947-1	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
61	Deery S E, Lancaster R T, Gubala A M, O'Donnell T F. X, Kwolek C J, Conrad M F, Cambria R P, and Patel V I (2018) Early Experience with Fenestrated Endovascular Compared to Open Repair of Complex Abdominal Aortic Aneurysms in a High-Volume Open Aortic Center. Annals of Vascular Surgery 48, 151-158	Extremely small study (n=18) in which multivariable regression would not have been appropriate. Furthermore, it is unclear what time period mortality was measured in.

No.	Study	Reason for exclusion
62	Desai Mital, Choke Edward, Sayers Robert D, Nath Mintu, and Bown Matthew J (2016) Sex-related trends in mortality after elective abdominal aortic aneurysm surgery between 2002 and 2013 at National Health Service hospitals in England: less benefit for women compared with men. European heart journal 37(46), 3452-3460	No comparisons were made between EVAR and OSR.
63	Dias Paulo, Sampaio Sergio, Rocha E Silva, Augusto , Roncon de Albuquerque, and R (2010) The need for reintervention is not higher after EVAR: an eight years single center experience. Revista portuguesa de cirurgia cardio-toracica e vascular : orgao oficial da Sociedade Portuguesa de Cirurgia Cardio-Toracica e Vascular 17(4), 245-50	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
64	Diehm Nicolas, Tsoukas Athanassios I, Katzen Barry T, Benenati James F, Baum Samuel, Pena Constantino, and Dick Florian (2008) Matched-pair analysis of endovascular versus open surgical repair of abdominal aortic aneurysms in young patients at low risk. Journal of vascular and interventional radiology : JVIR 19(5), 645-51	Undocumented, apparently subjective matching procedure
65	Donald, G. W., Ghaffarian, A. A., Isaac, F. et al. (2018) Preoperative frailty assessment predicts loss of independence after vascular surgery. Journal of Vascular Surgery 68(5): 1382-1389	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
66	Donas Konstantinos P, Eisenack Markus, Panuccio Giuseppe, Austermann Martin, Osada Nani, and Torsello Giovanni (2012) The role of open and endovascular treatment with fenestrated and chimney endografts for patients with juxtarenal aortic aneurysms. Journal of vascular surgery 56(2), 285-90	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
67	Duque Santos, A , Reyes Valdivia, A , Romero Lozano, M A, Aracil Sanus, E , Ocana Guaita, J , and Gandarias C (2018) Outcomes of open and endovascular repair of inflammatory abdominal aortic aneurysms. Vascular 26(2), 203-208	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
68	Endicott Kendal M, Emerson Dominic, Amdur Richard, and Macsata Robyn (2017) Functional status as a predictor of outcomes in open and endovascular abdominal aortic aneurysm repair. Journal of vascular surgery 65(1), 40-45	Study pools outcomes of EVAR and OSR and examines risk factors. No comparisons were made between treatments.
69	Esce Antoinette, Medhekar Ankit, Fleming Fergal, Noyes Katia, Glocker Roan, Ellis Jennifer, Raman Kathleen, Stoner Michael, and Doyle Adam (2018) Superior 3-Year Value of Open and Endovascular Repair of Abdominal Aortic Aneurysm with High-Volume Providers. Annals of vascular surgery 46, 17-29	Health economic analysis that only considers outcomes of people who survived after the perioperative period. This is not within the scope of this review.
70	Eslami MH, Rybin D, Doros G (2015) Comparison of a Vascular Study Group of New England risk prediction model with established risk prediction models of in-hospital mortality after elective abdominal aortic aneurysm repair. J Vasc Surg. 62(5):1125-33.e2.	Study pools the outcomes of EVAR and open repair and looks at risk factors for the whole group.

No.	Study	Reason for exclusion
71	Fotis Theofanis, Tsoumakidou Georgia, Katostaras Theophanis, Kalokairinou Athina, Konstantinou Evangelos, Kiki Vozides, and Perdikides Theodosios (2008) Cost and effectiveness comparison of endovascular aneurysm repair versus open surgical repair of abdominal aortic aneurysm: a single-center experience. Journal of vascular nursing : official publication of the Society for Peripheral Vascular Nursing 26(1), 15-21	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
72	Freyrie A, Gargiulo M, Gallitto E, Faggioli G L, Testi G, Giovanetti F, and Stella A (2012) Abdominal aortic aneurysms with short proximal neck: comparison between standard endograft and open repair. The Journal of cardiovascular surgery 53(5), 617-23	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
73	Garcia-Madrid C, Josa M, Riambau V, Mestres C A, Muntana J, and Mulet J (2004) Endovascular versus open surgical repair of abdominal aortic aneurysm: a comparison of early and intermediate results in patients suitable for both techniques. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 28(4), 365-72	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
74	Gattuso, R., Picone, V., Belli, C. et al. (2019) Treatment of abdominal aortic aneurysms and coexisting cancer: Endovascular versus traditional approach. Italian Journal of Vascular and Endovascular Surgery 26(2): 76-80	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
75	Giles Kristina A, Schermerhorn Marc L, O'Malley A James, Cotterill Philip, Jhaveri Ami, Pomposelli Frank B, and Landon Bruce E (2009) Risk prediction for perioperative mortality of endovascular vs open repair of abdominal aortic aneurysms using the Medicare population. Journal of vascular surgery 50(2), 256-62	More recent publications from the same study group, assessing a similar population (derived from the same sampling frame) were available.
76	Gnus Jan, Ferenc Stanislaw, Dziewiszek Malgorzata, Rusiecki Leslaw, and Witkiewicz Wojciech (2015) Comparison of Endovascular Aneurysm Repair with Open Repair in Patients with Abdominal Aortic Aneurysm in Our Own Material in Years 2002-2011. Advances in clinical and experimental medicine : official organ Wroclaw Medical University 24(3), 475-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
77	Goodney Philip P, Tavris Dale, Lucas F Lee, Gross Thomas, Fisher Elliott S, and Finlayson Samuel R. G (2010) Causes of late mortality after endovascular and open surgical repair of infrarenal abdominal aortic aneurysms. Journal of vascular surgery 51(6), 1340-1347.e1	Insufficient details were provided to ascertain study methodology.
78	Goodyear Stephen J, Yow Heng, Saedon Mahmud, Shakespeare Joanna, Hill Christopher E, Watson Duncan, Marshall Colette, Mahmood Asif, Higman Daniel, and Imray Christopher He (2013) Risk stratification by pre-operative cardiopulmonary exercise testing improves outcomes following elective abdominal aortic	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
	aneurysm surgery: a cohort study. Perioperative medicine (London, and England) 2(1), 10	
79	Greenberg Roy K, Chuter Timothy A. M, Sternbergh W Charles, 3rd, Fearnot Neal E, and Zenith Investigators (2004) Zenith AAA endovascular graft: intermediate-term results of the US multicenter trial. Journal of vascular surgery 39(6), 1209-18	Although authors stated that regression was performed, they did not report data in a reasonable format to inform this NICE review.
80	Gupta, A. K., Alshaikh, H. N., Dakour-Aridi, H. et al. (2019) Real-world cost analysis of endovascular repair versus open repair in patients with nonruptured abdominal aortic aneurysms. Journal of Vascular Surgery 03: 03	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
81	Gurakar, M., Locham, S., Alshaikh, H. N. et al. (2019) Risk factors and outcomes for bowel ischemia after open and endovascular abdominal aortic aneurysm repair. Journal of Vascular Surgery 05: 05	No protocol-specified outcomes were reported.
82	Health Quality, and Ontario (2009) Fenestrated endovascular grafts for the repair of juxtarenal aortic aneurysms: an evidence-based analysis. Ontario health technology assessment series 9(4), 1-51	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
83	Hicks Caitlin W, Obeid Tammam, Arhuidese Isibor, Qazi Umair, and Malas Mahmoud B (2016) Abdominal aortic aneurysm repair in octogenarians is associated with higher mortality compared with nonoctogenarians. Journal of vascular surgery 64(4), 956-965.e1	Study includes people with ruptured AAAs in its analysis. Ruptured AAAs are not within the scope of this review.
84	Hill Bradley B, Wolf Yehuda G, Lee W Anthony, Arko Frank R, Olcott Cornelius th, Schubart Peter J, Dalman Ronald L, Harris E John, Fogarty Thomas J, and Zarins Christopher K (2002) Open versus endovascular AAA repair in patients who are morphological candidates for endovascular treatment. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 9(3), 255-61	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
85	Hinterseher Irene, Kuffner Herold, Koch Rainer, Gabel Gabor, Saeger Hans D, and Smelser Diane (2012) Comparison of survival rates for abdominal aortic aneurysm treatment methods. World journal of surgery 36(4), 917-22	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
86	Ho P, Yiu W K, Cheung G C. Y, Cheng S W. K, Ting A C. W, and Poon J T. C (2006) Systematic review of clinical trials comparing open and endovascular treatment of abdominal aortic aneurysm. Surgical Practice 10(1), 24-37	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
87	Hoel Andrew W, Faerber Adrienne E, Moore Kayla O, Ramkumar Niveditta, Brooke Benjamin S, Scali Salvatore T, Sedrakyan Art, and Goodney Philip P (2017) A pilot study for long-term outcome assessment after aortic aneurysm repair using Vascular Quality Initiative data matched to Medicare claims. Journal of vascular surgery 66(3), 751-759.e1	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
88	Hoshina K, Hosaka A, Takayama T, Kato M, Ohkubo N, Okamoto H, Shigematsu K, and Miyata T (2012) Outcomes after open surgery and endovascular aneurysm repair for abdominal aortic aneurysm in patients with massive neck atheroma. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 43(3), 257-61	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
89	Hsieh, W. C., Kan, C. D., Hsieh, C. C. et al. (2019) Improved outcomes from endovascular aortic repair in younger patients: Towards improved risk stratification. Vascular: 1708538119843420	Meta-analysis of observational data including unadjusted
90	Hughes Kakra, Abdulrahman Hamdi, Prendergast Tahira, Rose David A, Ongu'ti Sharon, Tran Daniel, Cornwell Edward E, 3rd , Obisesan Thomas, and Amankwah Kwame S (2015) Abdominal aortic aneurysm repair in nonagenarians. Annals of vascular surgery 29(2), 183-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
91	Hwang Deokbi, Park Sujin, Kim Hyung-Kee, Lee Jong-Min, and Huh Seung (2017) Reintervention Rate after Open Surgery and Endovascular Repair for Nonruptured Abdominal Aortic Aneurysms. Annals of vascular surgery 43, 134-143	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
92	Hynes Niamh, and Sultan Sherif (2007) A prospective clinical, economic, and quality-of-life analysis comparing endovascular aneurysm repair (EVAR), open repair, and best medical treatment in high-risk patients with abdominal aortic aneurysms suitable for EVAR: the Irish patient trial. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 14(6), 763-76	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
93	Iannelli G, Monaco M, Di Tommaso , L , Piscione F, Stassano P, Mainenti P P, Laurino S, and Spampinato N (2005) Endovascular vs. open surgery of abdominal aortic aneurysm in high-risk patients: a single center experience. The Thoracic and cardiovascular surgeon 53(5), 291-4	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
94	Joh Jin Hyun, Park Yun-Young, Cho Sung-Shin, and Park Ho-Chul (2016) National trends for open and endovascular repair of aneurysms in Korea: 2004-2013. Experimental and therapeutic medicine 12(5), 3333-3338	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
95	Jonker Frederik H. W, Schlosser Felix J. V, Dewan Michael, Huddle Matthew, Sergi Michael, Indes Jeffrey E, Dardik Alan, and Muhs Bart E (2010) Abdominal aortic aneurysm repair in obese patients: improved outcome after endovascular treatment compared with open surgery. Vascular and endovascular surgery 44(2), 105-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
96	Joo, Hyun-Chel, Lee, Seung-Hyun, Chang, Byung-Chul et al. (2019) Late open conversion after endovascular abdominal aortic repair: a 20-year experience. The Journal of cardiovascular surgery 60(1): 73-80	Study reports outcomes of EVAR only – no comparison with OSR

No.	Study	Reason for exclusion
97	Jordan William D, Alcocer Francisco, Wirthlin Douglas J, Westfall Andrew O, and Whitley David (2003) Abdominal aortic aneurysms in "high-risk" surgical patients: comparison of open and endovascular repair. Annals of surgery 237(5), 623-30	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
98	Kalra Kanika, and Arya Shipra (2017) A comparative review of open and endovascular abdominal aortic aneurysm repairs in the national operative quality improvement database. Surgery 162(5), 979-988	Narrative review of studies.
99	Karkkainen, Jussi M., Sandri, Giuliano de A., Tenorio, Emanuel R. et al. (2018) Prospective assessment of health-related quality of life after endovascular repair of pararenal and thoracoabdominal aortic aneurysms using fenestrated-branched endografts. Journal of vascular surgery	Study reports outcomes of EVAR only – no comparison with OSR
100	Karmy-Jones Riyad, Bloch Robert, and Nicholls Stephen (2009) A comparison of endovascular repair versus open repair of abdominal aortic aneurysms in a community setting. Innovations (Philadelphia, and Pa.) 4(5), 261-4	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
101	Karthikesalingam Alan, Holt Peter J. E, Patterson Benjamin O, Vidal-Diez Alberto, Sollazzo Giuseppe, Poloniecki Jan D, Hinchliffe Robert J, and Thompson Matthew M (2013) Elective open suprarenal aneurysm repair in England from 2000 to 2010 an observational study of hospital episode statistics. PloS one 8(5), e64163	No comparisons were made between EVAR and OSR.
102	Karthikesalingam, A., Grima, M. J., Holt, P. J. et al. (2018) Comparative analysis of the outcomes of elective abdominal aortic aneurysm repair in England and Sweden. The British journal of surgery 105(5): 520-528	No protocol-specified outcomes in adjusted analyses (90-day mortality and single Cox model for 0–5yr)
103	Kato Takayoshi, Tamaki Mototsugu, Tsunekawa Tomohiro, Motoji Yusuke, Hirakawa Akihiro, Okawa Yasuhide, and Tomita Shinji (2017) Health-related quality of life prospectively evaluated by the 8-item short form after endovascular repair versus open surgery for abdominal aortic aneurysms. Heart and vessels 32(8), 960-968	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
104	Katsargyris A, Oikonomou K, Klonaris C, Topel I, and Verhoeven E L. G (2013) Comparison of outcomes with open, fenestrated, and chimney graft repair of juxtarenal aneurysms: Are we ready for a paradigm shift?. Journal of Endovascular Therapy 20(2), 159-169	Systematic review: individual studies were assessed to ascertain eligibility for inclusion in this NICE review.
105	Kayssi Ahmed, DeBord Smith, Ann , Roche-Nagle Graham, and Nguyen Louis L (2015) Health-related quality-of-life outcomes after open versus endovascular abdominal aortic aneurysm repair. Journal of vascular surgery 62(2), 491-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
106	Kennedy N A, Flynn L M, Berg R M, Lorelli D R, Rama K, and Rizk Y (2010) The evaluation of morbidity and mortality in abdominal aortic	No propensity score matching, multivariable regression, or other

No.	Study	Reason for exclusion
	aneurysm repair patients as related to body mass index. American Journal of Surgery 199(3), 369-371	statistical methods for adjusting for differences in case mix were performed.
107	Kisis Kaspars, Krievins Dainis, Naskovica Karina, Gedins Marcis, Savlovskis Janis, Ezite Natalija, Lietuvietis Edvins, and Zarins Kristaps (2012) Quality of life after endovascular abdominal aortic aneurysm repair: nellix sac-anchoring endoprosthesis versus open surgery. Medicina (Kaunas, and Lithuania) 48(6), 286-91	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
108	Komshian, S., Farber, A., Patel, V. I. et al. (2019) Patients with end-stage renal disease have poor outcomes after endovascular abdominal aortic aneurysm repair. Journal of Vascular Surgery 69(2): 405-413	EVAR only
109	Kontopodis Nikolaos, Antoniou Stavros A, Georgakarakos Efstratios, and Ioannou Christos V (2015) Endovascular vs Open Aneurysm Repair in the Young: Systematic Review and Meta-analysis. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 22(6), 897-904	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
110	Kontopodis, Nikolaos, Tavlas, Emmanouil, Georgakarakos, Efstratios et al. (2018) Has Anatomic Complexity of Abdominal Aortic Aneurysms Undergoing Open Surgical Repair Changed after the Introduction of Endovascular Treatment? Systematic Review and Meta-analysis of Comparative Studies. Annals of vascular surgery 52: 292-301	Study reports outcomes of OSR only – no comparison with EVAR
111	Krishnamoorthi H, Jeon-Slaughter H, Wall A, Banerjee S, Ramanan B, Timaran C, Modrall J G, and Tsai S (2018) Rate of secondary intervention after open versus endovascular abdominal aortic aneurysm repair. Journal of Surgical Research 232, 99-106	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
112	Kulig Piotr, Lewandowski Krzysztof, Ziaja Damian, Zaniewski Maciej, and Kulig Jan (2016) Endovascular Aneurysm Repair or Open Aneurysm Repair for the Treatment of Abdominal Aortic Aneurysm - The Latest Update. Polski przeglad chirurgiczny 88(3), 166-74	Narrative review discussing various observational studies.
113	Langenberg Jasper C. M, Kluytmans Jan A. J. W, de Groot , Hans G W, Ho Gwan H, Veen Eelco J, Buimer M G, van der Laan , and Lijckle (2018) Surgical Site and Graft Infections in Endovascular and Open Abdominal Aortic Aneurysm Surgery. Surgical infections 19(4), 424-429	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
114	Lareyre, F., Carboni, J., Chikande, J. et al. (2019) Association of Platelet to Lymphocyte Ratio and Risk of 30-Day Postoperative Complications in Patients Undergoing Abdominal Aortic Surgical Repair. Vascular and Endovascular Surgery 53(1): 5-11	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
115	Lederle F A, Stroupe K T, Open Versus Endovascular Repair Veterans Affairs Cooperative	No propensity score matching, multivariable regression, or other

No.	Study	Reason for exclusion
	S, and Group (2012) Cost-effectiveness at two years in the VA Open Versus Endovascular Repair Trial. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 44(6), 543-8	statistical methods for adjusting for differences in case mix were performed.
116	Lederle Frank A (2004) Abdominal aortic aneurysm--open versus endovascular repair. The New England journal of medicine 351(16), 1677-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
117	Lederle Frank A, Freischlag Julie A, Kyriakides Tassos C, Matsumura Jon S, Padberg Frank T, Jr , Kohler Ted R, Kougias Panagiotis, Jean-Claude Jessie M, Cikrit Dolores F, Swanson Kathleen M, and Group Over Veterans Affairs Cooperative Study (2012) Long-term comparison of endovascular and open repair of abdominal aortic aneurysm. The New England journal of medicine 367(21), 1988-97	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
118	Lee Hong-Gi, Clair Daniel G, and Ouriel Kenneth (2013) Ten-year comparison of all-cause mortality after endovascular or open repair of abdominal aortic aneurysms: a propensity score analysis. World journal of surgery 37(3), 680-7	Study includes people who received EVAR before 1999.
119	Lee Kevin, Tang Elaine, Dubois Luc, Power Adam H, DeRose Guy, and Forbes Thomas L (2015) Durability and survival are similar after elective endovascular and open repair of abdominal aortic aneurysms in younger patients. Journal of vascular surgery 61(3), 636-41	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
120	Lee S Y, Peacock M R, Farber A, Shah N K, Eslami M H, Kalish J A, Rybin D, Komshian S, and Siracuse J J (2017) Perioperative Infections after Open Abdominal Aortic Aneurysm Repair Lead to Increased Risk of Subsequent Complications. Annals of Vascular Surgery 44, 203-210	Study only assesses outcomes of people who underwent OSR.
121	Li, B., Khan, S., Salata, K. et al. (2019) A systematic review and meta-analysis of the long-term outcomes of endovascular versus open repair of abdominal aortic aneurysm. Journal of Vascular Surgery 27: 27	Meta-analysis of observational data including unadjusted
122	Lijftogt N, Vahl A C, Wilschut E D, Elsman B H. P, Amodio S, van Zwet , E W, Leijdekkers V J, Wouters M W. J. M, Hamming J F, Dutch Society of Vascular Surgery, the Steering Committee of the Dutch Surgical Aneurysm Audit, the Dutch Institute for Clinical, and Auditing (2017) Adjusted Hospital Outcomes of Abdominal Aortic Aneurysm Surgery Reported in the Dutch Surgical Aneurysm Audit. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 53(4), 520-532	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
123	Lilja F, Mani K, and Wanhainen A (2017) Editor's Choice - Trend-break in Abdominal Aortic Aneurysm Repair With Decreasing Surgical Workload. European journal of vascular and	Study about how AAA management has changed in Sweden over different time periods. No comparisons were performed between EVAR and OSR.

No.	Study	Reason for exclusion
	endovascular surgery : the official journal of the European Society for Vascular Surgery 53(6), 811-819	
124	Limet R, and Creemers E (2000) Comparison between open and closed repair for abdominal aortic aneurysms: a word of caution. Acta chirurgica Belgica 100(1), 12-5	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
125	Lindstrom, I., Khan, N., Vanttinen, T. et al. (2019) Psoas Muscle Area and Quality Are Independent Predictors of Survival in Patients Treated for Abdominal Aortic Aneurysms. Annals of Vascular Surgery 56: 183-193.e3	Not a comparison of EVAR and OSR
126	Locham, Satinderjit, Dakour-Aridi, Hanaa, Nejim, Besma et al. (2018) Outcomes and cost of open versus endovascular repair of intact thoracoabdominal aortic aneurysm. Journal of vascular surgery 68(4): 948-955.e1	Thoracoabdominal aortic aneurysms (with no subgroup for type IV cases)
127	Lottman Patricia E. M, Laheij Robert J. F, Cuypers Philip W. M, Bender Mart, and Buth Jacob (2004) Health-related quality of life outcomes following elective open or endovascular AAA repair: a randomized controlled trial. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 11(3), 323-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
128	Maeda H, Umezawa H, Hattori T, Nakamura T, Umeda T, Kobayashi H, Kawachi H, Iida A, and Shiono M (2013) Early and late outcomes of inflammatory abdominal aortic aneurysms: comparison with the outcomes after open surgical and endovascular aneurysm repair in literature reviews. International angiology : a journal of the International Union of Angiology 32(1), 67-73	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
129	Maeda Koji, Ohki Takao, Kanaoka Yuji, Baba Takeshi, Kaneko Kenjirou, and Shukuzawa Kota (2017) Comparison between Open and Endovascular Repair for the Treatment of Juxtarenal Abdominal Aortic Aneurysms: A Single-Center Experience with Midterm Results. Annals of vascular surgery 41, 96-104	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
130	Majd P, Ahmad W, Galas N, and Brunkwall J S (2018) Patients Older Than 80 Years Can Reach Their Normal Life Expectancy After Abdominal Aortic Aneurysm Repair: A Comparison Between Endovascular Aneurysm Repair and Open Surgery. Journal of Endovascular Therapy 25(2), 247-251	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
131	Majd Payman, Ahmad Wael, Becker Ingrid, and Brunkwall Jan Sigge (2017) Ten-Year Single-Center Results of Abdominal Aortic Aneurysm Treatment: Endovascular versus Open Repair. Annals of vascular surgery 44, 113-118	Study includes people who received EVAR before 1999.
132	Makaroun Michel S, Chaikof Elliot, Naslund Thomas, and Matsumura Jon S (2002) Efficacy of a bifurcated endograft versus open repair of	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
	abdominal aortic aneurysms: a reappraisal. Journal of vascular surgery 35(2), 203-10	
133	Malas Mahmoud B, Jordan William D, Cooper Michol A, Qazi Umair, Beck Adam W, Belkin Michael, Robinson William, and Fillinger Mark (2015) Performance of the Aorfix endograft in severely angulated proximal necks in the PYTHAGORAS United States clinical trial. Journal of vascular surgery 62(5), 1108-17	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
134	Mani Kevin, Bjorck Martin, Lundkvist Jonas, and Wanhainen Anders (2008) Similar cost for elective open and endovascular AAA repair in a population-based setting. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 15(1), 1-11	Health economic analysis that explores risk factors that lead to increased treatment costs.
135	Manis George, Feuerman Martin, and Hines George L (2006) Open aneurysm repair in elderly patients not candidates for endovascular repair (EVAR): Comparison with patients undergoing EVAR or preferential open repair. Vascular and endovascular surgery 40(2), 95-101	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
136	Manunga Jesse, Sullivan Timothy, Garberich Ross, Alden Peter, Alexander Jason, Skeik Nedaa, Titus Jessica, Stephenson Elliott, and Cragg Andrew (2018) Single-center experience with complex abdominal aortic aneurysms treated by open or endovascular repair using fenestrated/branched endografts. Journal of vascular surgery 68(2), 337-347	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
137	Markar, Sheraz R., Vidal-Diez, Alberto, Sounderajah, Viknesh et al. (2018) A population-based cohort study examining the risk of abdominal cancer after endovascular abdominal aortic aneurysm repair. Journal of vascular surgery	Study conflates elective and emergency cases
138	Marques De Marino, Pablo, Martinez Lopez, Isaac, Cernuda Artero, Inaki et al. (2018) Renal function after abdominal aortic aneurysm repair in patients with baseline chronic renal insufficiency: open vs. endovascular repair. International angiology : a journal of the International Union of Angiology 37(5): 377-383	No protocol-specified outcomes in adjusted analyses (single Cox model for 0–3.5yr)
139	Martinez, Rennier, Gaffney, Lukas, Parreco, Joshua et al. (2018) Nationally Representative Readmission Factors Associated with Endovascular versus Open Repair of Abdominal Aortic Aneurysm. Annals of vascular surgery 53: 105-116	No protocol-specified outcomes
140	Matsumura JS, Brewster DC, Makaroun MS, et al. (2003) A multicenter controlled clinical trial of open versus endovascular treatment of abdominal aortic aneurysm. Journal of Vascular Surgery. 37(2):262-271	No recruitment dates stated; likely to include participants who underwent EVAR before 1999.
141	May J, White G H, Waugh R, Ly C N, Stephen M S, Jones M A, and Harris J P (2001) Improved survival after endoluminal repair with second-generation prostheses compared with open repair	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
	in the treatment of abdominal aortic aneurysms: a 5-year concurrent comparison using life table method. Journal of vascular surgery 33(2 Suppl), S21-6	
142	Mazzaccaro Daniela, Nano Giovanni, Settembrini Alberto M, Carmo Michele, Dallatana Raffaello, Salvati Simone, Malacrida Giovanni, and Settembrini Piergiorgio G (2017) Open and endovascular elective treatment of abdominal aortic aneurysms: a real-world experience. Surgery today 47(11), 1347-1355	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
143	Mehta Manish, Byrne W John, Robinson Handel, Roddy Sean P, Paty Philip S. K, Kreienberg Paul B, Feustel Paul, Darling R Clement, and 3rd (2012) Women derive less benefit from elective endovascular aneurysm repair than men. Journal of vascular surgery 55(4), 906-13	Study assessed differences between men and women, and went on to assess risk factors for mortality in women who underwent EVAR.
144	Mehta Manish, Roddy Sean P, Darling R Clement, 3rd, Ozsvath Kathleen J, Kreienberg Paul B, Paty Philip S. K, Chang Benjamin B, and Shah Dhiraj M (2005) Infrarenal abdominal aortic aneurysm repair via endovascular versus open retroperitoneal approach. Annals of vascular surgery 19(3), 374-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
145	Menezes Fabio Husemann, Ferrarezi Barbara, Souza Moises Amancio de, Cosme Susyanne Lavor, and Molinari Giovani Jose Dal Poggetto (2016) Results of Open and Endovascular Abdominal Aortic Aneurysm Repair According to the E-PASS Score. Brazilian journal of cardiovascular surgery 31(1), 22-30	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
146	Michel Morgane, Becquemin Jean-Pierre, Marzelle Jean, Quelen Celine, Durand-Zaleski Isabelle, and participants Window Trial (2018) Editor's Choice - A Study of the Cost-effectiveness of Fenestrated/branched EVAR Compared with Open Surgery for Patients with Complex Aortic Aneurysms at 2 Years. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 56(1), 15-21	Secondary publication of included study (Michel et al., 2016), but outcome measures which are not in line with the scope of this review.
147	Miranda S P, Miranda P C, Volpato M G, Folino M C, Kambara A M, Rossi F H, and Izukawa N M (2014) Open vs. Endovascular repair of abdominal aortic aneurysm: A comparative analysis. Jornal Vascular Brasileiro 13(4), 276-284	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
148	Mistry P P, Becker P, Van Marle , and J (2007) A prospective comparison of secondary interventions and mortality in open and endovascular infrarenal abdominal aortic aneurysm repair. South African journal of surgery. Suid-Afrikaanse tydskrif vir chirurgie 45(2), 39-42	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
149	Moore W S, Kashyap V S, Vescera C L, and Quinones-Baldrich W J (1999) Abdominal aortic aneurysm: a 6-year comparison of endovascular versus transabdominal repair. Annals of surgery 230(3), 298-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
150	Moore Wesley S (2003) The Guidant Ancure bifurcation endograft: five-year follow-up. Seminars in vascular surgery 16(2), 139-43	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
151	Moore Wesley S, Matsumura Jon S, Makaroun Michel S, Katzen Barry T, Deaton David H, Decker Maria, Walker Gary, Investigators E VT, and Guidant (2003) Five-year interim comparison of the Guidant bifurcated endograft with open repair of abdominal aortic aneurysm. Journal of vascular surgery 38(1), 46-55	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
152	Morisaki Koichi, Matsumoto Takuya, Matsubara Yutaka, Inoue Kentaro, Aoyagi Yukihiko, Matsuda Daisuke, Tanaka Shinichi, Okadome Jun, and Maehara Yoshihiko (2016) Elective endovascular vs. open repair for abdominal aortic aneurysm in octogenarians. Vascular 24(4), 348-54	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
153	Mujib M, Alcocer F, Passman M, Matthews T C, and Jordan W D (2013) Secondary procedures and long-term morbidity and mortality following endovascular and open repair of abdominal aortic aneurysm. Vascular Disease Management 10(7), E124-E129	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
154	Nordon I M, Hinchliffe R J, Holt P J, Loftus I M, and Thompson M M (2009) Modern treatment of juxtarenal abdominal aortic aneurysms with fenestrated endografting and open repair--a systematic review. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 38(1), 35-41	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
155	Onohara, T., Kyuragi, R., Inoue, K. et al. (2018) Late-Onset Malignant Neoplasms and Their Prognostic Factors after Abdominal Aortic Aneurysm Repair. Annals of Vascular Surgery	No protocol-specified outcomes
156	Overbey Douglas M, Glebova Natalia O, Chapman Brandon C, Hosokawa Patrick W, Eun John C, and Nehler Mark R (2017) Morbidity of endovascular abdominal aortic aneurysm repair is directly related to diameter. Journal of vascular surgery 66(4), 1037-1047.e7	Data on desired outcomes were not extractable from the study manuscript.
157	Pane B, Spinella G, Signori A, Musio D, Perfumo M G, Lucertini G, Rousas N, and Palombo D (2014) Early and long-term outcomes after open or endovascular repair for abdominal aortic aneurysms in high-risk patients. The Journal of cardiovascular surgery 55(2), 257-63	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
158	Paolini David, Chahwan Santiago, Wojnarowski Dennis, Pigott John P, LaPorte Frankie, and Comerota Anthony J (2008) Elective endovascular and open repair of abdominal aortic aneurysms in octogenarians. Journal of vascular surgery 47(5), 924-7	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
159	Patel Ajay P, Langan Eugene M, 3rd , Taylor Spence M, Gray Bruce H, Carsten Christopher G, Cull David L, Snyder Bruce A, Stanbro Marcus D,	No propensity score matching, multivariable regression, or other

No.	Study	Reason for exclusion
	Youkey Jerry R, and Sullivan Timothy M (2003) An analysis of standard open and endovascular surgical repair of abdominal aortic aneurysms in octogenarians. The American surgeon 69(9), 744-748	statistical methods for adjusting for differences in case mix were performed.
160	Patel Virendra I, Lancaster Robert T, Conrad Mark F, Lamuraglia Glenn M, Kwolek Christopher J, Brewster David C, and Cambria Richard P (2011) Comparable mortality with open repair of complex and infrarenal aortic aneurysm. Journal of vascular surgery 54(4), 952-9	Study does not compare interventions: instead it is a comparison between different types of AAA (e.g. complex, non-complex) and their outcomes.
161	Pecoraro, Felice, Gloekler, Steffen, Mader, Caecilia E. et al. (2018) Mortality rates and risk factors for emergent open repair of abdominal aortic aneurysms in the endovascular era. Updates in surgery 70(1): 129-136	Study includes people with ruptured AAAs in its analysis. Ruptured AAAs are not within the scope of this review.
162	Peterson Brian G, Matsumura Jon S, Brewster David C, Makaroun Michel S, Excluder Bifurcated Endoprosthesis, and Investigators (2007) Five-year report of a multicenter controlled clinical trial of open versus endovascular treatment of abdominal aortic aneurysms. Journal of vascular surgery 45(5), 885-90	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
163	Pfeiffer T, Reiher L, Grabitz K, and Sandmann W (1998) Open surgery or endovascular treatment of the abdominal aortic aneurysm--quality assurance is urgently needed. Journal des maladies vasculaires 23(5), 393-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
164	Piffaretti Gabriele, Mariscalco Giovanni, Riva Francesca, Fontana Federico, Carrafiello Gianpaolo, and Castelli Patrizio (2014) Abdominal aortic aneurysm repair: long-term follow-up of endovascular versus open repair. Archives of medical science : AMS 10(2), 273-82	Study pools data on unruptured and ruptured AAA in its analyses.
165	Prault Trent L, Stevens Scott L, Freeman Michael B, Cassada David, Hardin Rob, and Goldman Mitchell H (2004) Open versus endo: early experience with endovascular abdominal aortic aneurysm repair beyond the clinical trials. The heart surgery forum 7(5), E459-61	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
166	Prinssen Monique, Buskens Erik, de Jong , Sjors E, Buth Jacob, Mackaay Albert J, van Sambeek , Marc R, Blankensteijn Jan D, and participants Dream trial (2007) Cost-effectiveness of conventional and endovascular repair of abdominal aortic aneurysms: results of a randomized trial. Journal of vascular surgery 46(5), 883-890	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
167	Quinney Brent E, Parmar Gaurav M, Nagre Shardul B, Patterson Mark, Passman Marc A, Taylor Steve, Chambers James, and Jordan William D (2011) Long-term single institution comparison of endovascular aneurysm repair and open aortic aneurysm repair. Journal of vascular surgery 54(6), 1592-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
168	Rasheed Khurram, Cullen John P, Seaman Matthew J, Messing Susan, Ellis Jennifer L, Glocker Roan J, Doyle Adam J, and Stoner Michael C (2016) Aortic anatomic severity grade correlates with resource utilization. Journal of vascular surgery 63(3), 569-76	This is a health economic analysis that explores risk factors that lead to increased treatment costs.
169	Ren Shiyan, Fan Xueqiang, Ye Zhidong, and Liu Peng (2012) Long-term outcomes of endovascular repair versus open repair of abdominal aortic aneurysm. Annals of thoracic and cardiovascular surgery : official journal of the Association of Thoracic and Cardiovascular Surgeons of Asia 18(3), 222-7	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
170	Revuelta Suero, S., Martinez Lopez, I., Hernandez Mateo, M. et al. (2019) Outcomes of the Repair of 1000 Abdominal Aortic Aneurysms in the Endovascular Era. Annals of Vascular Surgery 22: 22	Study includes people who received EVAR before 1999.
171	Robinson William P, Huang Wei, Rosen Amy, Schanzer Andres, Fang Hua, Anderson Frederick A, and Messina Louis M (2015) The Agency for Healthcare Research and Quality Inpatient Quality Indicator #11 overall mortality rate does not accurately assess mortality risk after abdominal aortic aneurysm repair. Journal of vascular surgery 61(1), 44-9	Study assesses the utility of risk assessment tools across different hospitals and does not make comparisons between treatment options.
172	Rocha, Rodolfo V., Friedrich, Jan O., Elbatarny, Malak et al. (2018) A systematic review and meta-analysis of early outcomes after endovascular versus open repair of thoracoabdominal aortic aneurysms. Journal of vascular surgery 68(6): 1936-1945.e5	Thoracoabdominal aortic aneurysms (with no subgroup for type IV cases)
173	Rose John, Evans Christopher, Barleben Andrew, Bandyk Dennis, Wilson S Eric, Chang David C, and Lane John (2014) Comparative safety of endovascular aortic aneurysm repair over open repair using patient safety indicators during adoption. JAMA surgery 149(9), 926-32	Outcomes of patients with unruptured and ruptured AAAs were pooled together in the analyses.
174	Rucinska, Z., Juzwiszyn, J., Bolanowska, Z. et al. (2018) The evaluation of the postoperative course in patients operated due to abdominal aortic aneurysm as urgent or elective procedure. Polski Przeglad Chirurgiczny 90(6): 1-5	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
175	Sajid Muhammad S, Desai Mittal, Haider Zishan, Baker Daryll M, and Hamilton George (2008) Endovascular aortic aneurysm repair (EVAR) has significantly lower perioperative mortality in comparison to open repair: a systematic review. Asian journal of surgery 31(3), 119-23	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
176	Sala-Almonacil Vicente A, Zaragoza-Garcia Jose M, Ramirez-Montoya Mauricio, Molina-Nacher Vicente, Martinez-Perello Inmaculada, and Gomez-Palones Francisco J (2017) Fenestrated and chimney endovascular aneurysm repair versus open surgery for complex abdominal aortic aneurysms. The Journal of cardiovascular surgery 58(6), 801-813	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
177	Salata, Konrad, Hussain, Mohamad A., de Mestral, Charles et al. (2018) Prevalence of Elective and Ruptured Abdominal Aortic Aneurysm Repairs by Age and Sex From 2003 to 2016 in Ontario, Canada. JAMA network open 1(7): e185418	No protocol-specified outcomes
178	Sandford R M, Choke E, Bown M J, and Sayers R D (2014) What is the best option for elective repair of an abdominal aortic aneurysm in a young fit patient?. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 47(1), 13-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
179	Sandridge Layne C, Baglioni A J, Jr , Kongable Gail L, and Harthun Nancy L (2006) Evaluation of the effect of endovascular options on infrarenal abdominal aortic aneurysm repair. The American surgeon 72(8), 700-6	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
180	Schermerhorn Marc L, Giles Kristina A, Hamdan Allen D, Dalhberg Suzanne E, Hagberg Robert, and Pomposelli Frank (2008) Population-based outcomes of open descending thoracic aortic aneurysm repair. Journal of vascular surgery 48(4), 821-7	Superseded by an included study (Schermerhorn et al., 2015)
181	Schouten O, Dunkelgrun M, Feringa H H. H, Kok N F. M, Vidakovic R, Bax J J, and Poldermans D (2007) Myocardial Damage in High-risk Patients Undergoing Elective Endovascular or Open Infrarenal Abdominal Aortic Aneurysm Repair. European Journal of Vascular and Endovascular Surgery 33(5), 544-549	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
182	Schouten O, Lever T M, Welten G M. J. M, Winkel T A, Dols L F. C, Bax J J, van Domburg , R T, Verhagen H J. M, and Poldermans D (2008) Long-term cardiac outcome in high-risk patients undergoing elective endovascular or open infrarenal abdominal aortic aneurysm repair. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 36(6), 646-52	No protocol-specified outcomes were reported.
183	Shahverdyan R, Majd M P, Thul R, Braun N, Gawenda M, and Brunkwall J (2015) F-EVAR does not Impair Renal Function more than Open Surgery for Juxtarenal Aortic Aneurysms: Single Centre Results. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 50(4), 432-41	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
184	Sidloff D A, Saratzis A, Sweeting M J, Michaels J, Powell J T, Thompson S G, and Bown M J (2017) Sex differences in mortality after abdominal aortic aneurysm repair in the UK. The British journal of surgery 104(12), 1656-1664	Study compares outcomes between men and women, and makes no comparisons between treatment options.
185	Siracuse J J, Schermerhorn M L, Meltzer A J, Eslami M H, Kalish J A, Rybin D, Doros G, Farber A, Vascular Study Group of New, and England (2016) Comparison of outcomes after endovascular and open repair of abdominal aortic	Study includes people with ruptured AAA in its analyses.

No.	Study	Reason for exclusion
	aneurysms in low-risk patients. The British journal of surgery 103(8), 989-94	
186	Siracuse Jeffrey J, Gill Heather L, Graham Ashley R, Schneider Darren B, Connolly Peter H, Sedrakyan Art, and Meltzer Andrew J (2014) Comparative safety of endovascular and open surgical repair of abdominal aortic aneurysms in low-risk male patients. Journal of vascular surgery 60(5), 1154-8	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
187	Soler, R., Bartoli, M. A., Faries, C. et al. (2019) Fenestrated endovascular aneurysm repair and open surgical repair for the treatment of juxtarenal aortic aneurysms. Journal of Vascular Surgery.	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
188	Soulez Gilles, Therasse Eric, Monfared Amir Abbas Tahami, Blair Jean-Francois, Choiniere Manon, Elkouri Stephane, Beaudoin Nathalie, Giroux Marie-France, Cliche Andree, Lelorier Jacques, and Oliva Vincent L (2005) Pain and quality of life assessment after endovascular versus open repair of abdominal aortic aneurysms in patients at low risk. Journal of vascular and interventional radiology : JVIR 16(8), 1093-100	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
189	Speicher Paul J, Barbas Andrew S, and Mureebe Leila (2014) Open versus endovascular repair of ruptured abdominal aortic aneurysms. Annals of vascular surgery 28(5), 1249-57	Study only assesses outcomes of people with ruptured AAA.
190	Steinmetz E, Abello N, Kretz B, Gauthier E, Bouchot O, and Brenot R (2010) Analysis of outcome after using high-risk criteria selection to surgery versus endovascular repair in the modern era of abdominal aortic aneurysm treatment. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 39(4), 403-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
191	Stone William M, Fankhauser Grant T, Bower Thomas C, Oderich Gustavo S, Oldenburg W Andrew, Kalra Manju, Naidu Sailendra, and Money Samuel R (2012) Comparison of open and endovascular repair of inflammatory aortic aneurysms. Journal of vascular surgery 56(4), 951-6	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
192	Stroupe Kevin T, Lederle Frank A, Matsumura Jon S, Kyriakides Tassos C, Jonk Yvonne C, Ge Ling, Freischlag Julie A, Open Versus Endovascular Repair Veterans Affairs Cooperative S, and Group (2012) Cost-effectiveness of open versus endovascular repair of abdominal aortic aneurysm in the OVER trial. Journal of vascular surgery 56(4), 901-9.e2	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
193	Suckow Bjoern D, Goodney Philip P, Columbo Jesse A, Kang Ravinder, Stone David H, Sedrakyan Art, Cronenwett Jack L, and Fillinger Mark F (2018) National trends in open surgical, endovascular, and branched-fenestrated endovascular aortic aneurysm repair in Medicare patients. Journal of vascular surgery 67(6), 1690-1697.e1	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
194	Sultan Sherif, and Hynes Niamh (2011) Clinical efficacy and cost per quality-adjusted life years of pararenal endovascular aortic aneurysm repair compared with open surgical repair. Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists 18(2), 181-96	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
195	Takagi Hisato, and Umemoto Takuya (2011) A meta-analysis of randomized and risk-adjusted observational studies of endovascular versus open repair for ruptured abdominal aortic aneurysm. Vascular and endovascular surgery 45(8), 717-9	Systematic review: individual studies were reviewed to ascertain eligibility.
196	Tarride Jean-Eric, Blackhouse Gord, De Rose , Guy , Novick Teresa, Bowen James M, Hopkins Robert, O'Reilly Daria, and Goeree Ron (2008) Cost-effectiveness analysis of elective endovascular repair compared with open surgical repair of abdominal aortic aneurysms for patients at a high surgical risk: A 1-year patient-level analysis conducted in Ontario, Canada. Journal of vascular surgery 48(4), 779-87	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
197	Teivelis Marcelo Passos, Malheiro Daniel Tavares, Hampe Marcio, Dalio Marcelo Bellini, and Wolosker Nelson (2016) Endovascular Repair of Infrarenal Abdominal Aortic Aneurysm Results in Higher Hospital Expenses than Open Surgical Repair: Evidence from a Tertiary Hospital in Brazil. Annals of vascular surgery 36, 44-54	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
198	Teufelsbauer Harald, Polterauer Peter, Lammer Johannes, Huk Ihor, Nanobachvili Josif, and Kretschmer Georg (2006) Repair of abdominal aortic aneurysms: the benefits of offering both endovascular and open surgical techniques. Perspectives in vascular surgery and endovascular therapy 18(3), 238-46	This is a narrative review of studies that includes the authors' opinions
199	Teufelsbauer Harald, Prusa Alexander M, Wolff Klaus, Polterauer Peter, Nanobashvili Josif, Prager Manfred, Holzenbein Thomas, Thurnher Siegfried, Lammer Johannes, Schemper Michael, Kretschmer Georg, and Huk Ihor (2002) Endovascular stent grafting versus open surgical operation in patients with infrarenal aortic aneurysms: a propensity score-adjusted analysis. Circulation 106(7), 782-7	Although propensity analysis was performed, it used to assess risk factors and wasn't used to compare EVAR with OSR.
200	The Japanese Society For Vascular Surgery Database Management Committee, Member and Ncd Vascular Surgery Data Analysis, Team (2018) Vascular Surgery in Japan: 2011 Annual Report by the Japanese Society for Vascular Surgery. Annals of vascular diseases 11(3): 377-397	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
201	Thomas Dustin, Anderson David, Hulten Edward, McRae Fiora, Ellis Shane, Malik Jamil A, Villines Todd C, and Slim Ahmad M (2015) Open versus endovascular repair of abdominal aortic aneurysm: Incidence of cardiovascular events in 632 patients in a department of defense cohort over 6-year follow-up. Vascular 23(3), 234-9	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

No.	Study	Reason for exclusion
202	Thompson, Simon G., Bown, Matthew J., Glover, Matthew J. et al. (2018) Screening women aged 65 years or over for abdominal aortic aneurysm: a modelling study and health economic evaluation. Health technology assessment (Winchester, England) 22(43): 1-142	Not a comparison of EVAR and OSR
203	Trenner M, Kuehnl A, Salvermoser M, Reutersberg B, Geisbuesch S, Schmid V, and Eckstein H H (2018) Editor's Choice - High Annual Hospital Volume is Associated with Decreased in Hospital Mortality and Complication Rates Following Treatment of Abdominal Aortic Aneurysms: Secondary Data Analysis of the Nationwide German DRG Statistics from 2005 to 2013. European Journal of Vascular and Endovascular Surgery 55(2), 161	Although propensity analysis was performed, it used to assess risk factors and wasn't used to compare EVAR with OSR.
204	Trenner, M., Kuehnl, A., Reutersberg, B. et al. (2018) Nationwide analysis of risk factors for in-hospital mortality in patients undergoing abdominal aortic aneurysm repair. The British journal of surgery 105(4): 379-387	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
205	Tsilimparis Nikolaos, Perez Sebastian, Dayama Anand, Ricotta Joseph J, and 2nd (2012) Age-stratified results from 20,095 aortoiliac aneurysm repairs: should we approach octogenarians and nonagenarians differently?. Journal of the American College of Surgeons 215(5), 690-701	No comparisons made across interventions. Instead comparisons are made across different age groups of people treated by EVAR
206	Turnbull Irene C, Criado Frank J, Sanchez Luis, Sadek Mikel, Malik Rajesh, Ellozy Sharif H, Marin Michael L, and Faries Peter L (2010) Five-year results for the Talent enhanced Low Profile System abdominal stent graft pivotal trial including early and long-term safety and efficacy. Journal of vascular surgery 51(3), 537-2	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
207	Unsgard, R. G., Altreuther, M., Lange, C. et al. (2019) Five-year results of endovascular aortic repair used according to instructions for use give a good general outcome for abdominal aortic aneurysm. SAGE Open Medicine 7: 2050312119853434	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
208	Valdivia, A. R., Fuente, M. F., Santos, A. D. et al. (2019) Impact of the Aortic Graft on Arterial Stiffness and Inflammatory Biomarkers after Endovascular Aortic Repair or Open Surgical Repair in Abdominal Aortic Aneurysm Disease. Annals of Vascular Surgery 23: 23	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
209	van Bochove , Cornelis A, Burgers Laura T, Vahl Anco C, Birnie Erwin, van Schothorst , Marien G, and Redekop William K (2016) Cost-effectiveness of open versus endovascular repair of abdominal aortic aneurysm. Journal of vascular surgery 63(3), 827-38.e2	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
210	Wahlgren Carl Magnus, Malmstedt Jonas, Swedish Vascular, and Registry (2008) Outcomes of endovascular abdominal aortic aneurysm repair compared with open surgical repair in high-risk	Although authors stated that regression was performed, they did not report data in a reasonable format to inform this NICE review.

No.	Study	Reason for exclusion
	patients: results from the Swedish Vascular Registry. Journal of vascular surgery 48(6), 1382-9	
211	Wang, J. C., Chien, W. C., Chung, C. H. et al. (2019) Association between surgical repair of aortic aneurysms and the diagnosis of intracranial aneurysms. Journal of Vascular Surgery 13: 13	No protocol-specified outcomes were reported.
212	Wang, Jen-Chun, Chien, Wu-Chien, Tzeng, Nian-Sheng et al. (2019) Surgical repair of aortic aneurysms and reduced incidence of dementia. International journal of cardiology 278: 46-50	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
213	Wanhainen A, Bylund N, Björck M. (2008) Outcome after abdominal aortic aneurysm repair in Sweden 1994-2005. Br J Surg. 95(5):564-70.	Study includes people who received EVAR before 1999.
214	White G H, May J, McGahan T, Yu W, Waugh R C, Stephen M S, and Harris J P (1996) Historic control comparison of outcome for matched groups of patients undergoing endoluminal versus open repair of abdominal aortic aneurysms. Journal of vascular surgery 23(2), 201-2	Insufficient details were provided to ascertain study methodology.
215	Williams Christopher R, and Brooke Benjamin S (2017) Effectiveness of open versus endovascular abdominal aortic aneurysm repair in population settings: A systematic review of statewide databases. Surgery 162(4), 707-720	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.
216	Wilt Timothy J, Lederle Frank A, Macdonald Roderick, Jonk Yvonne C, Rector Thomas S, and Kane Robert L (2006) Comparison of endovascular and open surgical repairs for abdominal aortic aneurysm. Evidence report/technology assessment (144), 1-113	Systematic review of multiple study designs: individual studies were reviewed to ascertain suitability.
217	Yoshitani, Kenji, Masui, Kenichi, Kawaguchi, Masahiko et al. (2018) Clinical Utility of Intraoperative Motor-Evoked Potential Monitoring to Prevent Postoperative Spinal Cord Injury in Thoracic and Thoracoabdominal Aneurysm Repair: An Audit of the Japanese Association of Spinal Cord Protection in Aortic Surgery Database. Anesthesia and analgesia 126(3): 763-768	Thoracoabdominal aortic aneurysms (with no subgroup for type IV cases)
218	Zeebregts C J, Geelkerken R H, van der Palen , J , Huisman A B, de Smit , P , van Det , and R J (2004) Outcome of abdominal aortic aneurysm repair in the era of endovascular treatment. The British journal of surgery 91(5), 563-8	Multivariate analyses or propensity score matching were not used to compare EVAR with OSR.
220	Zhang Chang-Lie, Song Zhi-Hong, and Wang Fan (2016) Comparison of Efficacy of Endovascular Aneurysm Repair Versus Open Surgical Repair in Middle/High-Risk Patients With Abdominal Aortic Aneurysm. American journal of therapeutics 23(1), e37-43	No propensity score matching, multivariable regression, or other statistical methods for adjusting for differences in case mix were performed.

Appendix D Instrument for appraising risk of bias

Critical appraisal for this review was performed according to the bespoke instrument shown in Table 4. It amalgamates key criteria from The categories draw on the generic provisions of the Cochrane Collaboration's 'Risk Of Bias In Non-randomized Studies of Interventions' (ROBINS-I) tool (Sterne et al., 2016) and the more technically focused criteria in the 'Quality of Effectiveness Estimates from Non-Randomised Studies' (QuEENS) checklist (Faria et al., 2015). See 2.4 for a description of how we developed the instrument.

Table 4: Bespoke instrument for appraising risk of bias in casemix-adjusted observational studies

Criteria	Domain addressed	Risk of bias	Definitions
1. Selection			
1.1. Were cohorts from the same time period?	QuEENS 11b	LOW	Contemporaneous recruitment over <5yrs (≥5 yrs if year of operation controlled for)
		MODERATE	≥5-yr recruitment with no adjustment for year of operation
		HIGH	Cohorts drawn from different periods in time (e.g. historical controls)
1.2. Were cohorts from the same place?	QuEENS 11b	LOW	Same hospital, area or country
		HIGH	Different sampling frames
1.3. Is the definition of AAA the same across cohorts?	QuEENS 11a	LOW	Differences unlikely
		MODERATE	All infrarenal and complex AAA included, with no adjustment for anatomy
		HIGH	Differences likely (especially where only 1 cohort may include complex AAA)
2. Confounding			
2.1. Does study control appropriately for demographics?	ROBINS-I 1.4 QuEENS 11c	LOW	At least age and sex
		MODERATE	Just age or sex
		HIGH	Neither age nor sex, or invalid adjustment methods
2.2. Does study control appropriately for comorbidity and/or fitness?	ROBINS-I 1.4 QuEENS 11c	LOW	Good range of individual comorbidities or validated index (e.g. Charlson, Elixhauser)
		MODERATE	Limited number of individual comorbidities
		HIGH	None, or invalid adjustment methods
2.3. Does study control appropriately for AAA characteristics?	ROBINS-I 1.4 QuEENS 11c	LOW	At least diameter and some measure of extent (unless study is limited to 1 extent)
		MODERATE	Just diameter or extent
		HIGH	None, or invalid adjustment methods

Criteria	Domain addressed	Risk of bias	Definitions
2.4. Could any adjustment variables have been affected by the intervention?	ROBINS-I 1.6	LOW	No post-intervention variables controlled for that may mediate treatment effect
		HIGH	Post-intervention variable(s) controlled for that may mediate treatment effect
3. Data collection			
3.1. Is method of data collection likely to have identified suitable participants accurately?	New	LOW	Medical records
		MODERATE	Detailed surgical registries (diagnosis and procedure codes specified)
		HIGH	Administrative registries with high-level diagnosis and procedure codes
3.2. Is method of data collection likely to record perioperative outcomes accurately?	New	LOW	Medical records or detailed surgical registries
		HIGH	Administrative registries with high-level diagnosis and procedure codes
3.3. Is method of data collection likely to record long-term outcomes accurately?	New	LOW	Direct use of reliable routine data registries
		MODERATE	Surgical registries or medical records with linkage to reliable routine data registries
		HIGH	Surgical registries or medical records (+/- questionnaires, etc.)
4. Analysis – general			
4.1. Were any checks conducted on model specification and/or fit?	QuEENS 5	LOW	Residual plots and/or formal tests (e.g. misspecification, autocorrelation, etc.)
		HIGH	None reported
4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for?	ROBINS-I 5.1–5.5	LOW	Few missing data or amount and types of missingness similar across cohorts
		HIGH	Possible differential missingness with no valid adjustment or not considered
4.3. Have different methods been compared within the study?	QuEENS 1	LOW	Methods with different assumptions re selection on unobservables compared
		MODERATE	Different methods compared but all rely on the same assumption about selection
		HIGH	No different methods compared
5. Analysis – matching			
5.1. Is the matching algorithm reported and reasonable?	QuEENS 12–14	LOW	E.g. nearest-neighbour or caliper/radius matching with reasonable assumptions
		HIGH	Unreported or invalid methods
5.2. Was overlap / common support appropriately assessed?	QuEENS 7	LOW	Checks on distribution of propensity score with trimming where necessary
		MODERATE	Comparisons of minima and maxima
		HIGH	No assessment reported

Criteria	Domain addressed	Risk of bias	Definitions
5.3. Has balancing of the covariates been demonstrated?	QuEENS 8	LOW	Normalised differences in covariates reported, with none >0.25
		MODERATE	Conventional hypothesis tests, with no evidence of significant differences
		HIGH	Meaningful differences in 1 or more important covariates or not reported
6. Analysis – simple multivariable models			
6.1. Is sample size adequate relative to number of covariates considered?	New	LOW	Number of events is ≥10 times greater than number of variables considered
		HIGH	Number of events is <10 times greater than number of variables considered
6.2. Were interactions between treatment and other covariates considered?	New	LOW	Yes
		HIGH	No
Overall			
Domain scores represent the average score given for each question (rounded mean where 1=low, 2=medium & 3=high) Overall judgement based on rules stated here.		LOW	0 domains high and <2 moderate
		MODERATE	0 domains high and >1 moderate **or** 1 domain high and <2 moderate
		HIGH	>1 domain high **or** 1 domain high and >1 moderate

Appendix E Evidence tables for included studies

E.1 Infrarenal AAAs

Full citation	Behrendt CA, Sedrakyan A, Christian H, et al. (2017) Short-term and long-term results of endovascular and open repair of abdominal aortic aneurysms in Germany. J Vasc Surg. 66(6):1704-1711.e3. doi: 10.1016/j.jvs.2017.04.040.
Study details	Study design: Retrospective cohort study Location(s): Germany Study period: October 2008 to April 2015 Aim of the study: to determine the short- and long-term outcomes of EVAR and OSR of unruptured and ruptured AAA and to assess whether recently reported results from RCTs reflect real world practice.
Participants	Sample size: EVAR group, n=3,493; OSR group, n=1,457 Inclusion criteria: patients who received EVAR or OSR for unruptured or ruptured infrarenal AAA were included. Exclusion criteria: authors state that missing values were excluded from the analysis. No further details were provided Baseline characteristics: Mean age (range): EVAR group, 74 (69-79) years; OSR group, 71 (66-76) years Gender: EVAR group, 85.4% male; OSR group, 82.8% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 16.5%; OSR group, 14.6% Hypertension: EVAR group, 70.0%; OSR group, 69.5% Dyslipidaemia: EVAR group, 39.0%; OSR group, 36.0% COPD: EVAR group, 14.5%; OSR group, 16.4% History of myocardial infarction: EVAR group, 9.1%; OSR group, 9.3% History of stroke: EVAR group, 1.8%; OSR group, 1.2%
Methods	Data collection: Data were collected from databases of the third largest health insurance provider in Germany (DAK-G). Patients who underwent AAA repair were identified using ICD10 codes and procedure codes in the database. For the identified cases that matched basic criteria, investigators collected data on demographics, procedures done while in hospital, coded comorbidities, and reason for discharge. For long-term outcomes investigators censored patients whose insurance contract expired within the study period: similar percentages of censored cases were reported across the 2 treatment arms.

Full citation	**Behrendt CA, Sedrakyan A, Christian H, et al. (2017) Short-term and long-term results of endovascular and open repair of abdominal aortic aneurysms in Germany. J Vasc Surg. 66(6):1704-1711.e3. doi: 10.1016/j.jvs.2017.04.040.**
	Analysis: multivariate Cox proportional hazard regression; selection of the model and range of adjusting covariates were based on statistical significance of variables in the bivariate model. Automatic backwards selection was used for the final parsimonious model.
Intervention	EVAR
Comparison	Open surgical repair
Outcomes	In-hospital mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation 1.2. Were cohorts from the same place? Low risk – study sample derived from a German Health insurance provider database. 1.3. Is the definition of AAA the same across cohorts? Low risk – all participants had AAA repair of infrarenal AAA according to clinical codes. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – model adjusted for multiple variables including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – model adjusted for multiple comorbidities. 2.3. Does study control appropriately for AAA characteristics? High risk – the study did not control for AAA characteristics. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – data obtained from a health insurance provider database. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – data obtained from a health insurance provider database. 3.3. Is method of data collection likely to record long-term outcomes accurately? High risk – data obtained from a health insurance provider database. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – None reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk - missing values were excluded from the analysis. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A

Full citation	**Behrendt CA, Sedrakyan A, Christian H, et al. (2017) Short-term and long-term results of endovascular and open repair of abdominal aortic aneurysms in Germany. J Vasc Surg. 66(6):1704-1711.e3. doi: 10.1016/j.jvs.2017.04.040.**
	5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of mortality events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – No interactions considered **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Bush RL, Johnson ML, Collins TC, et al. (2006) Open versus endovascular abdominal aortic aneurysm repair in VA hospitals. J Am Coll Surg. 202(4):577-87.** **Note this study includes the same population as Johnson et al 2006; however a different type of analysis was performed.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: May 2001 to September 2003 Aim of the study: to examine outcomes after elective aneurysm repair
Participants	Sample size: EVAR group, n=717; OSR group, n= 1,187 Inclusion criteria: all people who underwent EVAR or OSR of unruptured AAA were included. Exclusion criteria: people with ruptured AAA, thoracic or thoracoabdominal aortic aneurysm, or those who underwent conversion from EVAR to OSR were excluded. Baseline characteristics: Mean age (SD): EVAR group, 71.6 (7.8) years; OSR group,70.2 (7.9) years Gender: EVAR group, 99.6% male; OSR group, 99.1% male Mean aneurysm diameter: not reported COPD: EVAR group, 26.6%; OSR group, 26.0% Chronic heart failure: EVAR group, 3.6%; OSR group, 2.4% Renal insufficiency: EVAR group, 1.1%; OSR group, 1.0% Cerebrovascular accident with neuro-deficit: EVAR group, 5.3%; OSR group, 6.0% Diabetes: EVAR group, 14.4%; OSR group, 13.2%

Full citation	**Bush RL, Johnson ML, Collins TC, et al. (2006) Open versus endovascular abdominal aortic aneurysm repair in VA hospitals. J Am Coll Surg. 202(4):577-87.** **Note this study includes the same population as Johnson et al 2006; however a different type of analysis was performed.**
	Malignancy: EVAR group, 1.1%; OSR group, 1.0%
Methods	Data collection: Data collection: data were extracted from a detailed surgical registry run by the run by the military Veterans Health Administration: (A Veterans Affairs component of the National Surgical Quality Improvement Program; NSQIP). The NSQIP database requires hospitals to provide complete 30-day follow-up on at least 95% of patients. To supplement the information in the NSQIP records investigators used unique identifiers to link records with other Veterans Affairs databases: including the patient treatment file (which contains abstracts of all patients discharged), the outpatient clinic file (which contains records for every outpatient visit), and the VA beneficiary identification record locator system death file Analysis: Multivariate logistic regression was performed. All independent variables that were found to be significantly associated with morbidity and mortality outcomes (p values <0.1) using univariate analyses were included in the multivariate regression models. Authors stated that this level was selected arbitrarily to capture as many possible confounding factors as might be strongly associated with both the selection of EVAR and postoperative outcomes. Age was tested for linear association and found to have the best empiric fit as a categorical variable (greater than or equal to 80 years). Models were assessed for goodness of fit by the Hosmer-Lemeshow test and for discrimination by the c-index.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, mortality at 1 year and adverse events (including cardiac, neurologic, pulmonary, renal dysfunction, wound, graft failure and bleeding requiring blood transfusion).
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – both cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – both cohorts were drawn from the same time period. 1.3. Is the definition of AAA the same across cohorts? High risk – Exclusion of conversions from EVAR to OSR introduces bias. **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – confirmed that the study controlled for age but did not provide any information about whether gender was controlled for. 2.2. Does study control appropriately for comorbidity and/or fitness? High risk – authors do not provide any details about which demographic variables were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – authors do not provide any details about whether AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? High risk – authors do not provide any details about whether mediating variables were controlled for. **Data collection**

Full citation	**Bush RL, Johnson ML, Collins TC, et al. (2006) Open versus endovascular abdominal aortic aneurysm repair in VA hospitals. J Am Coll Surg. 202(4):577-87.** **Note this study includes the same population as Johnson et al 2006; however a different type of analysis was performed.**
	3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – a detailed surgical registry was used to identify participants with diagnosis and procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low – a detailed surgical registry was used to collect data on outcomes. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate – a surgical registry was used with linkage to routine data registries. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – model specification/fit was assessed using the Hosmer–Lemeshow test and the C-statsitc. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors stated (in another publication) that given the robust nature of the NSQIP and other databases used the likelihood of missing essential covariates is low. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – authors did not provide details about covariates in the model. Thus it is not possible to ascertain whether this quality assessment criterion was met. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Bush RL, Johnson ML, Hedayati N, et al. (2007) Performance of endovascular aortic aneurysm repair in high-risk patients: results from the Veterans Affairs National Surgical Quality Improvement Program. J Vasc Surg. 45(2):227-233; discussion 233-5.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: May 2001 to December 2004

Full citation	**Bush RL, Johnson ML, Hedayati N, et al. (2007) Performance of endovascular aortic aneurysm repair in high-risk patients: results from the Veterans Affairs National Surgical Quality Improvement Program. J Vasc Surg. 45(2):227-233; discussion 233-5.**
	Aim of the study: evaluate outcomes after elective EVAR performed in high-risk veterans
Participants	Sample size: EVAR group, n=788; OSR group, n=1,580 Inclusion criteria: patients considered **high-risk** who underwent EVAR or OSR for unruptured AAA were included. Minimum criteria for entry into our study included age ≥60 years and ASA classification 3 or 4. Patients were further classified according to the comorbidity variables of history of cardiac, respiratory, or hepatic disease, cardiac revascularization, renal insufficiency, and low serum albumin. Exclusion criteria: patients with secondary diagnostic codes for ruptured AAA or thoracic or thoracoabdominal aortic aneurysm were excluded from the analysis. Patients with codes representing open repair after EVAR were also excluded from primary analysis. Baseline characteristics: Mean age (SD): EVAR group, 72.9 (6.7) years; OSR group, 71.8 (6.4) years Gender: EVAR group, 99.4% male; OSR group, 99.2% male Mean aneurysm diameter: not reported High-risk respiratory condition: EVAR group, 57.7%; OSR group, 58.8% High-risk hepatic condition: EVAR group, 4.6%; OSR group, 5.0% High-risk cardiac condition: EVAR group, 75.6%; OSR group, 75.8% High-risk renal condition: EVAR group, 4.2%; OSR group, 6.7% Previous cardiac revascularisation: EVAR group, 21.3%; OSR group, 20.3%
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). Patients undergoing elective repair were identified using ICD9 diagnostic codes, as well as procedure codes. At the time of surgery, patients are enrolled in NSQIP, and baseline demographic, preoperative laboratory, and clinical information was collected by dedicated trained nurse reviewers. Additional perioperative data were subsequently collected by the nurses, including 30-day morbidity and mortality information. To supplement the information in the NSQIP records with longer-term utilisation and vital statistics data, investigators linked the dataset with reliable other routine data sources (VA Patient Treatment File, VA Outpatient Clinic File, A Beneficiary Identification Record Locator System). Analysis: multivariate logistic regression; selection of the model and range of adjusting covariates were based on statistical significance (p value<0.1) of variables in univariate analyses. Authors state that the significance level was selected arbitrarily to capture as many confounding variables as possible. Models were assessed for goodness to fit by the Hosmer-Lemeshow statistic and for discrimination by the c-index.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, 1-year mortality, and perioperative adverse events (including adverse cardiac events, renal dysfunction, pulmonary complications, wound complications, neurologic complications, postoperative bleeding requiring transfusion, and graft failure)

Full citation	**Bush RL, Johnson ML, Hedayati N, et al. (2007) Performance of endovascular aortic aneurysm repair in high-risk patients: results from the Veterans Affairs National Surgical Quality Improvement Program. J Vasc Surg. 45(2):227-233; discussion 233-5.**
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – populations drawn from the same time period 1.2. Were cohorts from the same place? Low risk – study sample derived from an American surgical registry 1.3. Is the definition of AAA the same across cohorts? High risk – Exclusion of conversions from EVAR to OSR introduces bias. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – demographic factors including age and gender were controlled for. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – authors stated that they adjusted for high-risk comorbidities. 2.3. Does study control appropriately for AAA characteristics? High risk – no details provided as to whether investigators controlled for AAA characteristics. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – data collected from a detailed surgical registry. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data collected from a detailed surgical registry. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – data were obtained from a detailed surgical registry with linkage to reliable routine data registries. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – models were assessed for goodness to fit by the Hosmer–Lemeshow statistic. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – not reported 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of mortality events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions considered **Overall risk of bias:** High risk

Full citation	**Bush RL, Johnson ML, Hedayati N, et al. (2007) Performance of endovascular aortic aneurysm repair in high-risk patients: results from the Veterans Affairs National Surgical Quality Improvement Program. J Vasc Surg. 45(2):227-233; discussion 233-5.**
	Directness: Partially applicable (high risk only)

Full citation	**Chadi SA, Rowe BW, Vogt KN, et al. (2012) Trends in management of abdominal aortic aneurysms. J Vasc Surg. 55(4):924-8.**
Study details	Study design: retrospective cohort study Location(s): Canada Study period: June 2000-May 2010 Aim of the study: evaluate patients undergoing elective repair of infrarenal AAAs and the longitudinal trends in surgical management
Participants	Sample size: EVAR group, n=875; OSR group, n=1,067 Inclusion criteria: all people who underwent EVAR or OSR for unruptured infrarenal AAA at a university-affiliated medical centre were included. Exclusion criteria: people with pararenal and suprarenal, visceral arterial a, isolated iliac, infected and ruptured aneurysms were excluded. Baseline characteristics: Mean age (SD): EVAR group, 71 (8.04) years; OSR group, 75 (8.05) years Gender: EVAR group, 87.5% male; OSR group, 82.2% male Mean aneurysm diameter: not reported Comorbidities: not reported
Methods	Data collection: data were obtained by reviewing an internally managed database of the university-affiliated medical centre. It is assumed that this database incorporated electronic health records. Analysis: authors state that “multivariable logistic regression was performed while adjusting for various preoperative variables”. No further details were provided.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality
Study Appraisal using NICE’s bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – recruitment over ≥5 yrs, but year of operation controlled for in analysis 1.2. Were cohorts from the same place? Low risk – all participants received treatment at the same university-affiliated medical centre. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA was similar across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – study controls for age.

Full citation	Chadi SA, Rowe BW, Vogt KN, et al. (2012) Trends in management of abdominal aortic aneurysms. J Vasc Surg. 55(4):924-8.
	2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – a limited number of comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – participants were identified using an internally-managed hospital database. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed using an internally-managed hospital database. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long-term outcomes were assessed **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no indication that missing data were considered. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Choke E, Lee K, McCarthy M, et al. (2012) Risk models for mortality following elective open and endovascular abdominal aortic aneurysm repair: a single institution experience. Eur J Vasc Endovasc Surg. 44(6): 549-54.
Study details	Study design: Prospective cohort study Location(s): UK

Full citation	**Choke E, Lee K, McCarthy M, et al. (2012) Risk models for mortality following elective open and endovascular abdominal aortic aneurysm repair: a single institution experience. Eur J Vasc Endovasc Surg. 44(6): 549-54.**
	Study period: January 2000 to October 2010 Aim of the study: to develop and validate an "in house" risk model for predicting perioperative mortality following elective AAA repair and to compare this with other models.
Participants	Sample size: EVAR group, n=589; OSR group, n= 564 Inclusion criteria: patients undergoing EVAR or OSR at a single medical centre were included. Exclusion criteria: not reported Baseline characteristics: NB – authors did not report demographic characteristics according to treatment groups Age: 69.6% were <70 years Gender: 88.9% male Mean aneurysm diameter: not reported Diabetes: 91.4% Any myocardial infarct: 75.8% Respiratory disease: 82.0% Receiving antihypertensive mediation: 50.5% Receiving statins: 74.3%
Methods	Data collection: data were prospectively collected from a single medical centre using proformas. Analysis: multivariate logistic regression; only variables that were found to be statistically significant (p value <0.1) on univariate analysis were entered into the multivariate model using a forward stepwise logistic regression analysis, to identify risk factors for perioperative mortality.
Intervention	EVAR
Comparator	OSR
Outcomes	Perioperative mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – study sample derived from a sampling frame of patients who underwent repair at 1 medical centre. 1.3. Is the definition of AAA the same across cohorts? High risk – authors report that 93.7% of aneurysms were infrarenal and the remainder were juxtarenal but they do not report the distribution across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – investigators only adjusted for age. 2.2. Does study control appropriately for comorbidity and/or fitness? High risk – no indication that comorbidities were adjusted for.

Full citation	**Choke E, Lee K, McCarthy M, et al. (2012) Risk models for mortality following elective open and endovascular abdominal aortic aneurysm repair: a single institution experience. Eur J Vasc Endovasc Surg. 44(6): 549-54.**
	2.3. Does study control appropriately for AAA characteristics? High risk – no indication that AAA characteristics were adjusted for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – data were collected prospectively using proformas. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data were collected prospectively using proformas. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long-term data assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – model fit was ascertained using the Hosmer–Lemeshow test. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – due to the nature of data collection it is unlikely that there was missing outcome data in the study cohort. 4.3. Have different methods been compared within the study? High risk - different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**de la Motte L, Jensen LP, Vogt K, et al. (2013) Outcomes after elective aortic aneurysm repair: a nationwide Danish cohort study 2007-2010. Eur J Vasc Endovasc Surg. 46(1):57-64.**
Study details	Study design: Retrospective cohort study Location(s): Denmark

Full citation	**de la Motte L, Jensen LP, Vogt K, et al. (2013) Outcomes after elective aortic aneurysm repair: a nationwide Danish cohort study 2007-2010. Eur J Vasc Endovasc Surg. 46(1):57-64.**
	Study period: January 2007 to December 2010 Aim of the study: to assess outcomes after treatment for asymptomatic AAA in Denmark in a period when both OSR and EVAR have been routine procedures.
Participants	Sample size: EVAR group, n=525; OSR group, n=1,176 Inclusion criteria: people who underwent elective AAA repair for asymptomatic unruptured AAA were included. Exclusion criteria: people with codes indicating the following were excluded: ruptured AAA, previous AAA repair, bypass from aorta to iliac artery for aneurysm, bypass from aorta to bilateral iliac arteries for aneurysm, bypass from aorta to iliac and contralateral femoral artery for aneurysm, bypass from aorta to femoral artery for aneurysm, bypass from aorta to bilateral femoral arteries for aneurysm, repair supracoeliac or juxtarenal AAA. Baseline characteristics: Mean age (range): EVAR group, 74 (69-78) years; OSR group, 70.5 (66-75) years Gender: EVAR group, 90% male; OSR group, 80% male Mean aneurysm diameter: not reported Smoking: EVAR group, 85%; OSR group, 84% Diabetes: EVAR group, 14%; OSR group, 9% Hypertension: EVAR group, 63%; OSR group, 67% Cardiac morbidity: EVAR group, 30%; OSR group, 18% Pulmonary morbidity: EVAR group, 23%; OSR group, 13% Cerebral morbidity: EVAR group, 14%; OSR group, 11%
Methods	Data collection: investigators obtained nationwide data on patients treated for asymptomatic unruptured AAA from the Danish Vascular Registry: a validated database of all procedures performed at vascular departments in Denmark. A manual search on each individual patient, using their unique social security number was done to match the registry data with data from the National patient register. Data were censored at the end of October 2011. Analysis: multivariate Cox regression; forward stepwise selection was used to input variables variables that were found to be statistically significant (p value <0.1) on univariate analysis into the multivariate model.
Intervention	EVAR
Comparator	OSR
Outcomes	All-cause mortality at 1 year
Study Appraisal	**Selection** 1.1. Were cohorts from the same time period? Low risk – populations drawn from the same time period.

Full citation	de la Motte L, Jensen LP, Vogt K, et al. (2013) Outcomes after elective aortic aneurysm repair: a nationwide Danish cohort study 2007-2010. Eur J Vasc Endovasc Surg. 46(1):57-64.
using NICE's bespoke risk of bias assessment tool	1.2. Were cohorts from the same place? Low risk – study sample derived from a sampling frame of patients who underwent vascular surgery in Denmark. 1.3. Is the definition of AAA the same across cohorts? High risk – EVAR cases explicitly limited to infrarenal, whereas OSR cases included supracoeliac and juxtarenal AAAs **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – model only adjusted for age. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – model adjusted for a good range of comorbidity variables including ASA scores. 2.3. Does study control appropriately for AAA characteristics? High risk – no indication that AAA characteristics were adjusted for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – data acquired from a detailed vascular surgery registry with diagnosis and procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data acquired from a detailed vascular surgery registry. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – data acquired from a detailed vascular surgery registry with linkage **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no evidence that checks were performed on model specification and/or fit. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no demonstration that missing data were taken into account in the analyses. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered

Full citation	**de la Motte L, Jensen LP, Vogt K, et al. (2013) Outcomes after elective aortic aneurysm repair: a nationwide Danish cohort study 2007-2010. Eur J Vasc Endovasc Surg. 46(1):57-64.**
	6.2 Were interactions between treatment and other covariates considered? **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Elkouri S, Gloviczki P, McKusick MA, et al. (2004) Perioperative complications and early outcome after endovascular and open surgical repair of abdominal aortic aneurysms. J Vasc Surg. 39(3):497-505.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: December 1999 to December 2001 Aim of the study: to compare the early results of elective EVAR with open repair that was performed during the same period at a sincle medical centre.
Participants	Sample size: EVAR group, n=94; OSR group, n= 261 Inclusion criteria: all patients who underwent elective infrarenal AAA repair at a single medical centre were included. Exclusion criteria: patients with juxtarenal AAA, associated planned visceral or renal revascularization, mycotic or false aneurysms, associated aortic dissection, or ruptured aneurysms were excluded. Baseline characteristics: Mean age (range): EVAR group, 77 (61-98) years; OSR group, 73 (52-90) years Gender: EVAR group, 90.4% male; OSR group, 87.7% male Mean aneurysm diameter: not reported Details on comorbidities were not available in the full study manuscript (online supplement only).
Methods	Data collection: investigators retrospectively reviewed the clinical and radiologic records of all patient who underwent elective AAA repair procedures. A minimum of 30 days of follow-up was obtained for all patients for the 30-day morbidity and mortality complications. Analysis: Multivariate logistic regression and Cox proportional hazards regression were used to analyse the association between type of surgical procedure and the 30-day outcomes of cardiac, pulmonary, and graft complications, as well as reintervention within 30 days. Multiple models were used to adjust for age, gender and high-risk status (a higher risk of complications from OSR because of associated comorbidities or because of relative contraindications to OSR).
Intervention	EVAR
Comparator	OSR

Full citation	Elkouri S, Gloviczki P, McKusick MA, et al. (2004) Perioperative complications and early outcome after endovascular and open surgical repair of abdominal aortic aneurysms. J Vasc Surg. 39(3):497-505.
Outcomes	Adverse events within 30 days (cardiac, pulmonary and graft complications reported separately), and reintervention within 30 days
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants were treated at the same medical centre. 1.3. Is the definition of AAA the same across cohorts? High risk – EVAR cohort defined as patients at higher risk due to comorbidities and contraindications to OSR. **Confounding** 2.1. Does study control appropriately for demographics? Low risk –study controls for demographics including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – a limited number of comorbidities were controlled using an unvalidated bespoke risk tool. 2.3. Does study control appropriately for AAA characteristics? High risk – there is no indication that AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – medical records were used to identify participants. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed by examining medical records. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no details about any checks were provided. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors do not provide any details about how missing data on outcomes were managed. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered.

Full citation	**Elkouri S, Gloviczki P, McKusick MA, et al. (2004) Perioperative complications and early outcome after endovascular and open surgical repair of abdominal aortic aneurysms. J Vasc Surg. 39(3):497-505.**
	6.2 Were interactions between treatment and other covariates considered? High risk – no interactions between treatment and other covariates were considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Feringa HHH, Karagiannis S, Vidakovic R, et al. (2007) Comparison of the Incidences of Cardiac Arrhythmias, Myocardial Ischemia, and Cardiac Events in Patients Treated With Endovascular Versus Open Surgical Repair of Abdominal Aortic Aneurysms. Am J Cardiol. 100:1479-1484**
Study details	Study design: prospective cohort study Location(s): The Netherlands Study period: 2002 to 2006 Aim of the study: to examine differences in cardiac arrhythmias, perioperative myocardial ischemia, troponin T release, and cardiovascular events between endovascular and open repair of abdominal aortic aneurysms
Participants	Sample size: EVAR group, n=49; OSR group, n=126 Inclusion criteria: elective open or endovascular repair of infrarenal AAAs Exclusion criteria: Patients with a cardiac pacemaker, left ventricular hypertrophy, left or right bundle branch block, or atrial fibrillation were excluded. Patients who participated in clinical intervention trials in or outside the Erasmus Medical Centre were also excluded Baseline characteristics: Age >70 years: EVAR group, 65%; OSR group, 52% Gender: EVAR group, 86% male; OSR group, 83% male Mean aneurysm diameter: not reported Congestive heart failure: EVAR group, 2%; OSR group, 4% Renal failure: EVAR group, 4%; OSR group, 4% Diabetes mellitus: EVAR group, 8%; OSR group, 15%
Methods	Data collection: data was prospectively collected Analysis: In multivariate analysis, adjustments were made for age, gender, diabetes, renal failure, coronary artery disease (i.e., history of angina or myocardial infarction or stress-induced ischemia), history of cerebrovascular disease, hypertension, β blockers, statins, and propensity scores.
Intervention	EVAR

Full citation	**Feringa HHH, Karagiannis S, Vidakovic R, et al. (2007) Comparison of the Incidences of Cardiac Arrhythmias, Myocardial Ischemia, and Cardiac Events in Patients Treated With Endovascular Versus Open Surgical Repair of Abdominal Aortic Aneurysms. Am J Cardiol. 100:1479-1484**
Comparator	OSR
Outcomes	Long-term mortality, long-term cardiac events
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk 1.2. Were cohorts from the same place? Low risk 1.3. Is the definition of AAA the same across cohorts? Low risk – differences are unlikely **Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk 2.3. Does study control appropriately for AAA characteristics? High risk - none 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – no explanation of the long-term follow-up plan **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – none reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk 4.3. Have different methods been compared within the study? High risk – none reported **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? High risk – no checks reported 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – 11 covariates and 18–69 events 6.2 Were interactions between treatment and other covariates considered? High risk **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Gupta PK, Ramanan B, Lynch TG, et al. (2012) Endovascular repair of abdominal aortic aneurysm does not improve early survival versus open repair in patients younger than 60 years. Eur J Vasc Endovasc Surg. 43(5):506-12.**
Study details	Study design: Retrospective cohort study Location(s): USA Study period: January 2007 to December 2009 Aim of the study: to compare 30-day outcomes after EVAR and OSR for unruptured infrarenal aortic aneurysm in patients younger than 60 years.
Participants	Sample size: EVAR group, n=369; OSR group, n=282 Inclusion criteria: people **under 60 years** who underwent elective EVAR or OSR repair of infrarenal AAAs were included. Exclusion criteria: not reported Baseline characteristics: Median age (IDQ): EVAR group, 56 (54-58) years; OSR group, 56 (53-58) years Gender: EVAR group, 90.8% male; OSR group, 80.5% male Mean aneurysm diameter: not reported Diabetes on insulin: EVAR group, 5.4%; OSR group, 7.8% Hypertension: EVAR group, 75.9%; OSR group, 76.2% Angina within 1 month: EVAR group, 2.7%; OSR group, 1.1% Cardiac surgery prior: EVAR group, 16.8%; OSR group, 13.8% Congestive heart failure: EVAR group, 1.9%; OSR group, 0.7% Myocardial infarction: EVAR group, 1.1%; OSR group, 1.1%
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). The NSQIP collects data on 136 variables and requires hospitals to provide complete 30-day follow-up on at least 95% of patients. Investigators also examined inpatient records and outpatients charts and attempted to contact patients by telephone. If no response is obtained, the Social Security Death Index and the National Obituary Archives are queried to investigate the potential of a death. Analysis: Stepwise multivariate logistic regression was performed. Authors stated that the type of aortic repair (EVAR or OSR) was forced into the logistic regression analysis. Both the C-statistic and the p-value for the Hosmer-Lemeshow goodness of fit test were obtained to determine if there was a satisfactory fit of the model.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality and morbidity (including The latter included deep wound infection, organ space infection, wound dehiscence, pneumonia, reintubation, on ventilator >48 h, pulmonary embolus, deep venous thrombosis, renal insufficiency, acute renal failure, stroke, coma, peripheral

Full citation	**Gupta PK, Ramanan B, Lynch TG, et al. (2012) Endovascular repair of abdominal aortic aneurysm does not improve early survival versus open repair in patients younger than 60 years. Eur J Vasc Endovasc Surg. 43(5):506-12.**
	nerve deficiency, graft/prosthesis failure, cardiac arrest, myocardial infarction, transfusion >4 units packed red blood cells (PRBCs) within 72 h, sepsis, and septic shock or return to the operating room)
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – cohorts were derived from the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? Low risk – all aneurysms were unruptured infrarenal AAA. **Confounding** 2.1. Does study control appropriately for demographics? High risk – model appears to have not adjusted for age or gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – model adjusted for relevant comorbidities. 2.3. Does study control appropriately for AAA characteristics? High risk – no indication that AAA characteristics were adjusted for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – participants were identified data from detailed registries supplemented by medical record examination. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcome were obtained by supplementing registry with information from medical records, communication with patients and direct utilisation of the Social Security Death Index and the National Obituary Archives. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – outcome were obtained by supplementing registry with information from medical records, communication with patients and direct utilisation of the Social Security Death Index and the National Obituary Archives. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk - Both the C-statistic and the p-value for the Hosmer–Lemeshow goodness of fit test were performed. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors highlight that the NSQIP database required hospitals to provide complete 30-day follow-up on at least 95% of patients. 4.3. Have different methods been compared within the study? High risk - different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A

Full citation	Gupta PK, Ramanan B, Lynch TG, et al. (2012) Endovascular repair of abdominal aortic aneurysm does not improve early survival versus open repair in patients younger than 60 years. Eur J Vasc Endovasc Surg. 43(5):506-12.
	Analysis – simple multivariable models 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** partially applicable

Full citation	Hicks CW, Black JH, Arhuidese I, et al. (2015) Mortality variability after endovascular versus open abdominal aortic aneurysm repair in a large tertiary vascular center using a Medicare-derived risk prediction model. J Vasc Surg. 61(2):291-7
Study details	Study design: retrospective cohort study Location(s): USA Study period: November 2003 to August 2012 Aim of the study: to compare the perioperative morbidity and mortality observed with EVAR vs open AAA repair at a single large tertiary vascular centre with the predicted mortality as generated by application of the Giles risk stratification model.
Participants	Sample size: EVAR group, n=214; OSR group, n= 83 Inclusion criteria: all people who underwent elective infrarenal repair of AAA at a single tertiary institution were included. Exclusion criteria: people with connective tissue disorders, inflammatory aneurysms, and ruptured aneurysms were excluded. Baseline characteristics: Mean age (SD): EVAR group, 74.3 (0.54) years; OSR group, 69.2 (0.86) years Gender: EVAR group, 80.8% male; OSR group, 75.9% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 17.8%; OSR group, 14.5% Hypertension: EVAR group, 84.1%; OSR group, 90.4% Dyslipidaemia: EVAR group, 77.6%; OSR group, 75.9% COPD: EVAR group, 25.7%; OSR group, 30.1% Congestive heart failure: EVAR group, 11.2%; OSR group, 4.8% Cancer: EVAR group, 25.2%; OSR group, 21.7%

Full citation	**Hicks CW, Black JH, Arhuidese I, et al. (2015) Mortality variability after endovascular versus open abdominal aortic aneurysm repair in a large tertiary vascular center using a Medicare-derived risk prediction model. J Vasc Surg. 61(2):291-7**
Methods	Data collection: the electronic health records of patients who underwent repair of unruptured infrarenal AAA were retrospectively reviewed by two independent study team members to collect data on patient demographics, symptoms, comorbidities, surgical technique, postoperative outcomes, and mortality. Patient comorbidities were abstracted based on physician documentation within the electronic medical record. Analysis: multivariable logistic regression was performed accounting for age, gender, and comorbidities (congestive heart failure, COPD, coronary artery disease, and chronic renal insufficiency). It is unclear whether stepwise regression was performed to enter covariates into the regression model.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – study cohorts all received treatment at the same tertiary centre. 1.3. Is the definition of AAA the same across cohorts? Low risk – all participants had unruptured infrarenal AAA. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controls for demographic variables including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – investigators controlled for a broad variety of comorbidities. 2.3. Does study control appropriately for AAA characteristics? High risk – there is no indication that AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – participants were identified by reviewing medical records. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed by reviewing medical records. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long-term outcomes assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – it is not apparent that checks were performed on model specification/fit. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – although authors do not mention missing data, the nature in which data were collected is unlikely to have introduced any bias. 4.3. Have different methods been compared within the study? High risk – different methods were not compared.

Full citation	**Hicks CW, Black JH, Arhuidese I, et al. (2015) Mortality variability after endovascular versus open abdominal aortic aneurysm repair in a large tertiary vascular center using a Medicare-derived risk prediction model. J Vasc Surg. 61(2):291-7**
	Analysis – matching 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk - number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions with covariates were considered. **Overall risk of bias:** Moderate risk **Directness:** directly applicable

Full citation	**Hua HT, Cambria RP, Chuang SK et al. (2005) Early outcomes of endovascular versus open abdominal aortic aneurysm repair in the National Surgical Quality Improvement Program-Private Sector (NSQIP-PS). J Vasc Surg. 2005 Mar;41(3):382-9**
Study details	Study design: retrospective cohort study Location(s): USA Study period: January 2000 to October 2003 Aim of the study: to compare early outcomes EVAR versus OSR in a contemporary large, multicentre cohort.
Participants	Sample size: EVAR group, n=460; OSR group, n=582 Inclusion criteria: people who underwent elective repair of infrarenal AAA were included. Exclusion criteria: not reported Baseline characteristics: Mean age: EVAR group, 74.0 years; OSR group, 71.2 years Gender: EVAR group, 84.6% male; OSR group, 79.6% male Mean aneurysm diameter: not reported Congestive heart failure: EVAR group, 2.83%; OSR group, 2.0% Myocardial infarction: EVAR group, 1.79%; OSR group, 0.78% Hypertension: EVAR group, 69.6%; OSR group, 74.5% Stroke with deficit: EVAR group, 7.39%; OSR group, 5.32% Diabetes: EVAR group, 12.7%; OSR group, 11.0%

Full citation	Hua HT, Cambria RP, Chuang SK et al. (2005) Early outcomes of endovascular versus open abdominal aortic aneurysm repair in the National Surgical Quality Improvement Program-Private Sector (NSQIP-PS). J Vasc Surg. 2005 Mar;41(3):382-9
	COPD: EVAR group, 25.4%; OSR group, 17.9% Acute renal failure: EVAR group, 0.43%; OSR group, 0.69%
Methods	Data collection: procedure codes were obtain the data files of patients who underwent elective AAA repair by querying a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). No further details were provided. Analysis: multivariate logistic regression was performed. Only variables that were found to be significant on univariate analysis (p value < 0.05) were entered into the logistic regression model
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, and adverse events (cardiac, pulmonary, renal, neurologic, infectious, and hematologic complications), and length of stay
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants were identified from the same national database. 1.3. Is the definition of AAA the same across cohorts? Low risk – there is no indication that the definition of AAA was different across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – the study controls for age and gender 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – the study controls for a broad range of relevant comorbidities. 2.3. Does study control appropriately for AAA characteristics? High risk – no AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate – participants were identified using a detailed surgical registry with diagnosis and procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed using a detailed surgical registry. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long-term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks specified. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors did not state how missing data was handled. 4.3. Have different methods been compared within the study? High risk – different methods were not compared.

Full citation	**Hua HT, Cambria RP, Chuang SK et al. (2005) Early outcomes of endovascular versus open abdominal aortic aneurysm repair in the National Surgical Quality Improvement Program-Private Sector (NSQIP-PS). J Vasc Surg. 2005 Mar;41(3):382-9**
	Analysis – matching 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were compared. **Overall risk of bias:** High risk **Directness: directly applicable**

Full citation	**Huang Y, Gloviczki P, Oderich GS, et al. (2015) Outcome after open and endovascular repairs of abdominal aortic aneurysms in matched cohorts using propensity score modeling. J Vasc Surg; 62(2):304-11.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: January 2000 to December 2011 Aim of the study: to compare the outcomes of EVAR and OSR of unruptured infrarenal AAA.
Participants	Sample size of matched cohort: EVAR group, n=558; OSR group, n=558 Inclusion criteria: people who underwent EVAR or OSR for unruptured infrarenal AAA were included. Indications for repair of the asymptomatic AAA included rapid growth of the aneurysm (>0.5 cm/y), AAA size ≥5.5 cm in diameter, and smaller AAAs with enlarged (>3 cm) associated iliac aneurysm. Exclusion criteria: people with symptomatic or ruptured AAAs, those with concomitant renal revascularizations, and those who had inter-renal or suprarenal aortic clamping were excluded. Baseline characteristics: Mean age (SD): EVAR group 74 (7.1) years; OSR group, 72 (8.0) years Gender: EVAR group, 86% male; OSR group, 86% male Mean aneurysm diameter (SD): EVAR group, 5.7 (1.0) cm; OSR group, 5.9 (1.2) cm History of cancer: EVAR group, 22%; OSR group, 21%

Full citation	**Huang Y, Gloviczki P, Oderich GS, et al. (2015) Outcome after open and endovascular repairs of abdominal aortic aneurysms in matched cohorts using propensity score modeling. J Vasc Surg; 62(2):304-11.**
Methods	Data collection: data were collected from an aortic registry of the Mayo Clinic (a non-profit academic medical centre). This retrospectively recorded database included data on demographics, maximum external diameter comorbidities, procedures, mortalities, complications, reinterventions, and ruptures. Follow-up information was obtained from the medical records and mailing questionnaires. The patient's vital status was established from charts, mailing questionnaires, death certificate, or autopsy report. Analysis: a propensity score using logistic regression was estimated considering predictors of gender, the year of intervention, and SVS comorbidity scores of cardiac, renal, pulmonary, hypertension, and age. The C statistic was used to assess goodness to fit. Subsequently, propensity score-matched cohorts of patients treated by EVAR and OSR were created. In the matched cohort, the propensity score and surgical risk were included as covariates in all models (logistic and Cox). In-hospital/30-day events were assessed using logistic regression, whereas longer-term outcomes were assessed using Cox regression.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality, 30-day mortality, long-term mortality, hospital length of stay, length of stay in ICU, adverse events and reinterventions
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – recruitment over ≥5 yrs, but year of operation controlled for in analysis 1.2. Were cohorts from the same place? Low risk – both cohorts were derived from an aortic registry of the Mayo Clinic. 1.3. Is the definition of AAA the same across cohorts? Low risk – all participants had unruptured infrarenal AAA. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – investigators controlled for demographic variables, including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – investigators controlled for a broad variety of comorbidities. 2.3. Does study control appropriately for AAA characteristics? Low risk – study controlled for aneurysm size (in the multivariate analyses). 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – participants were identified using registry data and examination of medical records. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – follow-up information was obtained from the medical records and mailing questionnaires. 3.3. Is method of data collection likely to record long-term outcomes accurately? High risk – the patient's vital status was established from charts, mailing questionnaires, death certificate, or autopsy report, with no attempt to use administrative data. **Analysis – general**

Full citation	Huang Y, Gloviczki P, Oderich GS, et al. (2015) Outcome after open and endovascular repairs of abdominal aortic aneurysms in matched cohorts using propensity score modeling. J Vasc Surg; 62(2):304-11.
	4.1. Were any checks conducted on model specification and/or fit? Low risk – C statistic was used to assess goodness to fit. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors state that “rates of late mortality, complication, reintervention, and rupture might be underestimated in this retrospective study because of missing adverse events and loss of follow-up”. 4.3. Have different methods been compared within the study? Moderate - different methods were compared but both relied on the same assumption about selection. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? High risk – no matching algorithm is reported. 5.2. Was overlap / common support appropriately assessed? High risk – no assessment reported. 5.3. Has balancing of the covariates been demonstrated? Moderate - Conventional hypothesis tests were performed, with no evidence of significant differences. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Lee WA, Carter JW, Upchurch G, et al. (2004) Perioperative outcomes after open and endovascular repair of intact abdominal aortic aneurysms in the United States during 2001. J Vasc Surg. 39 (3):491-6.
Study details	Study design: retrospective cohort study Location(s): USA Study period: January to December 2001 Aim of the study: to compare the perioperative outcomes of endovascular and open surgical AAA repair in an unselected sample of patients in a single calendar year using a national administrative database.
Participants	Sample size: EVAR group, n=2,565; OSR group, n=4,607 Inclusion criteria: people who underwent repair of unruptured infrarenal AAAs were included. Exclusion criteria: people younger than 50 years and those with secondary diagnostic codes for ruptured AAA, aortic dissection, thoracic or thoracoabdominal aortic aneurysm, coarctation of the aorta, Marfan syndrome and other congenital anomalies, gonadal dysgenesis, Turner syndrome , and polyarteritis nodosa were excluded. Baseline characteristics:

Full citation	Lee WA, Carter JW, Upchurch G, et al. (2004) Perioperative outcomes after open and endovascular repair of intact abdominal aortic aneurysms in the United States during 2001. J Vasc Surg. 39 (3):491-6.
	Mean age (SD): EVAR group, 73.4 (7.8) years; OSR group, 71.9 (7.7) years Gender: EVAR group, 84.4% male; OSR group, 78.1% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 11%; OSR group, 11% Hypertension: EVAR group, 57%; OSR group, 53% Renal insufficiency: EVAR group, 25%; OSR group, 29% Ischaemic heart disease: EVAR group, 20%; OSR group, 14% Cerebrovascular occlusive disease: EVAR group, 0.7%; OSR group, 0.4%
Methods	Data collection: participants were identified and data were obtained using ICD9 diagnostic and procedure codes to query a national administrative database (the National Inpatient Sample) Analysis: multivariate logistic regression was performed. Only variables that were found to be significant on univariate analysis (significance level not specified) were entered into the logistic regression model. No further details were provided.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality, adverse events, and discharge to home.
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period 1.2. Were cohorts from the same place? Low risk – cohorts were drawn from the same national database. 1.3. Is the definition of AAA the same across cohorts? High risk – procedure codes used do not distinguish between infrarenal and complex AAA; likely to be many more complex cases in OSR cohort, given era. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controls for age and sex 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – study controls for number of comorbidities rather than specific comorbidities of interest. 2.3. Does study control appropriately for AAA characteristics? High risk – study does not control for AAA characteristics. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – an administrative registry with high-level diagnosis and procedure codes was used.

Full citation	Lee WA, Carter JW, Upchurch G, et al. (2004) Perioperative outcomes after open and endovascular repair of intact abdominal aortic aneurysms in the United States during 2001. J Vasc Surg. 39 (3):491-6.
	3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – an administrative registry with high-level diagnosis and procedure codes was used. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long term outcomes were assessed **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks for model specification/fit were performed. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors do not discuss whether there there was any missing data and how this was handled. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were compared **Overall risk of bias:** High risk **Directness:** Directly applicable

Full citation	Liang NL, Reitz KM, Makaroun MS, et al. (2018) Comparable perioperative mortality outcomes in younger patients undergoing elective open and endovascular abdominal aortic aneurysm repair. J Vasc Surg. 2018 May;67(5):1404-1409.e2.
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2003 to 2014 Aim of the study: to compare perioperative and short-term outcomes for EVAR and OSR in younger patients using a large national disease and procedure-specific data set
Participants	Sample size: EVAR group, n=1,928; OSR group, n= 713 Inclusion criteria: people 65 years of age or younger undergoing first-time EVAR or OSR of unruptured infrarenal AAA were included.

Full citation	**Liang NL, Reitz KM, Makaroun MS, et al. (2018) Comparable perioperative mortality outcomes in younger patients undergoing elective open and endovascular abdominal aortic aneurysm repair. J Vasc Surg. 2018 May;67(5):1404-1409.e2.**
	Exclusion criteria: pararenal EVAR chimney or fenestrated operations, OSRs involving suprarenal clamping and pararenal or thoracoabdominal aneurysms, repairs performed for isolated iliac aneurysm were excluded. Furthermore, EVAR patients who were deemed medically unfit for OSR were excluded. Baseline characteristics: Median age (IQR): EVAR group, 62 (59-64) years; OSR group, 61 (58-64) years Gender: EVAR group, 88.0% male; OSR group, 85.3% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 21.1%; OSR group, 13.5% Hypertension: EVAR group, 80.7%; OSR group, 79.5% Coronary artery disease: EVAR group, 28.9%; OSR group, 24.1% Heart failure: EVAR group, 7.2%; OSR group, 4.6% Emphysema: EVAR group, 27.3%; OSR group, 28.7% History of CABG: EVAR group, 31.4%; OSR group, 27.3%
Methods	Data collection: investigators used data from the national Vascular Quality Initiative (VQI) EVAR and OSR registries to identify relevant participants and assess their postoperative outcomes. Note: authors report that details of patients treated by EVAR and OSR were recorded in 2 distinct registries but the outcome measure variables were consistent across both data sets. Analysis: Inverse probability weighting was performed using propensity scores. Initially, the propensity for receiving treatment was fit using logistic regression to adjust for clinical and comorbid characteristics between the EVAR and OSR groups. Covariates were included following a stepwise inclusion method, or forced into the model if deemed clinically important. The comparability of the two initial cohorts was confirmed by examining distributions of propensity scores. An inverse probability of treatment weight based on the propensity score was then calculated for each subject and applied to both cohorts; stabilised weights were used to correct for outliers. Propensity weighted mortality, adverse events and reintervention rates were then calculated and compared between groups.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, 30-day morbidity, adverse events and reintervention
Study Appraisal using NICE's bespoke risk of bias	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – although 2 different surgical databases were used, both databases were from the same country and collected the same types of data. 1.3. Is the definition of AAA the same across cohorts? Low risk –similar definitions were used across cohorts. **Confounding**

Full citation	Liang NL, Reitz KM, Makaroun MS, et al. (2018) Comparable perioperative mortality outcomes in younger patients undergoing elective open and endovascular abdominal aortic aneurysm repair. J Vasc Surg. 2018 May;67(5):1404-1409.e2.
assessment tool	2.1. Does study control appropriately for demographics? Low risk – study controls for demographic variables including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – study controls for a broad range of relevant comorbidities. 2.3. Does study control appropriately for AAA characteristics? Low risk – investigators controlled for AAA characteristics including diameter. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – patients identified using detailed surgical registries. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes assessed using detailed surgical registries. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – 1 year survival was based on Social Security Death Index-linked death records. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no tests reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors stated that analysis of the data is also complicated by significant amounts of missing data for 1-year follow-up. It is not clear if this was adjusted for. 4.3. Have different methods been compared within the study? High – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? High risk – no details were provided about the matching algorithm used. 5.2. Was overlap / common support appropriately assessed? Low risk - Checks were performed by plotting propensity distribution densities for each treatment arm. 5.3. Has balancing of the covariates been demonstrated? Low risk – standardised differences were reported with none of the weighted differences exceeding 0.1. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** partially applicable

Full citation	**Locham S, Lee R, Nejim B, et al. (2017) Mortality after endovascular versus open repair of abdominal aortic aneurysm in the elderly. J Surg Res. 215:153-159.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2011 to 2014 Aim of the study: determine the predictors of 30-d mortality after AAA repair in elderly population
Participants	Sample size: EVAR group, n=3,869; OSR group, n=360 Inclusion criteria: patients 70 years or over with unruptured infrarenal AAA who underwent EVAR or OSR were included. Exclusion criteria: not reported Baseline characteristics: Mean age (SD): EVAR group, 78.4 (5.6) years; OSR group, 76.8 (4.8) years Gender: EVAR group, 79.4% male; OSR group, 68.9% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 15.5%; OSR group, 11.7% History of COPD: EVAR group, 18.6%; OSR group, 19.7% History of chronic heart failure: EVAR group, 1.6%; OSR group, 1.9% Hypertension: EVAR group, 80.9%; OSR group, 82.2% Progressive renal insufficiency: EVAR group, 49.9%; OSR group, 46.6% Renal failure þ dialysis: EVAR group, 1.2%; OSR group, 0.6%
Methods	Data collection: investigators identified participants and obtained data on their outcomes by querying the American College of Surgeons version of the National Surgical Quality Improvement Program (ACS-NSQIP) database which contained information such as the proximal and distal extents of the aneurysm, specific operative characteristics, and 30-d postoperative vascular outcomes in both inpatient and outpatient settings. The selected cohort was later linked/merged with the general version of the NSQIP to obtain information on demographics and comorbidities. Analysis: multivariate logistic regression was performed to explore risk factors associated with 30-day mortality. It is unclear how risk factors were selected into the logistic regression model. The final model was evaluated by Hosmere and Lemeshow test and area under the curve.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality
Study Appraisal using NICE's	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical database

Full citation	**Locham S, Lee R, Nejim B, et al. (2017) Mortality after endovascular versus open repair of abdominal aortic aneurysm in the elderly. J Surg Res. 215:153-159.**
bespoke risk of bias assessment tool	1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA was the same across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controls for demographic variables, including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a good range of individual comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? Low risk – authors stated that they controlled for aneurysm diameter and distal extent of the aneurysm in the logistic regression model. 2.4. Could any adjustment variables have been affected by the intervention? High risk - perioperative transfusion included in model. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – a detailed surgical registry was used with procedure and diagnosis codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed using a detailed surgical registry was used with procedure and diagnosis codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long-term outcomes were assed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – model specification/fit was assessed using the Hosmer–Lemeshow test as well as the C-statistic. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – there is no indication that missing data were taken into account. 4.3. Have different methods been compared within the study? High risk - different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** Moderate risk **Directness:** directly applicable

Full citation	**Malas M, Arhuidese I, Qazi U, et al. (2014) Perioperative mortality following repair of abdominal aortic aneurysms: application of a randomized clinical trial to real-world practice using a validated nationwide data set. JAMA Surg. 149(12):1260-5.**
Study details	Study design: Location(s): USA Study period: January 2005 to December 2011 Aim of the study: to compare 30-day mortality from are recent trial comparing EVAR and OSR with data from a national registry and to assess temoral trends in perioperative mortality
Participants	Sample size: EVAR group, n=15,807; OSR group, n=5,308 Inclusion criteria: people who underwent EVAR or OSR of unruptured isolated infrarenal aortic aneurysms or aortoilliac aneurysms were included. Exclusion criteria: people with ruptured or symptomatic aneurysms were excluded. Baseline characteristics: Mean age (SD): EVAR group, 74.2 (8.4) years; OSR group, 71.1 (8.2) years Gender: EVAR group, 82.0% male; OSR group, 74.4% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 15.3%; OSR group, 12.8% Hypertension: EVAR group, 80.3%; OSR group, 82.5% COPD: EVAR group, 19.3%; OSR group, 19.1% Myocardial infarction: EVAR group, 1.0%; OSR group, 1.3% Angina: EVAR group, 1.9%; OSR group, 1.6% Chronic heart failure: EVAR group, 1.4%; OSR group, 0.8% Renal failure: EVAR group, 1.2%; OSR group, 0.8% Malignancy: EVAR group, 0.6%; OSR group, 0.2%
Methods	Data collection: participants were identified using diagnosis and procedure codes, and data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). Analysis: multivariate logistic regression models were built to identify predictors of outcomes. Authors state that likelihood ratio tests were used to test the predictive value of each covariate in the build-up of the final model. Predictive covariates and clinically relevant risk factors were included in the final model. Sensitivity analyses were carried out by removing variables with missing data and comparing it with the complete values in the model.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality

Full citation	**Malas M, Arhuidese I, Qazi U, et al. (2014) Perioperative mortality following repair of abdominal aortic aneurysms: application of a randomized clinical trial to real-world practice using a validated nationwide data set. JAMA Surg. 149(12):1260-5.**
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA was similar across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – the study controls for gender, but not age. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a good range of individual comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – no AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – participants identified a detailed surgical registry with procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data obtained from a detailed surgical registry with procedure codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk - authors highlight that data were missing in a non-systematic manner, and sensitivity analyses showed that results were consistent with the complete case analyses. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk

Full citation	**Malas M, Arhuidese I, Qazi U, et al. (2014) Perioperative mortality following repair of abdominal aortic aneurysms: application of a randomized clinical trial to real-world practice using a validated nationwide data set. JAMA Surg. 149(12):1260-5.**
	Directness: directly applicable

Full citation	**Nguyen BN, Neville RF, Rahbar R, et al. (2013) Comparison of outcomes for open abdominal aortic aneurysm repair and endovascular repair in patients with chronic renal insufficiency. Send to Ann Surg. 258(3):394-9.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2005 to 2010 Aim of the study: compare outcomes of EVAR versus OSR in patients with chronic renal insufficiency.
Participants	Sample size: EVAR group, n=3,886; OSR group, n=1,256 Inclusion criteria: people with chronic renal insufficiency (pre-treatment eGFR of less than 60 mL/min/1.73 m^2) who underwent EVAR or OSR for unruptured infrarenal AAA were included. Note: some patients underwent emergency treatment (likely to be attributed to symptomatic aneurysms). Exclusion criteria: ruptured AAA, fenestrated EVAR, combined thoracic endovascular aneurysm repair, open repairs that included any visceral bypasses, or additional procedures for lower extremities were excluded. Baseline characteristics: Mean age (SD): EVAR group, 77.5 (7.6) years; OSR group, 74.4 (7.7) years Gender: EVAR group, 78.5% male; OSR group, 67.4% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 16.5%; OSR group, 13.3% Coronary artery disease: EVAR group, 1.1%; OSR group, 1.6% Congestive heart failure: EVAR group, 2.0%; OSR group, 1.4% COPD: EVAR group, 18.9%; OSR group, 19.8% Acute renal insufficiency: EVAR group, 0.6%; OSR group, 0.7%
Methods	Data collection: people who underwent elective repair of infrarenal AAA were identified using procedure codes to query a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). Upon identification of relevant population, data files were reviewed to identify people who had chronic renal insufficiency, who were selected for inclusion in the study. Patient demographics, preoperative comorbidity data, and outcome data were extracted from the NSQIP database.

Full citation	**Nguyen BN, Neville RF, Rahbar R, et al. (2013) Comparison of outcomes for open abdominal aortic aneurysm repair and endovascular repair in patients with chronic renal insufficiency. Send to Ann Surg. 258(3):394-9.**
	Analysis: Multivariate logistic regression was performed to explore whether treatment type was a significant predictor of postoperative outcomes. Authors stated that multivariate regression corrected for all preoperative variables which were found to be significantly different between groups.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, and adverse events (renal dysfunction, pulmonary complications, and cardiovascular events)
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA was similar across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controlled for age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – study controlled for a limited number of comorbidities. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – participants were identified using detailed surgical registries with diagnosis and procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – detailed surgical registries were used with diagnosis and procedure codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long-term outcomes were assessed **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks on model specification were performed. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors did not discuss missing data. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A

Full citation	**Nguyen BN, Neville RF, Rahbar R, et al. (2013) Comparison of outcomes for open abdominal aortic aneurysm repair and endovascular repair in patients with chronic renal insufficiency. Send to Ann Surg. 258(3):394-9.**
	5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** partially applicable

Full citation	**Sugimoto M, Koyama A, Niimi K, et al. (2017) Long-term Comparison of Endovascular and Open Repair of Abdominal Aortic Aneurysms: Retrospective Analysis of Matched Cohorts with Propensity Score. Ann Vasc Surg.43:96-103.**
Study details	Study design: retrospective cohort study Location(s): Japan Study period: June 2007 to October 2014 Aim of the study: to compare our long-term outcomes of EVAR and OSR, eliminating the differences of patients' backgrounds with propensity score matching.
Participants	Sample size of unmatched cohort: EVAR group, n=386; OSR group, n=351 Sample size of matched cohort: EVAR group, n=157; OSR group, n=157 Inclusion criteria: all patients who underwent EVAR or OSR for unruptured infrarenal AAA >5.0cm who had over 1-year follow-up data available were included. Exclusion criteria: patients with suprarenal, pararenal, mycotic or ruptured were excluded. Baseline characteristics (of matched cohort): Mean age (range): EVAR group, 75 (70-79) years; OSR group, 74 (71-79) years Gender: EVAR group, 86.6% male; OSR group, 86.0% male Mean aneurysm diameter (SD): EVAR group, 5.34 (0.88) cm OSR group, 5.34 (1.05) cm Diabetes: EVAR group, 12.7%; OSR group, 10.2% Hypertension: EVAR group, 68.2%; OSR group, 72.6% Coronary artery disease: EVAR group, 32.5%; OSR group, 35.0% Stroke: EVAR group, 12.7%; OSR group, 10.8%

Full citation	**Sugimoto M, Koyama A, Niimi K, et al. (2017) Long-term Comparison of Endovascular and Open Repair of Abdominal Aortic Aneurysms: Retrospective Analysis of Matched Cohorts with Propensity Score. Ann Vasc Surg.43:96-103.**
	Malignancy: EVAR group, 22.3%; OSR group, 19.7%
Methods	Data collection: participants were identified through review of medical records and data on comorbidities, AAA diameters and postoperative outcomes were collected from patients' medical records and assembled in a dedicated database. Analysis: Matching according to propensity scores was performed. Propensity scores were calculated using multivariate regression considering the following variables: age, gender, hypertension, coronary arterial disease, COPD, diabetes, stroke, malignancy, haemodialysis, ejection fraction, preoperative serum creatinine, and $FEV_{1.0}$%. Upon matching the cohorts, univariate analyses (t-tests, Mann-Whitney U-test, chi-squared test, and Fisher's exact test) were performed.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, mortality at 1 year, late adverse events (occurring 3 to 12 months), and late reinterventions (occurring 3 to 12 months)
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – all participants were treated at the same medical centre. 1.3. Is the definition of AAA the same across cohorts? High risk – earlier publication of the same cohort (prior to matching) notes that 'cases of juxta-renal AAA that were treated via fenestrated EVAR or the chimney technique were excluded' whereas OSR on the same anatomy would not have been. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – both age and gender were controlled for in the analyses. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – good range of relevant comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – although sample size data were controlled for, authors did not control for aneurysm size or any other aneurysm characteristic in their matching and subsequent analyses. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – participants were identified by reviewing medical records. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – medical records were reviewed 3.3. Is method of data collection likely to record long-term outcomes accurately? High risk – only medical records were used without any confirmation from other data sources. **Analysis – general**

Full citation	**Sugimoto M, Koyama A, Niimi K, et al. (2017) Long-term Comparison of Endovascular and Open Repair of Abdominal Aortic Aneurysms: Retrospective Analysis of Matched Cohorts with Propensity Score. Ann Vasc Surg.43:96-103.**
	4.1. Were any checks conducted on model specification and/or fit? High risk – authors did not provide any details about checks for model specification/fit. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors stated that only patients with follow-up data of more than 1 year were included. Authors reported that the few patients who were lost to follow-up were censored at their last visits. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? High risk – no matching algorithm was provided or discussed. 5.2. Was overlap / common support appropriately assessed? High risk – no overlap was assessed. 5.3. Has balancing of the covariates been demonstrated? Moderate risk – conventional hypothesis tests were performed showing no evidence of significant differences between groups. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Zabrocki L, Marquardt F, Albrecht K, et al. (2018) Acute kidney injury after abdominal aortic aneurysm repair: current epidemiology and potential prevention. Int Urol Nephrol. 50(2):331-337.**
Study details	Study design: retrospective cohort study Location(s): Germany Study period: 2007 to 2011 Aim of the study: to evaluate whether patients receiving EVAR or OSR differed with respect to frequency and severity of acute kidney injuryafter adjusting by propensity score matching.
Participants	Sample size: EVAR group, n=91; OSR group, n=91 Inclusion criteria: all patients who underwent repair of unruptured or ruptured infrarenal AAAs at a single tertiary centre were included. Note: EVAR was offered as a first option to patients considered high risk for OAR due to their comorbidities. Exclusion criteria: patients with thoracoabdominal aneurysm, supra-, juxta- or pararenal AAA, ruptured AAA, repair of recurrent AAA, end-stage renal disease and AKI just prior to AAA repair were excluded. Baseline characteristics (of matched cohort):

Full citation	**Zabrocki L, Marquardt F, Albrecht K, et al. (2018) Acute kidney injury after abdominal aortic aneurysm repair: current epidemiology and potential prevention. Int Urol Nephrol. 50(2):331-337.**
	Mean age (SD): EVAR group, 74 (7) years; OSR group, 72 (7) years Gender: EVAR group, 11.2% male; OSR group, 16.8% male Mean aneurysm diameter: EVAR group, 6.3 (1.1) cm; OSR group, 6.2 (1.1) cm Diabetes: EVAR group, 14%; OSR group, 15% Hypertension: EVAR group, 76%; OSR group, 78% Hypercholesterolaemia: EVAR group, 36%; OSR group, 33% Severe cardiac disease: EVAR group, 42%; OSR group, 38% Severe pulmonary disease: EVAR group, 22%; OSR group, 29% History of cancer: EVAR group, 17%; OSR group, 20%
Methods	Data collection: patients were identified using mandatory administrative and reimbursement ICD10 codes in combination with procedure codes from hospital databases. Upon identifying relevant patients, Data were obtained from the electronic hospital records, 30 day and 3-month follow-up data after AAA repair from hospital or primary care physician's records. Analysis: Propensity score matching was used to control for substantial differences in demographic factors and comorbidity due to non-random assignment of patients to EVAR or OSR. Propensity scores were calculated using multivariate logistic regression considering the following variables: age, gender, urgent admission (likely to be symptomatic), diabetes, hypertension, severe cardiac and lung disease, history of cancer, CKD, and diameter of AAA. Nearest-neighbour matching was subsequently used to match patients. To ensure close matches, investigators required that the propensity score of EVAR and OAR patients agreed on five decimals. The goodness-of-fit of the propensity score was assessed by C-statistics and Hosmer–Lemeshow test. In addition to matched comparisons between groups, authors also performed multivariate logistic regression to explore risk factors (including type of treatment) associated with acute kidney injury.
Intervention	EVAR
Comparator	OSR
Outcomes	Acute kidney injury
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants were treated at the same tertiary medical centre. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA appears to be similar across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – the study controls for demographics including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a good range of comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? Low risk – AAA characteristics (diameter) were controlled for.

Full citation	**Zabrocki L, Marquardt F, Albrecht K, et al. (2018) Acute kidney injury after abdominal aortic aneurysm repair: current epidemiology and potential prevention. Int Urol Nephrol. 50(2):331-337.**
	2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – diagnosis and procedure codes, supplemented by medical record review, were used to identify relevant participants. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – perioperative outcomes were assessed by reviewing medical records. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A risk – no long-term outcomes assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – the C-statistic and Hosmer–Lemeshow test were used to assess model specification/fit 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – although authors do not discuss missing data, the way in which data was collected minimises the risk of missing data on outcomes and relevant covariates. 4.3. Have different methods been compared within the study? Moderate risk - different methods were compared but they relied on the same assumption about selection. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Low risk – nearest neighbour matching was performed as mentioned above. 5.2. Was overlap / common support appropriately assessed? High risk – no assessment reported. 5.3. Has balancing of the covariates been demonstrated? Moderate risk - conventional hypothesis tests were performed, with no evidence of significant differences between groups. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** Moderate risk **Directness:** directly relevant

E.2 All AAAs

Full citation	**Casey K, Hernandez-Boussard T, Mell MW, Lee JT. (2013) Differences in readmissions after open repair versus endovascular aneurysm repair. J Vasc Surg. 57(1):89-95**
Study details	Study design: retrospective cohort study

Full citation	**Casey K, Hernandez-Boussard T, Mell MW, Lee JT. (2013) Differences in readmissions after open repair versus endovascular aneurysm repair. J Vasc Surg. 57(1):89-95**
	Location(s): USA Study period: 2003 to 2008 Aim of the study: to determine reasons for all-cause readmissions within the first year after open repair and EVAR
Participants	Sample size: EVAR group, n=9,356; OSR group, n= 6,380 Inclusion criteria: adults who underwent open repair or endovascular repair of infrarenal AAA. Patients were identified using International Classification of Diseases, Ninth Revision (ICD-9) diagnosis codes for intact AAA. All patients at risk for 1-year readmission after elective AAA repair were included. Exclusion criteria: patients identified with ICD-9 codes for ruptured AAA. Age <40 years. Age >90 years. Baseline characteristics: Mean age (SD): EVAR group, 75.0 (8.3) years; OSR group, 72.1 (9.0) years Gender: EVAR group, 84.7% male; OSR group, 75.6% male Mean aneurysm diameter: not reported Private insurance: EVAR group, 15.3%; OSR group, 20.8% Charlson Index ≥2: EVAR group, 51.7%; OSR group, 55.5% Obesity: EVAR group, 7.2%; OSR group, 7.0% Congestive heart failure: EVAR group, 9.8%; OSR group, 14.1% Complicated DM: EVAR group, 1.7%; OSR group, 1.2% Peripheral vascular disease: EVAR group, 98.8%; OSR group, 98.5% End-stage renal disease: EVAR group, 8.6%; OSR group, 9.1%
Methods	Data collection: The State Inpatient Database (SID), contains a range of data collected from discharge inpatient hospital records, including demographics, ICD-9-Clinical Modification codes for primary and secondary diagnoses and procedures, admission source, length of stay, discharge disposition, inpatient mortality, and hospital characteristics. This database also allows for identification of patient readmissions ≤1 full year in California and attempts to capture patient data characteristics from the readmission, including primary and secondary diagnoses. Multiple readmissions from the same patient ≤1 year counted toward the total number of readmissions for that cohort. Analysis: Modified Cox proportional hazards modelling with adjustment for patient and hospital characteristics was used to adjust for patient mix on readmission rates between the two surgical procedures. Patient characteristics included age, sex, race, insurance status, obesity, complicated diabetes mellitus (DM), complicated hypertension, peripheral vascular disease (PVD), chronic obstructive pulmonary disease (COPD), and end-stage renal disease.
Intervention	EVAR
Comparator	OSR

Full citation	**Casey K, Hernandez-Boussard T, Mell MW, Lee JT. (2013) Differences in readmissions after open repair versus endovascular aneurysm repair. J Vasc Surg. 57(1):89-95**
Outcomes	Readmission rates
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk - ≥5-yr recruitment with no adjustment for year of operation 1.2. Were cohorts from the same place? Low risk – same country 1.3. Is the definition of AAA the same across cohorts? Moderate risk – recruitment methods would have led to all infrarenal and complex AAA included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk - good range of individual comorbidities 2.3. Does study control appropriately for AAA characteristics? High risk - none 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registry 3.2. Is method of data collection likely to record perioperative outcomes accurately? N/A 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – the administrative registry is likely to record readmission rates **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – none reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no details reported, and long-term nature of readmission outcome raise risk of loss to follow-up 4.3. Have different methods been compared within the study? High risk – none reported **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk 6.2 Were interactions between treatment and other covariates considered? High risk - no **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Chang DC, Parina RP, Wilson SE et al. (2015) Survival After Endovascular vs Open Aortic Aneurysm Repairs. JAMA Surg. 150 (12): 1160-6.**
Study details	Study design: Retrospective cohort study Location(s): USA Study period: 2001 to 2009 Aim of the study: to determine long-term outcomes of EVAR vs open repair on a population level.
Participants	Sample size: EVAR group, n=12,239; OSR group, n= 11,431 Inclusion criteria: all people who underwent EVAR or OSR of unruptured AAA at non-federal hospitals in California were included. Note: there is no indication that selection as limited to people with infrarenal AAA, and people with complex aneurysms may have been included. Exclusion criteria: people with concomitant thoracic aneurysm repairs or diagnosis of syphilitic, traumatic, thoracoabdominal, ruptured, or unspecified aortic aneurysms were excluded. Baseline characteristics: Mean age: EVAR group, 75.1 years; OSR group, 72.3 years Gender: EVAR group, 84.4% male; OSR group, 77.5% male Mean aneurysm diameter: not reported Private insurance: EVAR group, 15.4%; OSR group, 20.6% Charlson Comorbidity Index score of ≥2: EVAR group, 57.1%; OSR group, 62.7%
Methods	Data collection: participants were identified and data were collected by querying an administrative database maintained by California state authorities using ICD9 diagnosis and procedure codes. The administrative database was also inked to the Social Security Death Index to obtain mortality records. Analysis: Cox proportional hazards analyses were performed, controlling for repair type (open vs EVAR), age, race/ethnicity, gender, insurance types, Charlson Comorbidity Index score, calendar year, admission type (scheduled vs unscheduled), and hospital type
Intervention	EVAR
Comparator	OSR
Outcomes	Overall long-term mortality, and reinterventions
Study Appraisal using NICE's bespoke risk of bias	**Selection** 1.1. Were cohorts from the same time period? Low risk – ≥5-yr recruitment with adjustments made for year of operation 1.2. Were cohorts from the same place? Low risk – all participants came from California

Full citation	Chang DC, Parina RP, Wilson SE et al. (2015) Survival After Endovascular vs Open Aortic Aneurysm Repairs. JAMA Surg. 150 (12): 1160-6.
assessment tool	1.3. Is the definition of AAA the same across cohorts? Moderate risk – recruitment methods would have led to all infrarenal, complex, and possibly some symptomatic unruptured AAA being included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controlled for age, gender and ethnicity 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a validated comorbidity index (Charlson Comorbidity score) was considered in the regression model. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – an administrative database with high-level diagnosis and procedure codes was used to identify participants 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – an administrative database was used to assess outcomes 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – an administrative database was used with linkage to a reliable mortality registry **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no details of any checks for model fit were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no details provided. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered **Overall risk of bias:** High risk **Directness:** Directly applicable

Full citation	Davenport DL, and Xenos ES., et al. (2013) Deep venous thrombosis after repair of nonruptured abdominal aneurysm. J Vasc Surg. 2013 Mar;57(3):678-683
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2005 to 2009 Aim of the study: to examine venous thromboembolism (VTE) rates, timing, and risk factors after nonruptured open or endoluminal abdominal aortic aneurysm (AAA) repair.
Participants	Sample size: EVAR group, n=8,502; OSR group, n= 3,967 Inclusion criteria: people undergoing elective AAA repair were included. Note: there is no indication that selection as only limited to people with infrarenal AAA. Exclusion criteria: people <18 years old, people with ruptured AAA and people with admissions for trauma were excluded. Baseline characteristics (for all participants): Mean age (SD): 73.2 (8.7) years Gender: 80.2% male Mean aneurysm diameter: not reported Medically treated diabetes: 14.0% Medically treated hypertension:80.8% Prior cardiac stent or operation: 38.3% Chronic heart failure, cardiac arrest or infarct: 3.9% History of stroke, TIA, or hemiplegia: 15.3%
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). Patients who underwent repair procedures for complex AAAs were identified using ICD9 diagnostic codes and CPT procedure codes. Analysis: Stepwise multivariate logistic regression of the occurrence of VTE (in hospital) was performed considering over 40 preoperative variables contained in the NSQIP database. Intraoperative variables (including procedure type, operative duration, wound class, and intraoperative transfusion) were also considered in the regression model.
Intervention	EVAR
Comparator	OSR
Outcomes	Deep vein thrombosis
Study Appraisal	**Selection**

Full citation	Davenport DL, and Xenos ES., et al. (2013) Deep venous thrombosis after repair of nonruptured abdominal aneurysm. J Vasc Surg. 2013 Mar;57(3):678-683
using NICE's bespoke risk of bias assessment tool	1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all study participants were identified using the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? High risk – the definition of OSR includes juxta-/pararenal AAA whereas EVAR codes are limited to infrarenal. **Confounding** 2.1. Does study control appropriately for demographics? High risk – the study did not control for demographic variables. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – a limited number of individual comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? High risk – intra operative variables that may mediate the treatment effect (Intraoperative transfusion) were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – detailed surgical registries were used with diagnosis and procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – administrative codes with high level diagnosis and procedure codes were used to record outcomes. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long-term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks for model specification/fit were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors briefly mention that there was some missing data but do not provide information as to how the missing data was dealt with. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered **Overall risk of bias:** High risk

Full citation	**Davenport DL, and Xenos ES., et al. (2013) Deep venous thrombosis after repair of nonruptured abdominal aneurysm. J Vasc Surg. 2013 Mar;57(3):678-683**
	Directness: directly applicable

Full citation	**Egorova NN, Vouyouka AG, McKinsey JF, et al. (2011) Effect of gender on long-term survival after abdominal aortic aneurysm repair based on results from the Medicare national database. J Vasc Surg. 54(1):1-12.e6**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 1995 to 2006 Aim of the study: to determine the effect of gender on long-term survival differences after EVAR and OSR of AAA under elective and emergency circumstances
Participants	Sample size of matched cohort: EVAR group, n=42,320; OSR group, n=42,320 Inclusion criteria: all people who underwent EVAR or OSR for ruptured or unruptured AAA were included. NB: it is possible that the study sample included people with complex AAA morphologies. Exclusion criteria: the study excluded people with thoracic or thoracoabdominal aneurysms. Baseline characteristics: Mean age: EVAR group - men 76.2 years, women 77.45 years; OSR group - men 74.6 years, women 75.5 years Gender: EVAR group, 82.4% male; OSR group, 76.0% male Mean aneurysm diameter: not reported Coronary comorbidities: EVAR group - men 37.6%, women 30.6%; OSR group - men 49.4%, women 38.1%; Arrhythmia: EVAR group - men 26.6%, women 20.6%; OSR group - men 28.3%, women 22.9%; Coronary heart failure: EVAR group - men 15.0%, women 16.1%; OSR group - men 16.0%, women 17.9%; Renal failure: EVAR group - men 6.3%, women 6.1%; OSR group - men 4.3%, women 4.1%; Diabetes: EVAR group - men 16.9%, women 14.9%; OSR group - men 11.3%, women 10.0%; Cancer: EVAR group - men 8.2%, women 4.8%; OSR group - men 7.1%, women 3.9%; Hypertension: EVAR group - men 72.6%, women 78.2%; OSR group - men 60.3%, women 68.2%; Hyperlipidaemia: EVAR group - men 42.5%, women 41.3%; OSR group - men 23.1%, women 24.5%;
Methods	Data collection: data were retrospectively collected from the Medicare Inpatient Standard Analytical and Denominator databases. Participants were selected through a combination of diagnosis codes (ICD9). The hospitalisation with the first AAA procedure was identified as the index hospitalisation. Pre-index and index hospitalisations were used to assess baseline comorbidities; including all those identified in prior hospitalisations and the chronic conditions reported at the index hospitalisation.

Full citation	**Egorova NN, Vouyouka AG, McKinsey JF, et al. (2011) Effect of gender on long-term survival after abdominal aortic aneurysm repair based on results from the Medicare national database. J Vasc Surg. 54(1):1-12.e6**
	Analysis: Propensity matching was performed. To determine the propensity score, a logistic regression model was developed where the dependent variable was the type of procedure or gender. All baseline confounders, including gender, race, comorbidities age, year of surgery, hospital, and surgeon volume (as continuous variables) were included in the model as independent variables. The patients were matched by greedy algorithm using the 8- to 1-digit matching scheme without replacement. Perioperative complications and 30-day mortality were compared in matched groups using the McNemar test. Survival curves were constructed with Cox models. Propensity score and type of repair or gender were included in this regression analysis.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, perioperative adverse events, long term survival, long-term adverse events, and reintervention.
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? High risk – all 1995 to 2006 hospitalisations for OSR and hospitalisations from 2000 through 2006 for EVAR were included. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same databases of a nationwide health insurance provider. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA were included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – demographic variables including age and gender were controlled. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a good range of individual comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – no indication that AAA characteristics were adjusted for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – an administrative (healthcare insurance provider) database with high-level diagnosis and procedure codes was used in this study. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – an administrative (healthcare insurance provider) database with high-level diagnosis and procedure codes was used in this study. 3.3. Is method of data collection likely to record long-term outcomes accurately? High risk – an administrative (healthcare insurance provider) database with high-level diagnosis and procedure codes was used in this study. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – none reported.

Full citation	Egorova NN, Vouyouka AG, McKinsey JF, et al. (2011) Effect of gender on long-term survival after abdominal aortic aneurysm repair based on results from the Medicare national database. J Vasc Surg. 54(1):1-12.e6
	4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors did not demonstrate that missing data were considered in the analyses. 4.3. Have different methods been compared within the study? Moderate risk – different methods were compared but they relied on the same assumption about selection. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Low risk – a greedy algorithm that used an 8 to 1 matching scheme without replacement was used. 5.2. Was overlap / common support appropriately assessed? High risk – no assessment reported. 5.3. Has balancing of the covariates been demonstrated? Moderate risk – Conventional hypothesis tests, with no evidence of significant differences. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Eslami MH, Rybin DV, Doros G, Farber A. (2017) Description of a risk predictive model of 30-day postoperative mortality after elective abdominal aortic aneurysm repair. J Vasc Surg. 65:65-75
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2005 to 2011 Aim of the study: to describe such a 30-day postoperative risk prediction model using American College of Surgeons National Surgical Quality Improvement Project (NSQIP) data.
Participants	Sample size: EVAR group, n=4,635; OSR group, n= 14,282 Inclusion criteria: patients who underwent elective AAA repair Exclusion criteria: Patients who had a non-elective admission for AAA repair or were missing information for age, sex, discharge status (dead, discharged to home, or discharged to rehabilitation facility) or procedure type were excluded. We also excluded patients who had secondary procedures. Any missing data was treated as missing completely at random and not included in the analyses, and no data imputation was considered.

Full citation	**Eslami MH, Rybin DV, Doros G, Farber A. (2017) Description of a risk predictive model of 30-day postoperative mortality after elective abdominal aortic aneurysm repair. J Vasc Surg. 65:65-75**
	Baseline characteristics: Mean age: data not provided for the EVAR and OSR groups separately Gender: data not provided for the EVAR and OSR groups separately Mean aneurysm diameter: not reported
Methods	Data collection: The NSQIP database was queried. This quality improvement database that is risk-adjusted and includes information about both index admission and subsequent 30-day postoperative follow-up after the index operation. These risk adjustments are performed for the hospitals. Currently, over 500 hospitals in the United States participate in NSQIP, including university centres and private hospitals. These institutions select for analysis a random set of surgical cases, including those from general, vascular, plastic, and cardiac surgery. Cases are gathered on an 8-day cycle, trained staff abstract 135 pre-, intra-, and postoperative measures directly from the medical record and enter them as de-identified data in the database Analysis: multivariable logistic regression. In this parsimonious model, choice of operation significantly increased odds of mortality (OAR vs EVAR: odds ratio [OR], 2.71; 95% confidence interval [CI], 2.12-3.47; P<0.001). In this model, presence of chronic obstructive pulmonary disease (OR, 1.44; 95% CI, 1.10-1.90; P=0.008) and history of myocardial disease (OR, 1.36; 95% CI, 1.06-1.74; P=0.017) were significantly associated with increased mortality. Elevated creatinine also significantly increased the risk of mortality. End-stage renal disease (ESRD) was associated with POD in a univariate fashion, but this variable was not included in the final model. Age and sex also affected mortality; patients 70 years of age and older had significantly higher odds of death compared with those under 70 (OR, 2.24, 95% CI, 1.66-3.03; P<0.001) and women had increased odds of mortality compared with men (OR 1.46, 95% CI, 1.11-1.92; P=0.007). Functionally dependent patients had more than twice odds of mortality after elective AAA repair than functionally independent patients (OR, 2.29; 95% CI, 1.44-3.64; P<0.001). It was noted that weight loss, defined by ACS-NSQIP as ">10% loss of body weight in the last six months", was highly predictive of mortality after elective AAA repair.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk - ≥5-yr recruitment with no adjustment for year of operation 1.2. Were cohorts from the same place? Low risk – same country 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA were included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk

Full citation	Eslami MH, Rybin DV, Doros G, Farber A. (2017) Description of a risk predictive model of 30-day postoperative mortality after elective abdominal aortic aneurysm repair. J Vasc Surg. 65:65-75
	2.3. Does study control appropriately for AAA characteristics? High risk - none 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative database 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – administrative database 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk 4.3. Have different methods been compared within the study? High risk – no different methods compared **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk (20 covariates; 315 events) 6.2 Were interactions between treatment and other covariates considered? High risk **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Giles KA, Landon BE, Cotterill P, et al. (2011) Thirty-day mortality and late survival with reinterventions and readmissions after open and endovascular aortic aneurysm repair in Medicare beneficiaries. J Vasc Surg. 53(1):6-12
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2001 to 2004 Aim of the study: to evaluate the impact of reinterventions and readmission after initial AAA repair on 30-day and long-term mortality
Participants	Sample size of matched cohort: EVAR group, n=22,826; OSR group, n=22,826

Full citation	**Giles KA, Landon BE, Cotterill P, et al. (2011) Thirty-day mortality and late survival with reinterventions and readmissions after open and endovascular aortic aneurysm repair in Medicare beneficiaries. J Vasc Surg. 53(1):6-12**
	Inclusion criteria: All Medicare beneficiaries who underwent EVAR or OSR for unruptured AAA repair were included. Note: there is no indication that selection was limited to only people with infrarenal AAA. Exclusion criteria: not reported Baseline characteristics (in participants who did not received any reintervention): Mean age (SD): EVAR group, 76.2 (5.4 years; OSR group, 75.9 (5.2) years Gender: EVAR group, 84.7% male; OSR group, 85.6% male Mean aneurysm diameter: not reported Myocardial disease within 6 months: EVAR group, 1.9%; OSR group, 1.8% Congestive heart failure: EVAR group, 13.2%; OSR group, 13.1% Peripheral vascular disease: EVAR group, 20.9%; OSR group, 20.3% Cerebrovascular disease: EVAR group, 16.5%; OSR group, 16.2% Hypertension: EVAR group, 65.9%; OSR group, 65.7% Diabetes: EVAR group, 15.8%; OSR group, 15.8% COPD: EVAR group, 29.7%; OSR group, 29.1%
Methods	Data collection: data were retrospectively collected from the Medicare databases. Patient demographic characteristics were identified from the Medicare denominator file database. Comorbidities were ascertained using inpatient and outpatient claims up to 2 years before but not including the admission for repair. Reinterventions and readmissions were determined from inpatient and outpatient claims by assessing ICD9 coding, and mortality was determined from the Medicare denominator file. Analysis: To control for the non-random assignment of patients to OSR vs EVAR, investigators created matched cohorts of patients using a logistic regression model from demographics and pre-existing comorbidities. Investigators measured the rates of coexisting conditions using an Elixhauser algorithm that was adapted to also include diagnoses made in the outpatient setting. They matched each beneficiary who underwent EVAR to the beneficiary who underwent OSR with the closest estimated propensity score. To ensure close matches, investigators required that the estimated log-odds scores for EVAR of a patient who underwent EVAR and one who underwent OSR were within 0.60 SD of one another. They stated that this requirement ensured the removal of approximately 90% of the bias in estimates of effects due to differences in covariate distributions between the EVAR group and the OSR group.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality and reinterventions up to 7 years.
Study Appraisal using NICE's	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period.

Full citation	**Giles KA, Landon BE, Cotterill P, et al. (2011) Thirty-day mortality and late survival with reinterventions and readmissions after open and endovascular aortic aneurysm repair in Medicare beneficiaries. J Vasc Surg. 53(1):6-12**
bespoke risk of bias assessment tool	1.2. Were cohorts from the same place? Low risk – all participants were selected from the same databases of a nationwide health insurance provider. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA were included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – although authors did not explicitly state that age and gender were adjusted for they stated that they used all available demographic and clinical characteristics for beneficiaries at baseline as explanatory variables. These are likely to have included age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – authors stated that they used all available demographic and clinical characteristics for beneficiaries at baseline as explanatory variables. 2.3. Does study control appropriately for AAA characteristics? High risk – authors stated that they were not able to assess anatomic differences among patients. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables that could mediate the treatment effect appear to be controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – an administrative (healthcare insurance provider) database with high-level diagnosis and procedure codes was used in this study. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – an administrative (healthcare insurance provider) database was used. 3.3. Is method of data collection likely to record long-term outcomes accurately? High risk – an administrative (healthcare insurance provider) database was used. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors did not demonstrate that missing data were considered in the analyses. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Low risk – matching algorithm reported in previous study by the same authors and seems reasonable. 5.2. Was overlap / common support appropriately assessed? High risk – no indication that overlap was accurately assessed. 5.3. Has balancing of the covariates been demonstrated? High risk – no standardised differences reported.

Full citation	Giles KA, Landon BE, Cotterill P, et al. (2011) Thirty-day mortality and late survival with reinterventions and readmissions after open and endovascular aortic aneurysm repair in Medicare beneficiaries. J Vasc Surg. 53(1):6-12
	Analysis – simple multivariable models 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Jackson RS, Chang DC, and Freischlag JA. (2012) Comparison of long-term survival after open vs endovascular repair of intact abdominal aortic aneurysm among Medicare beneficiaries. JAMA. 307(15):1621-8**
Study details	Study design: retrospective cohort study Location(s): USA Study period: January 2003 to December 2007 Aim of the study: to compare long-term outcomes after EVAR and OSR of unruptured AAA
Participants	Sample size: EVAR group, n=3,826; OSR group, n=703 Inclusion criteria: people 65 and over with clinical diagnosis as well as procedure codes (ICD9 codes) corresponding to EVAR or OSR of unruptured AAA were included. Exclusion criteria: People younger than 65 years Baseline characteristics: Mean age (SD): EVAR group, 76.4 (6.3) years; OSR group, 75.2 (5.7) years Gender: EVAR group, 79.9% male; OSR group, 70.8% male Mean aneurysm diameter: not reported Diabetes: EVAR group, 18.8%; OSR group, 14.1% Hypertension: EVAR group, 59.1%; OSR group, 53.9% Chronic renal failure: EVAR group, 3.0%; OSR group, 4.1% Congestive heart failure: EVAR group, 12.3%; OSR group, 13.2% COPD: EVAR group, 24.3%; OSR group, 25.2%
Methods	Data collection: the study cohort was composed of eligible patients identified from the Medicare Standard Medicare Standard Analytic Files database using ICD9 codes. In the database, patient records contain longitudinal data from the index admission and subsequent hospital admissions, vital status, and date of death for deceased beneficiaries (based on linkage with social security administrative database). Comorbidities were assigned using Clinical Classifications Software (CCS): the software consolidates over 14,000 ICD9 codes and is developed by the Healthcare Cost and Utilization Project of the Agency for Healthcare Research and Quality. Analysis: Logistic regression was used to calculate a propensity score for each patient based on emergency presentation, age, calendar year of repair, gender, race, type of CCS diagnostic category, and number CCS diagnostic categories. Propensity scores were divided into quintiles by repair type, and propensity score quintile was treated as a categorical variable in the multivariable models. Multivariate logistic regression was used to assess 30-day mortality and hospital length of stay. Coz proportional hazards regression was used for all other outcomes. Two sensitivity analyses were performed to examine the effect of model assumptions on study results. The first sensitivity analysis examined the influence of possible misclassification of surgical complications as comorbidities in the portion of the cohort whose CCS categories were drawn from the same calendar years as the index AAA repair. The second sensitivity analysis examined the influence of emergency presentation on study results. In the primary analysis, emergency presentation was adjusted for as a component of the propensity score.
Intervention	EVAR

Full citation	**Jackson RS, Chang DC, and Freischlag JA. (2012) Comparison of long-term survival after open vs endovascular repair of intact abdominal aortic aneurysm among Medicare beneficiaries. JAMA. 307(15):1621-8**
Comparator	OSR
Outcomes	Mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – all cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – both cohorts were derived from the same national healthcare provider database. 1.3. Is the definition of AAA the same across cohorts? High risk –all infrarenal and complex AAA included, which is likely to have led to implicitly different inclusion criteria across treatment arms. I In addition, a higher proportion of participants in the OSR group (23.3%) were admitted under emergency conditions compared to the EVAR (14.7%), possibly due to symptomatic aneurysms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – the study controlled for demographic variables including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – the study controlled for a broad set of comorbidities. 2.3. Does study control appropriately for AAA characteristics? High risk – there is no indication that AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – participants were identified using administrative registries with high-level diagnosis and procedure codes. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – outcomes were assessed using administrative registries with high-level codes. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – surgical registries were used with linkage to a social security administrative database. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – none reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – casewise deletion was used to exclude patients with missing data. 4.3. Have different methods been compared within the study? Moderate risk - different methods were compared but all relied on the same assumption about selection **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? Moderate risk –box plots of propensity scores stratified by AAA repair type were examined, and median propensity score and interquartile range were compared between groups.

Full citation	**Jackson RS, Chang DC, and Freischlag JA. (2012) Comparison of long-term survival after open vs endovascular repair of intact abdominal aortic aneurysm among Medicare beneficiaries. JAMA. 307(15):1621-8**
	5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions appear to have been considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Jetty P, Hebert P, van Walraven C, et al. (2010) Long-term outcomes and resource utilization of endovascular versus open repair of abdominal aortic aneurysms in Ontario. J Vasc Surg. 51(3):577-83**
Study details	Study design: retrospective cohort study Location(s): Canada Study period: April 2002 to March 2007 Aim of the study: to compare population-based clinical outcomes and resource utilisation for EVAR and OSR of all elective AAA repairs in Ontario
Participants	Sample size: EVAR group, n=888; OSR group, n= 5,573 Inclusion criteria: people who underwent EVAR or OSR of unruptured infrarenal AAA in the province of Ontario were included. Exclusion criteria: people with ruptured aneurysms, thoracic aneurysms, isolated iliac aneurysms, and pseudoaneurysms were excluded. Non-Ontarians who underwent surgical repair in an Ontario hospital were excluded because their follow-up period is not captured by the Ontario health databases used for the study. Baseline characteristics: Median age (range): EVAR group, 76 (70-81) years; OSR group, 72 (66-77) years Gender: EVAR group, 86.2% male; OSR group, 80.3% male Mean aneurysm diameter: not reported Peripheral vascular disease: EVAR group, 49.1%; OSR group, 32.3% Cerebrovascular disease: EVAR group, 13.7%; OSR group, 9.9% Congestive heart failure: EVAR group, 42.3%; OSR group, 28.6% Pulmonary disease: EVAR group, 33.6%; OSR group, 21.7% Diabetes: EVAR group, 23.8%; OSR group, 18.0% Hypertension: EVAR group, 72.9%; OSR group, 60.9%

Full citation	**Jetty P, Hebert P, van Walraven C, et al. (2010) Long-term outcomes and resource utilization of endovascular versus open repair of abdominal aortic aneurysms in Ontario. J Vasc Surg. 51(3):577-83**
	Cancer: EVAR group, 17.7%; OSR group, 12.7%
Methods	Data collection: All study data were obtained from several population based administrative databases. The primary database was the Canadian Institute for Health Information-Discharge Abstract Database (CIHI-DAD) which contains ICD10 diagnosis codes and Canadian Classification of Health Interventions procedure codes. Other databases used for the study included a diabetes database, the Registered Persons Database (which captures the death dates), a myocardial infarction database, a health insurance plan database, and a prescriptions database. These databases were linked by a patient identifier number Analysis: Investigators performed regression based on propensity scores. Propensity scores were calculated using multivariable logistic regression with preoperative patient factors outlined in the Charlson Comorbidity Index as the predictor variables and the assignment to EVAR as the outcome variable. Clinically important terms were included in the model that generated the probability that a patient would receive either treatment. Patients were then classified into quintiles by their propensity score. This allowed investigators to evaluate the quality of the model by confirming balance. Finally, Cox proportional hazard modelling was used for the adjusted survival analysis using the propensity quintiles as stratifying variables.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, mortality at 5 years, length of stay, length of stay in ICU, discharge to long-term care
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants underwent AAA repair in hospitals from the Ontario region. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controlled for variables in the Charlson Comorbidity Index, including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a broad range of comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – study did not control for AAA characteristics. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registries with high-level diagnosis and procedure codes were used.

Full citation	Jetty P, Hebert P, van Walraven C, et al. (2010) Long-term outcomes and resource utilization of endovascular versus open repair of abdominal aortic aneurysms in Ontario. J Vasc Surg. 51(3):577-83
	3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – administrative registries with high-level diagnosis and procedure codes were used. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – data were collected via linkage to routine registries. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – none reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors do not discuss what approach was taken for missing data. 4.3. Have different methods been compared within the study? Moderate risk – different methods compared but all rely on the same assumption about selection **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? High risk – no checks reported 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were explored. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Johnson ML, Bush RL, Collins TC, et al. (2006) Propensity score analysis in observational studies: outcomes after abdominal aortic aneurysm repair. Am J Surg. 192(3):336-43.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: May 2001 and September 2003 Aim of the study: to compare outcomes of EVAR and OSR of unruptured AAA.
Participants	Sample size of unmatched cohort: EVAR group, n=717; OSR group, n=1,187 Sample size of matched cohort: EVAR group, n=670; OSR group, n=670 Inclusion criteria: all people wo underwent EVAR or OSR of unruptured AAA were included. Exclusion criteria: people with ruptured AAA, thoracic or thoracoabdominal aortic aneurysm, or those who underwent conversion from EVAR to OSR were excluded. Baseline characteristics: Mean age (SD): EVAR group, 71.6 (7.8) years; OSR group,70.2 (7.9) years Gender: EVAR group, 99.6% male; OSR group, 99.1% male Mean aneurysm diameter: not reported COPD: EVAR group, 26.6%; OSR group, 26.0% Chronic heart failure: EVAR group, 3.6%; OSR group, 2.4% Renal insufficiency: EVAR group, 1.1%; OSR group, 1.0% Cerebrovascular accident with neuro-deficit: EVAR group, 5.3%; OSR group, 6.0% Diabetes: EVAR group, 14.4%; OSR group, 13.2% Cancer: EVAR group, 1.1%; OSR group, 1.0%
Methods	Data collection: data were extracted from a detailed surgical registry run by the run by the military Veterans Health Administration: (A Veterans Affairs component of the National Surgical Quality Improvement Program; NSQIP). The NSQIP database requires hospitals to provide complete 30-day follow-up on at least 95% of patients. To supplement the information in the NSQIP records investigators used unique identifiers to link records with other Veterans Affairs databases: including the patient treatment file (which contains abstracts of all patients discharged), the outpatient clinic file (which contains records for every outpatient visit), and the VA beneficiary identification record locator system death file Analysis: Propensity score matching was used with no additional analyses performed. The propensity score (predicted probability of receiving EVAR) was obtained by performing a multivariable logistic regression of 32 independent variables. One-to-one matched samples were created and matched pairs were subsequently categorised into 5 groups or strata of increasing score. Patients with the lowest propensity score were most likely to receive OSR based on their baseline risk factors, whereas patients with higher propensity scores were more likely to receive EVAR. The clinical outcomes were subsequently compared between strata.

Full citation	**Johnson ML, Bush RL, Collins TC, et al. (2006) Propensity score analysis in observational studies: outcomes after abdominal aortic aneurysm repair. Am J Surg. 192(3):336-43.**
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, 1 year mortality and postoperative adverse events (included adverse cardiac events, renal dysfunction, pulmonary complications, wound complications, neurologic complications, bleeding requiring blood transfusion, and graft failure)
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – both cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – both cohorts were drawn from the same time period. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk - patient demographic data including age, gender and age were controlled for in the analyses. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – comorbidities, including those conditions known to have an influence on the risk of cardiovascular morbidity and mortality, were chosen for analysis. 2.3. Does study control appropriately for AAA characteristics? High risk – there is no indication that the study controlled for AAA characteristics. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – a detailed surgical registry was used to identify participants with diagnosis and procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low – a detailed surgical registry was used to collect data on outcomes. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate – a surgical registry was used with linkage to routine data registries. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors state that given the robust nature of the NSQIP and other databases used the likelihood of missing essential covariates is low. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching**

Full citation	Johnson ML, Bush RL, Collins TC, et al. (2006) Propensity score analysis in observational studies: outcomes after abdominal aortic aneurysm repair. Am J Surg. 192(3):336-43.
	5.1. Is the matching algorithm reported and reasonable? High risk – matching algorithm was discussed/reported. 5.2. Was overlap / common support appropriately assessed? High risk – insufficient checks were performed. 5.3. Has balancing of the covariates been demonstrated? Moderate risk – “baseline factors that were statistically significantly associated with type of surgery in unadjusted bivariate analyses were re-examined after propensity score stratification to confirm that baseline differences had been removed.” **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Karthikesalingam A, Holt PJ, Vidal-Diez A, et al. (2016) The impact of endovascular aneurysm repair on mortality for elective abdominal aortic aneurysm repair in England and the United States. J Vasc Surg. 64(2):321-327	
Study details	Study design: retrospective cohort study Location(s): UK and USA Study period: January 2005 to December 2010 Aim of the study: to compare in-hospital mortality between people who underwent EVAR and those who OSR for unruptured AAA repair, and subsequently compare outcomes between people in England and the United States.	
Participants	**UK cohort** Sample size: EVAR group, n=7,937; OSR group, n= 13,335 Inclusion criteria: all patients undergoing elective AAA repair. Exclusion criteria: ruptured AAA. Baseline characteristics: Median age (IQR): 74 (69-79) years Gender: 86.6% male Mean aneurysm diameter: not reported Diabetes: 11.3% Myocardial infarction: 5.2% Cerebrovascular disease: 2.0%	**USA cohort** Sample size: EVAR group, n=126,211; OSR group, n= 69,902 Inclusion criteria: all patients undergoing elective AAA repair. Exclusion criteria: ruptured AAA. Baseline characteristics: Median age (IQR): 72.7 (66.9-78.2) years Gender: 76.4% male Mean aneurysm diameter: not reported Diabetes: 17.8% Myocardial infarction: 14.7% Cerebrovascular disease: 5.7%

Full citation	**Karthikesalingam A, Holt PJ, Vidal-Diez A, et al. (2016) The impact of endovascular aneurysm repair on mortality for elective abdominal aortic aneurysm repair in England and the United States. J Vasc Surg. 64(2):321-327**	
	COPD: 15.8% Renal disease: 7.5% Cancer: 0.3%	COPD: 34.9% Renal disease: 10.0% Cancer: 0.4%
Methods	Data collection: Demographic and in-hospital outcome data were extracted from the UK Hospital Episode Statistics (HES) and the USA Nationwide Inpatient Sample (NIS) for all patients undergoing elective AAA repair. HES data captures the majority of people undergoing AAA repair in the UK whereas, NIS data is an anonymised, stratified sample of 20% of all discharges from U.S. hospitals. Participants were identified by examining diagnosis and procedure codes; such as ICD9 and OPCS4 codes. Analysis: logistic regression was performed adjusting for age, gender, social deprivation, comorbidity index scores hospital procedural volume (caseload), hospital bed capacity and teaching status. Backward selection procedures were used with comparison of models by the likelihood ratio test to ascertain whether individual covariates improved goodness-of-fit for prediction of in-hospital mortality. Note: Age- and gender-matched analyses were constructed to compare English and U.S. outcomes for in-hospital; however this was not considered relevant for this review.	
Intervention	EVAR	
Comparator	OSR	
Outcomes	In-hospital mortality	
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – recruitment over ≥5 yrs, but year of operation controlled for in analysis. 1.2. Were cohorts from the same place? Low risk – EVAR and OSR cohorts were compared in the context of the country in which they were performed. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – demographic variables including age and gender were controlled for. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – investigators adjusted for comorbidities by considering Charlson scores (a comorbidity index). 2.3. Does study control appropriately for AAA characteristics? High risk – the study did not control for AAA characteristics. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registries with high-level diagnosis and procedure codes were used.	

Full citation	Karthikesalingam A, Holt PJ, Vidal-Diez A, et al. (2016) The impact of endovascular aneurysm repair on mortality for elective abdominal aortic aneurysm repair in England and the United States. J Vasc Surg. 64(2):321-327
	3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – administrative registries with high-level diagnosis and procedure codes were used. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long-term outcomes assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no demonstration that missing data was considered. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk - number of events is ≥10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Lo RC, Bensley RP, Hamdan AD, et al. (2013) Gender differences in abdominal aortic aneurysm presentation, repair, and mortality in the Vascular Study Group of New England. J Vasc Surg. 57:1261-8**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2003 to 2011 Aim of the study: to describe differences in the presentation, choice of repair, and mortality among men and women undergoing AAA repair
Participants	Sample size: EVAR group, n=2,159; OSR group, n=1,867 Inclusion criteria: Patient presentation was categorized as intact (including patients who were symptomatic as well as those undergoing elective repair) or ruptured. Four subgroup analyses were performed: intact EVAR, ruptured EVAR, intact open repair, and ruptured open repair. For the purposes of multivariable modelling, intact EVAR served as the referent group. Exclusion criteria: None mentioned Baseline characteristics: Mean age for men (range): EVAR group, 74 (67-80) years; OSR group, 71 (64-77) years. Mean age for women (range): EVAR group, 77 (71-81) years; OSR group, 73 (68-78) years Gender: EVAR group, 80.0% male; OSR group, 74.6% male Mean aneurysm diameter: not reported Diabetes mellitus: Males: EVAR group, 1.16%; OSR group, 1.08%. Females: EVAR group, 4.42%; OSR group, 2.74% Coronary artery disease: Males: EVAR group, 2.14%; OSR group, 2.51%. Females: EVAR group, 6.28%; OSR group, 5.7% Congestive heart failure: Males: EVAR group, 0.64%; OSR group, 0.53%. Females: EVAR group, 2.79%; OSR group, 1.43% COPD: Males: EVAR group, 1.97%; OSR group, 2.37%. Females: EVAR group, 9.53%; OSR group, 9.49%
Methods	Data collection: retrospective review of open and endovascular AAA repairs in the Vascular Study Group of New England (VSGNE) database, a voluntary collaboration among vascular surgeons, cardiologists, and radiologists from 30 academic and community hospitals in the six New England states. Formed as a quality improvement initiative, VSGNE represents a pool of clinical data related to several frequently performed vascular procedures that have been collected since 2003. Analysis: Multivariable logistic regression was used to determine predictors of 30-day mortality. Individual survival curves for each presentation and treatment group were evaluated for differences in survival between men and women. Cox proportional hazards modelling was used to determine predictors of long-term mortality. Multivariable predictors of 30-day mortality were age, coronary artery disease, preoperative dialysis dependence, and presentation/ treatment subgroup (intact vs ruptured and EVAR vs open repair). Age >80 years increased the odds of 30-day mortality by 10.9 and rupture increased the odds by 48.4 for EVAR and 83.8 for open repair. Female gender did not reach statistical significance as a predictor of 30-day mortality.
Intervention	EVAR
Comparator	OSR

Full citation	Lo RC, Bensley RP, Hamdan AD, et al. (2013) Gender differences in abdominal aortic aneurysm presentation, repair, and mortality in the Vascular Study Group of New England. J Vasc Surg. 57:1261-8
Outcomes	30-day mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk - ≥5-yr recruitment with no adjustment for year of operation 1.2. Were cohorts from the same place? Low risk 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included. This is likely to have led to implicitly different inclusion criteria across treatment arms.. **Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk 2.3. Does study control appropriately for AAA characteristics? High risk - none 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – detailed surgical registry 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk - none reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk 4.3. Have different methods been compared within the study? High risk – none reported **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk 6.2 Were interactions between treatment and other covariates considered? High risk **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Locham S, Rizwan M, Dakour-Aridi H et al. (2018) Outcomes after elective abdominal aortic aneurysm repair in obese versus nonobese patients. J Vasc Surg. S0741-5214(18)30907-8
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2003 to 2017 Aim of the study: to use a nationally representative vascular database to compare in-hospital outcomes in obese versus non-obese patients undergoing elective EVAR or OSR.
Participants	Sample size: EVAR group, n=26,723; OSR group, n=6,359 Inclusion criteria: people who underwent elective OSR and EVAR were included. Note: people with complex aneurysm anatomies are likely to have been included. Exclusion criteria: people who underwent urgent or emergent AAA repair were excluded from the study. Baseline characteristics: Mean age: treatment specific ages not reported Gender: EVAR group, 81.4% male; OSR group, 73.8% male Mean aneurysm diameter: treatment specific ages not reported Diabetes: EVAR group, 20.5%; OSR group, 16.3% Hypertension: EVAR group, 83.3%; OSR group, 84.2% Coronary artery disease: EVAR group, 29.8%; OSR group, 27.0% Coronary heart failure: EVAR group, 11.5%; OSR group, 7.5% COPD: EVAR group, 32.5%; OSR group, 32.7% Dialysis: EVAR group, 1.0%; OSR group, 0.6% Prior bypass: EVAR group, 3.0%; OSR group, 4.4%
Methods	Data collection: investigators queried a large national database that contained detailed preoperative, operative, and postoperative characteristics of patients on several common vascular surgical procedures (the Vascular Quality Initiative: VQI) to identify people who underwent elective AAA repair and obtain information on their outcomes. Analysis: Multivariate logistic regression was performed. Covariates were chosen on the basis of clinical and statistical significance in univariate analysis (significance level not specified). The interaction between surgery and obesity was also evaluated. All models were tested using variation inflation factor, Hosmer-Lemeshow goodness of fit test, and area under the receiver operating characteristic curve (C-statistic).
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality and adverse events

Full citation	Locham S, Rizwan M, Dakour-Aridi H et al. (2018) Outcomes after elective abdominal aortic aneurysm repair in obese versus nonobese patients. J Vasc Surg. S0741-5214(18)30907-8
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national vascular registry. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included. This is likely to have led to implicitly different inclusion criteria across treatment arms.. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study adjusted for demographics variables, including age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a broad range of demographic variables were controlled for. 2.3. Does study control appropriately for AAA characteristics? Moderate risk – the study controlled for aneurysm diameter. 2.4. Could any adjustment variables have been affected by the intervention? High risk – investigators controlled for blood loss, which could mediate the treatment effect. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – a detailed vascular registry was used to identify participants. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were obtained from a detailed vascular registry. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long-term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – model specification/fit was checked using the Hosmer–Lemeshow test as well as the C-statistic 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors reported that the effect of missing data was minimal as the VQI performs thorough checks to maintain standards. All study hospitals are required to enter complete information if they are found to have a large number of missing data. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered.

Full citation	Locham S, Rizwan M, Dakour-Aridi H et al. (2018) Outcomes after elective abdominal aortic aneurysm repair in obese versus nonobese patients. J Vasc Surg. S0741-5214(18)30907-8
	6.2 Were interactions between treatment and other covariates considered? Low risk – the following interaction was considered OSR*obese **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Mark TL, Lawrence W, Coffey RM, et al. (2013) The value of linking hospital discharge and mortality data for comparative effectiveness research. Journal of comparative effectiveness research 2(2), 175-84
Study details	Study design: retrospective cohort study Location(s): USA Study period: July 2000 to January 2006 Aim of the study: to demonstrate the value of linking state community hospital discharge data to vital statistics death files for research by conducting a comparative effectiveness analysis of elective EVAR versus OSR.
Participants	Sample size of unmatched cohort: EVAR group, n=6,046; OSR group, n=7,606 Sample size of matched cohort: EVAR group, n=4,483; OSR group, n=4,483 Inclusion criteria: people who underwent elective EVAR or OSR for unruptured AAA were included. Note: there is no indication that selection was limited to only infrarenal AAAs. Exclusion criteria: people who underwent emergency AAA repair, both EVAR and OSR during the same admission, revision of a previous AAA repair, and those with less than 1 year of follow-up data available were excluded. Furthermore, people with the clinical codes indicating the following conditions were excluded: thoracoabdominal aneurysms, and/or visceral or renal bypass, polyarteritis nodosa, coarctation of the aorta, Marfan syndrome and other congenital anomalies. Finally, individuals who had multiple death records, a death date prior to an admission date, or multiple codes for sex were excluded. Baseline characteristics (of matched cohort): % <65: EVAR group, 22%; OSR group, 24% Gender: EVAR group, 8% male; OSR group, 6% male Mean aneurysm diameter: not reported COPD: EVAR group, 29%; OSR group,31% Diabetes with or without complications: EVAR group, 15%; OSR group, 14% Renal failure: EVAR group, 5%; OSR group, 5% Lymphoma: EVAR group, 1%; OSR group, 1% Metastatic cancer: EVAR group, 1%; OSR group, 0%

Full citation	Mark TL, Lawrence W, Coffey RM, et al. (2013) The value of linking hospital discharge and mortality data for comparative effectiveness research. Journal of comparative effectiveness research 2(2), 175-84
	Solid tumour without metastasis: EVAR group, 7%; OSR group, 7% Ischaemic heart disease: EVAR group, 43%; OSR group, 43% Cerebral vascular occlusive disease: EVAR group, 5%; OSR group, 5%
Methods	Data collection: investigators identified participants and obtained data on their outcomes by linking (using social security numbers) 2 administrative databases managed by the California Office of State-wide Health Planning and Development. Analysis: Propensity score matching was performed. Logistic regression was performed to derive propensity scores. The predicted propensity score was then used to create a 1:1 match between cohorts using greedy matching with a caliper of 0.25 standard deviation of the propensity score. Balance on the covariates was assessed by computing standardized differences for each covariate; ‘balanced’ was defined as <10 standardised differences. Supplementary, analyses using simple multivariable models were also performed.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital, 30-day, 1-year and 5-year mortality
Study Appraisal using NICE’s bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – recruitment over ≥5 yrs, but year of operation controlled for in analysis 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same regional database. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included. This is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? High risk – no details were provided. 2.2. Does study control appropriately for comorbidity and/or fitness? High risk – no details were provided. 2.3. Does study control appropriately for AAA characteristics? High risk – authors stated that information about aneurysm size or anatomical features was not available. 2.4. Could any adjustment variables have been affected by the intervention? High risk – no details were provided. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – participants were identified using an administrative database with high-level procedure codes. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – outcomes were assessed using an administrative database with high-level procedure codes. 3.3. Is method of data collection likely to record long-term outcomes accurately? Moderate risk – an administrative patient discharge database was used with linkage to a reliable routine registry.

Full citation	**Mark TL, Lawrence W, Coffey RM, et al. (2013) The value of linking hospital discharge and mortality data for comparative effectiveness research. Journal of comparative effectiveness research 2(2), 175-84**
	Analysis – general 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors stated that approximately 20-30% of patient discharge records did not have a valid social security number and therefore were linked between databases. 4.3. Have different methods been compared within the study? High risk – no different methods were compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Low risk – matching was performed using greedy matching with a caliper of 0.25 standard deviation of the propensity score. 5.2. Was overlap / common support appropriately assessed? High risk – no assessment reported. 5.3. Has balancing of the covariates been demonstrated? Low risk – conventional hypothesis tests were performed. Furthermore, standardised differences in covariates were also reported and considered by investigators. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Quintana MJ, Gich I, Librero J, Bellmunt-Montoya S, Escudero JR, Bonfill X. Variation in the choice of elective surgical procedure for abdominal aortic aneurysm in spain. Vascular Health and Risk Management. 2019;15:69.**
Study details	Study design: retrospective cohort study Location(s): Spain Study period: 2002 to 2012 Aim of the study: to identify current preferences among Spanish hospitals for OSR or EVAR and to determine changes in these preferences over the course of the study period
Participants	Sample size: EVAR group, n=4,010; OSR group, n=4,010 Inclusion criteria: primary diagnosis of unruptured AAA, elective admission, treated with OSR or EVAR Exclusion criteria: thoracic or thoracoabdominal AA, aortic dissection, patients who underwent both EVAR and OSR during a single intervention, ruptured AAA and emergency admission Baseline characteristics (whole cohort):

Full citation	**Quintana MJ, Gich I, Librero J, Bellmunt-Montoya S, Escudero JR, Bonfill X. Variation in the choice of elective surgical procedure for abdominal aortic aneurysm in spain. Vascular Health and Risk Management. 2019;15:69.**
	Mean age 71.4 Gender: 96.7% male Mean aneurysm diameter: not reported Mean Charlson index: 0.7
Methods	Data collection: data in the database “Minimum Basic Dataset at Hospital Discharge” (in Spanish: “Conjunto Mínimo Básico de Datos al Alta Hospitalaria” [CMBDAH]). This database contains data on patients discharged from Spanish public hospitals. ICD9 codes used reported. Analysis:. Generalised linear mixed models (multivariate logistic regressions with hospital random effect) to examine interhospital variation. Three multivariate models that included temporal and hospital effects were developed. The first model included adjusted variables; the second model also considered a potential time trend (year of surgery); and the third model added surgical volume (ie, number of procedures) performed at each hospital. The results were adjusted for individual factors (age, gender, and comorbidities) to assess the influence of surgical volumes.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality
Study Appraisal using NICE’s bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical registry 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included, which is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – cohort controls for age and sex. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – Charlson index. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables controlled for that may mediate treatment effect. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registries with high-level diagnosis and procedure codes. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – administrative registries with high-level diagnosis and procedure codes.

Full citation	Quintana MJ, Gich I, Librero J, Bellmunt-Montoya S, Escudero JR, Bonfill X. Variation in the choice of elective surgical procedure for abdominal aortic aneurysm in spain. Vascular Health and Risk Management. 2019;15:69.
	3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long-term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – Hosmer–Lemeshow test 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no details. 4.3. Have different methods been compared within the study? High risk – no different methods were compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – 601 events, 6 covariates. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Raval MV, and Eskandari MK. (2012) Outcomes of elective abdominal aortic aneurysm repair among the elderly: endovascular versus open repair. Surgery. 151(2):245-60.
Study details	Study design: retrospective cohort study Location(s): USA Study period: January 2005 to December 2008 Aim of the study: to analyse outcomes of endovascular EVAR and OSR of elective AAA
Participants	Sample size: EVAR group, n=2,350; OSR group, n=5,586 Inclusion criteria: people who underwent EVAR or OSR for unruptured AAA were included. Note: people with complex aneurysms may also have been included. Exclusion criteria: people with ruptured AAA, wound infection, pneumonia, ventilator dependence/reintubation, renal failure, stroke, and coma prior to AAA repair were excluded. Baseline characteristics (whole cohort): Mean age not reported

Full citation	Raval MV, and Eskandari MK. (2012) Outcomes of elective abdominal aortic aneurysm repair among the elderly: endovascular versus open repair. Surgery. 151(2):245-60.
	Gender: 81.1% male Mean aneurysm diameter: not reported Diabetes: 13.9% COPD: 18.3% Congestive heart failure: 1.0% Hypertension requiring treatment: 80.2% Neurologic disease/event: 15.5% Cancer: 0.8%
Methods	Data collection: investigators obtained data by querying the American College of Surgeons version of the National Surgical Quality Improvement Program (ACS-NSQIP) database using procedure codes. Mortality was determined by examination of medical records, contacting patients, and querying social Security Death Index and the National Obituary Archives. Analysis: multivariate logistic regression was performed. Predictor variables were entered into the regression models by forward stepwise selection except operative approach (which was "forced" into the models). Final models were assessed for ''goodness-of-fit'' using Hosmer and Lemeshow tests and model discrimination was evaluated using C-index.
Intervention	EVAR
Comparator	OSR
Outcomes	People of any age (whole cohort): 30-day mortality and adverse events within 30 days (wound, pulmonary, cardiac, renal and infectious complications) People > 80 years: 30-day mortality, length of stay > 7 days and adverse events within 30 days (wound, pulmonary, cardiac, renal and infectious complications)
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.4. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.5. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical registry 1.6. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included, which is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – cohort controls for age and sex. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a broad range of relevant comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? High risk –analysis controlled for duration of operation.

Full citation	Raval MV, and Eskandari MK. (2012) Outcomes of elective abdominal aortic aneurysm repair among the elderly: endovascular versus open repair. Surgery. 151(2):245-60.
	Data collection 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – participants were identified using detailed surgical registry with diagnosis and procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – detailed surgical registries were used to assess outcomes. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long-term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – model specification/fit was assessed using the C-statistic and the Hosmer–Lemeshow test 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk - variables with considerable amounts of missing data were accounted for in the analyses. 4.3. Have different methods been compared within the study? High risk – no different methods were compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Salata K, Hussain MA, de Mestral C, Greco E, Aljabri BA, Mamdani M, Forbes TL, Bhatt DL, Verma S, Al-Omran M. Comparison of Outcomes in Elective Endovascular Aortic Repair vs Open Surgical Repair of Abdominal Aortic Aneurysms. JAMA network open. 2019 Jul 3;2(7):e196578.
Study details	Study design: retrospective propensity-matched cohort study Location(s): Canada Study period: April 2003 to March 2016

Full citation	**Salata K, Hussain MA, de Mestral C, Greco E, Aljabri BA, Mamdani M, Forbes TL, Bhatt DL, Verma S, Al-Omran M. Comparison of Outcomes in Elective Endovascular Aortic Repair vs Open Surgical Repair of Abdominal Aortic Aneurysms. JAMA network open. 2019 Jul 3;2(7):e196578.**
	Aim of the study: to assess the differences between EVAR and OSR for elective AAA repair in long-term survival, major adverse cardiovascular event (MACE)–free survival, reintervention, and secondary rupture
Participants	Sample size: EVAR group, n=4,010; OSR group, n= 4,010 Inclusion criteria: all elective EVARs and OSRs of AAA performed in Ontario, Canada, in patients 40 years and older Exclusion criteria: Patients with multiple AAA repair procedures listed on their index admission Baseline characteristics Mean age: EVAR group, 73.1 years; OSR group, 72.8 years Gender: EVAR group, 82.5% male; OSR group, 81.6% male Mean aneurysm diameter: not reported Coronary artery disease: EVAR group, 16.8%; OSR group, 16.7% MI: EVAR group, 5.6%; OSR group, 5.5% Congestive heart failure: EVAR group, 13.1%; OSR group, 12.4% Hypertension: EVAR group, 80.9%; OSR group, 80.5% Diabetes: EVAR group, 29.2%; OSR group, 28.6% COPD: EVAR group, 40.2%; OSR group, 39.3% Chronic kidney disease: EVAR group, 2.2%; OSR group, 1.8%
Methods	Data collection: Linked data across 11 administrative health databases Analysis: The propensity score for repair approach was calculated using a logistic regression model incorporating all covariates as potential confounders. Patients who received EVAR or OSR were matched 1:1 using the greedy nearest-neighbour method with a calliper width of 0.2 SD units. Balance of covariates was assessed using standardized differences, with differences less than 0.1 indicating good balance. Residual confounding was assessed using the distribution of tracer variables not used to specify the propensity score.
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, 30-day cardiovascular events, long-term survival, reinterventions
Study Appraisal using NICE's bespoke risk of bias	**Selection** 1.1. Were cohorts from the same time period? Low risk – the recruitment was >5 years. However, the year of recruitment was controlled for 1.2. Were cohorts from the same place? Low risk – same country 1.3. Is the definition of AAA the same across cohorts? High risk – authors note that the algorithm used to identify cases is validated to find infrarenal EVAR and infrarenal, pararenal, and juxtarenal OSR.

Full citation	**Salata K, Hussain MA, de Mestral C, Greco E, Aljabri BA, Mamdani M, Forbes TL, Bhatt DL, Verma S, Al-Omran M. Comparison of Outcomes in Elective Endovascular Aortic Repair vs Open Surgical Repair of Abdominal Aortic Aneurysms. JAMA network open. 2019 Jul 3;2(7):e196578.**
assessment tool	**Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – broad range of individual comorbidities and Charlson score 2.3. Does study control appropriately for AAA characteristics? High risk – none 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registry with high-level diagnosis and procedure codes 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk –administrative registry with high-level diagnosis and procedure codes 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – direct use of reliable routine data registry **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – none reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk 4.3. Have different methods been compared within the study? High risk **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Low risk 5.2. Was overlap / common support appropriately assessed? High risk – no checks reported 5.3. Has balancing of the covariates been demonstrated? Low risk – standardised differences (all $\leq$ 0.05) **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** Moderate risk **Directness:** directly applicable

Full citation	**Schermerhorn ML, Buck DB, O'Malley. et al. (2015) Long-term outcomes of abdominal aortic aneurysm in the Medicare population. NEJM. 373:328-338** **Schermerhorn ML, Giles KA, Sachs T, et al. (2011) Defining perioperative mortality after open and endovascular Aortic Aneurysm Repair in the US Medicare Population. J Am Coll Surg. 212(3):349-355** **Schermerhorn ML, O'Malley AJ, Jhaveri A, et al. (2008) Endovascular vs. open repair of abdominal aortic aneurysms in the Medicare population. NEJM. 359:464-474**
Study details	Study design: retrospective propensity-matched cohort study Location(s): USA Study period: January 2001 to December 2008 Aim of the study: to compare endovascular repair with open repair with respect to the long-term outcomes of each procedure
Participants	Sample size: EVAR group, n=39,966; OSR group, n= 39,966 Inclusion criteria: patients were included in the study if they had been continuously enrolled in traditional Medicare Parts A and B for at least 2 years before the repair, had received a discharge diagnosis of abdominal aortic aneurysm, and had undergone open repair or endovascular repair. Exclusion criteria: ruptured abdominal aortic aneurysms, thoracic aneurysms, thoracoabdominal aortic aneurysms, or aortic dissections. In addition, they excluded those who had undergone visceral bypass or renal bypass Baseline characteristics Mean age: EVAR group, 75.7 years; OSR group, 75.5 years Gender: EVAR group, 77.7% male; OSR group, 77.6% male Mean aneurysm diameter: not reported MI in previous 6 months: EVAR group, 1.6%; OSR group, 1.6% MI in previous 7-24 months: EVAR group, 6.8%; OSR group, 6.7% Valvular heart disease: EVAR group, 8.7%; OSR group, 8.6% Congestive heart failure: EVAR group, 11.7%; OSR group, 11.6% Peripheral vascular disease: EVAR group, 19.8%; OSR group, 19.4% Neurovascular disease: EVAR group, 13.9%; OSR group, 13.9% Hypertension: EVAR group, 63.2%; OSR group, 62.9% Diabetes: EVAR group, 16.1%; OSR group, 15.9% COPD: EVAR group, 27.8%; OSR group, 27.8% Renal failure: EVAR group, 5.6%; OSR group, 5.5% End-stage renal disease: EVAR group, 0.4%; OSR group, 0.4% Obesity: EVAR group, 2.0%; OSR group, 2.0%

Full citation	**Schermerhorn ML, Buck DB, O'Malley. et al. (2015) Long-term outcomes of abdominal aortic aneurysm in the Medicare population. NEJM. 373:328-338** **Schermerhorn ML, Giles KA, Sachs T, et al. (2011) Defining perioperative mortality after open and endovascular Aortic Aneurysm Repair in the US Medicare Population. J Am Coll Surg. 212(3):349-355** **Schermerhorn ML, O'Malley AJ, Jhaveri A, et al. (2008) Endovascular vs. open repair of abdominal aortic aneurysms in the Medicare population. NEJM. 359:464-474**
Methods	Data collection: Medicare medical records Analysis: Propensity score matching
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, in-hospital mortality, long-term survival/mortality, length of stay, discharge to home, adverse events (including readmission)
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.4. Were cohorts from the same time period? Low risk – the recruitment was >5 years. However, the year of recruitment was controlled for 1.5. Were cohorts from the same place? Low risk – same country 1.6. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included, which is likely to have led to implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk 2.3. Does study control appropriately for AAA characteristics? High risk – none 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registry with high-level diagnosis and procedure codes 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk –administrative registry with high-level diagnosis and procedure codes 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – direct use of reliable routine data registry **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – none reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk 4.3. Have different methods been compared within the study? Low risk **Analysis – matching**

Full citation	**Schermerhorn ML, Buck DB, O'Malley. et al. (2015) Long-term outcomes of abdominal aortic aneurysm in the Medicare population. NEJM. 373:328-338** **Schermerhorn ML, Giles KA, Sachs T, et al. (2011) Defining perioperative mortality after open and endovascular** **Aortic Aneurysm Repair in the US Medicare Population. J Am Coll Surg. 212(3):349-355** **Schermerhorn ML, O'Malley AJ, Jhaveri A, et al. (2008) Endovascular vs. open repair of abdominal aortic aneurysms in the Medicare population. NEJM. 359:464-474**
	5.1. Is the matching algorithm reported and reasonable? Low risk 5.2. Was overlap / common support appropriately assessed? High risk – no checks reported 5.3. Has balancing of the covariates been demonstrated? Moderate risk – conventional hypothesis tests, with no evidence of significant differences (except for age, where difference in mean appears minor – 75.7 -v- 75.5) **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Schwarze ML, Shen Y, Hemmerich J, et al. (2009) Age-related trends in utilization and outcome of open and endovascular repair for abdominal aortic aneurysm in the United States, 2001-2006. J Vasc Surg. 50(4):722-729.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2001 to 2006 Aim of the study: to compare utilisation and age-specific outcomes between EVAR and OSR for the treatment of AAA
Participants	Sample size: EVAR group, n=90,925; OSR group, n=75,222 Inclusion criteria: people who underwent EVAR or OSR of unruptured AAA were included. Note: there is no indication that selection was limited to only people with infrarenal AAA. Exclusion criteria: people <50 years, those with ruptured AAA, aortic dissection, thoracic or thoracoabdominal aortic aneurysms, coarctation of the aorta, Marfan syndrome and other congenital anomalies, gonadal dysgenesis-Turner syndrome, or polyarteritis nodosa were excluded. Baseline characteristics: Mean age: not reported Gender: not reported Mean aneurysm diameter: not reported

Full citation	**Schwarze ML, Shen Y, Hemmerich J, et al. (2009) Age-related trends in utilization and outcome of open and endovascular repair for abdominal aortic aneurysm in the United States, 2001-2006. J Vasc Surg. 50(4):722-729.**
	Comorbidities: not reported
Methods	Data collection: participants were identified using diagnosis and procedure codes (such as ICD9 and OPCS4 codes) to query an administrative database that used high-level diagnosis and procedure codes: the Nationwide Inpatient Sample (NIS). NIS data is an anonymised, stratified sample of 20% of hospitals including specialty, community and public hospitals, and academic medical centres. All data on comorbidities and outcomes were extracted from the NIS database. Analysis: multivariate linear regression was used to examine the risk-adjusted association between the type of procedure and length of stay. Multivariate logistic regression was performed to assess the risk-adjusted effect of the procedure performed on in-hospital mortality, discharge to home, and the occurrence of adverse events.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality, discharge to home, adverse events, and length of stay
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national administrative database 1.3. Is the definition of AAA the same across cohorts? High risk – authors highlight that 'only anatomically suitable infrarenal AAAs were treated by EVAR. **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – only gender was controlled for in the analyses. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate – authors stated that the models adjusted for comorbidities but no further details were provided. 2.3. Does study control appropriately for AAA characteristics? High risk – no AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registries with high-level diagnosis and procedure codes were used. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – administrative registries with high-level diagnosis and procedure codes were used. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long-term outcomes assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were performed.

Full citation	**Schwarze ML, Shen Y, Hemmerich J, et al. (2009) Age-related trends in utilization and outcome of open and endovascular repair for abdominal aortic aneurysm in the United States, 2001-2006. J Vasc Surg. 50(4):722-729.**
	4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – the impact of missing data was not factored into the analyses. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – authors do not explicitly state how many covariates (namely comorbidities) were considered in the regression models. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Symonides B, Śliwczyński A, Gałązka Z, et al. (2018) Short- and long-term survival after open versus endovascular repair of abdominal aortic aneurysm-Polish population analysis. PLoS One. 2018 Jun 14;13(6):e0198966.**
Study details	Study design: retrospective cohort study Location(s): Poland Study period: January 2011 to March 2016 Aim of the study: to compare short and long-term mortality and readmissions in patients with unruptured AAA treated by EVAR or OSR
Participants	Sample size of unmatched cohort: EVAR group, n=5,469; OSR group, n=2,336 Sample size of unmatched cohort: EVAR group, n=2,336; OSR group, n=2,336 Inclusion criteria: people with unruptured AAA who underwent elective EVAR or OSR were included. Note: there is no indication that selection was limited to only people with infrarenal AAA. Exclusion criteria: ruptured and people with thoracoabdominal aneurysms were excluded Baseline characteristics of the matched cohorts: Mean age (SD): EVAR group, 68.7 (8.0) years; OSR group, 68.5 (7.7) years Gender: EVAR group, 85.3% male; OSR group, 84.8% male Mean aneurysm diameter: not reported

Full citation	**Symonides B, Śliwczyński A, Gałązka Z, et al. (2018) Short- and long-term survival after open versus endovascular repair of abdominal aortic aneurysm-Polish population analysis. PLoS One. 2018 Jun 14;13(6):e0198966.**
	Hypertension: EVAR group, 48.7%; OSR group, 48.8% Chronic renal failure: EVAR group, 2.7%; OSR group, 2.7% Diabetes: EVAR group, 10.3%; OSR group, 11.3% Coronary heart disease: EVAR group, 9.1%; OSR group, 9.5% Stroke: EVAR group, 1.6%; OSR group, 1.9%
Methods	Data collection: patients were identified and outcome data were collected using ICD9 codes to query reimbursement data in a database managed by the only public and obligatory health insurer in Poland (the National Health Fund). Authors highlighted that the database tracks all patient admissions, main diagnoses, concomitant diseases and medical procedures longitudinally throughout the entire country. Additionally, the database contains information on birth and death dates. Analysis: propensity matching was performed, and supplemented with Cox proportional hazard regression. Propensity score analysis was performed by matching patients while controlling for age, gender and concomitant diseases. Similarly, Cox proportional-hazards analyses were performed controlling for age, gender, concomitant diseases and readmissions (expressed as total number of events).
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality, readmissions, and long-term survival
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – cohorts were selected from the same national healthcare insurer database. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included, which is likely to have led to implicitly different inclusion criteria across treatment arms.. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – the study controlled for age and sex. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – authors stated that comorbidities were controlled for but did not specify which ones. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – in the propensity score matching analysis, no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – participants were identified using a health insurance provider's administrative database.

Full citation	**Symonides B, Śliwczyński A, Gałązka Z, et al. (2018) Short- and long-term survival after open versus endovascular repair of abdominal aortic aneurysm-Polish population analysis. PLoS One. 2018 Jun 14;13(6):e0198966.**
	3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – outcomes were assessed using a health insurance provider's administrative database. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – direct use of reliable routine data registry **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk- no checks were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no details were provided. 4.3. Have different methods been compared within the study? Moderate risk - Different methods were compared but all relied on the same assumption about selection. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? High risk – no algorithm was reported. 5.2. Was overlap / common support appropriately assessed? High risk – no details were provided by authors. 5.3. Has balancing of the covariates been demonstrated? Moderate risk – conventional hypothesis tests were performed, with no evidence of significant differences. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – it was not possible to ascertain whether the sample size was adequate relative to number of covariates considered because authors did not provide a list of all covariates considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness: Directly applicable**

Full citation	**Tarbunou YA, Smith JB, Kruse RL, Vogel TR. Outcomes associated with hyperglycemia after abdominal aortic aneurysm repair. Journal of vascular surgery. 2019 Mar 1;69(3):763-73.**
Study details	Study design: retrospective propensity-matched cohort study Location(s): USA Study period: September 2008 to March 2014 Aim of the study: to evaluate outcomes and complications associated with AAA repair in patients with postoperative hyperglycemia.
Participants	Sample size: EVAR group, n=1,486; OSR group, n= 992 Inclusion criteria: Patients who underwent endovascular or open repair for a nonruptured AAA

Full citation	**Tarbunou YA, Smith JB, Kruse RL, Vogel TR. Outcomes associated with hyperglycemia after abdominal aortic aneurysm repair. Journal of vascular surgery. 2019 Mar 1;69(3):763-73.**
	Exclusion criteria: Patients who were younger than 21 years, had an emergent or urgent admission, or had no postoperative medication or laboratory data in Health Facts and patients whose postoperative blood glucose levels were below 80 mg/dL (hypoglycemic) Baseline characteristics (whole group) Mean age: 69.4 Gender: 71.7% male Mean aneurysm diameter: not reported Coronary artery disease: 8.1% Diabetes: 22.1% Chronic kidney disease: 11.2%
Methods	Data collection: Patients were identified from Cerner Health Facts (Cerner Corporation, North Kansas City, Mo), a database composed of electronic clinical records from hospital systems that have Cerner Corporation's electronic health record (ICD-9-CM) Analysis: Multivariable logistic regression was used to examine the association between postoperative hyperglycemia and infection, mortality, and readmission after adjusting for patient and hospital characteristics. We calculated odds ratios (ORs) and 95% CIs. We assessed model discrimination with the C statistic. Model calibration over the range of risk was assessed with the Hosmer–Lemeshow goodness-of-fit test.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation 1.2. Were cohorts from the same place? Low risk – same country 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included, with no adjustment for anatomy. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – age and sex adjusted for 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – broad range of individual comorbidities 2.3. Does study control appropriately for AAA characteristics? High risk – none 2.4. Could any adjustment variables have been affected by the intervention? High risk – postoperative hyperglycaemia and medications adjusted for **Data collection**

Full citation	**Tarbunou YA, Smith JB, Kruse RL, Vogel TR. Outcomes associated with hyperglycemia after abdominal aortic aneurysm repair. Journal of vascular surgery. 2019 Mar 1;69(3):763-73.**
	3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – administrative registry with high-level diagnosis and procedure codes 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk –administrative registry with high-level diagnosis and procedure codes 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – Hosmer–Lemeshow and c-statistic 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no consideration of missing data 4.3. Have different methods been compared within the study? High risk **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – 58 events and 18 covariates 6.2 Were interactions between treatment and other covariates considered? High risk – none **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Wald R, Waikar SS, Liangos O, et al. (2006) Acute renal failure after endovascular vs open repair of abdominal aortic aneurysm. J Vasc Surg. 43(3):460-466;**
Study details	Study design: retrospective cohort study Location(s): USA Study period: January to December 2002 Aim of the study: to compare outcomes of people who underwent EVAR and OSR for unruptured AAA.
Participants	Sample size: EVAR group, n=2,651; OSR group, n=3,865 Inclusion criteria: people over 18 years with a primary diagnosis of unruptured AAA who underwent EVAR or OSR were included. Note: people with complex AAA or symptomatic AAA may have also been included. Exclusion criteria: patients who underwent an aorto-renal bypass in addition to AAA repair and those receiving dialysis were excluded

Full citation	**Wald R, Waikar SS, Liangos O, et al. (2006) Acute renal failure after endovascular vs open repair of abdominal aortic aneurysm. J Vasc Surg. 43(3):460-466;**
	Baseline characteristics: Mean age (SD): EVAR group, 73.5 (10.4) years; OSR group, 71.6 (10.8) years Gender: EVAR group, 85.3% male; OSR group, 77.0% male Mean aneurysm diameter: not reported Chronic kidney disease: EVAR group, 7.3%; OSR group, 8.4% Congestive heart failure: EVAR group, 7.2%; OSR group, 11.6% Chronic lung disease: EVAR group, 30.8%; OSR group, 36.7% Chronic liver disease: EVAR group, 0.7%; OSR group, 1.0% Diabetes: EVAR group, 13.5%; OSR group, 10.3%
Methods	Data collection: participants were identified using diagnosis and procedure codes (such as ICD9 and OPCS4 codes) to query an administrative database that used high-level diagnosis and procedure codes: the Nationwide Inpatient Sample (NIS). NIS data is an anonymised, stratified sample of 20% of acute care non-federal American hospitals in 33 states. All data on comorbidities and outcomes were extracted from the NIS database. Analysis: Multivariate logistic regression was the primary method of analysis. Authors stated that all demographic, clinical, and hospital-related variables were included in the regression models. Furthermore, investigators selected multiplicative interaction terms (eg, chronic kidney disease*procedure type) to evaluate for effect modification. The Hosmer and Lemeshow test was used to assess model calibration. To adjust for additional unmeasured confounding, investigators fit a logistic regression model using procedure type as the dependent variable to generate propensity scores for receipt of EVAR. Only covariates with p values <0.05 were included in this model. The propensity score was then used to rank individuals according to their likelihood of receiving a given treatment and the association between procedure type and acute renal failure was then evaluated across quintiles of the propensity score.
Intervention	EVAR
Comparator	OSR
Outcomes	
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national patient database. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – all infrarenal and complex AAA included, which is likely to have led to implicitly different inclusion criteria across treatment arms.. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – the study controlled for age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a broad number of relevant comorbidities were controlled for.

Full citation	Wald R, Waikar SS, Liangos O, et al. (2006) Acute renal failure after endovascular vs open repair of abdominal aortic aneurysm. J Vasc Surg. 43(3):460-466;
	2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – an administrative registry with high-level diagnosis and procedure codes was used. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – an administrative registry with high-level diagnosis and procedure codes was used. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long term outcomes were assessed **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – model fit was assessed using the Hosmer–Lemeshow test. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no indication that the effect missing data were considered. 4.3. Have different methods been compared within the study? Low risk – different methods were compared but all relied on the same assumption about selection. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? Low risk – investigators selected multiplicative interaction terms (eg, chronic kidney disease*procedure type) to evaluate for effect modification **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	Williams TK, Schneider EB, Black JH, et al. (2013) Disparities in outcomes for Hispanic patients undergoing endovascular and open abdominal aortic aneurysm repair. Ann Vasc Surg. 27(1):29-37.
Study details	Study design: retrospective cohort study

Full citation	Williams TK, Schneider EB, Black JH, et al. (2013) Disparities in outcomes for Hispanic patients undergoing endovascular and open abdominal aortic aneurysm repair. Ann Vasc Surg. 27(1):29-37.
	Location(s): USA Study period: 2005 to 2008 Aim of the study: to examine the influence of race and ethnicity on the outcomes of EVAR and OSR of unruptured AAA and its effect on costs.
Participants	Sample size: EVAR group, n=62,728; OSR group, n=24,253 Inclusion criteria: people who underwent repair of unruptured AAA were included. Patients who possessed ICD-9-CM codes for both EVAR and OSR were included in the EVAR group, as this likely represented patients undergoing open conversion during the same hospitalisation. Analyses were performed using an intention-to-treat approach. Note: some people were operated upon under emergency circumstances (likely to be due to symptomatic AAA). Furthermore, there is no indication that the study was limited to people with infrarenal AAA. Exclusion criteria: people < 18 years, > 99 years, those with ruptured aneurysms, thoracic aneurysms, mycotic aneurysms, syphilitic aneurysms, or traumatic aneurysms were excluded. Baseline characteristics: Mean age: not reported Gender: EVAR group, 82.6% male; OSR group, 75.6% male Mean aneurysm diameter: not reported Cerebrovascular disease: EVAR group, 4.7%; OSR group, 5.1% Renal disease: EVAR group, 8.9%; OSR group, 10.8% Congestive heart failure: EVAR group, 7.9%; OSR group, 10.1% COPD: EVAR group, 31.2%; OSR group, 38.1% Emergency admission: EVAR group, 8.3%; OSR group, 20.7%
Methods	Data collection: participants were identified using diagnosis and procedure codes (such as ICD9 and OPCS4 codes) to query an administrative database that used high-level diagnosis and procedure codes: the Nationwide Inpatient Sample (NIS). NIS data is an anonymised, stratified sample of 20% of acute care non-federal American hospitals in 33 states. All data on comorbidities and outcomes were extracted from the NIS database. Analysis: multivariate regression analysis was performed adjusting for age, gender, race, comorbidities utilising the Charlson comorbidity index, procedure type, insurance type, and hospital characteristics. No further details were provided.
Intervention	EVAR
Comparator	OSR
Outcomes	In-hospital mortality
Study Appraisal	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period.

Full citation	Williams TK, Schneider EB, Black JH, et al. (2013) Disparities in outcomes for Hispanic patients undergoing endovascular and open abdominal aortic aneurysm repair. Ann Vasc Surg. 27(1):29-37.
using NICE's bespoke risk of bias assessment tool	1.2. Were cohorts from the same place? Low risk – all participants were selected from the same international database. 1.3. Is the definition of AAA the same across cohorts? High risk – a considerably higher proportion of participants in the EVAR group were treated under emergency admissions, indicating the likelihood of symptomatic AAAs. This highlights implicitly different inclusion criteria across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controls for age and sex. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – comorbidities (using the Charlson comorbidity index) were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – no AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – an administrative registry with high-level diagnosis and procedure codes was used. 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – an administrative registry with high-level diagnosis and procedure codes was used. 3.3. Is method of data collection likely to record long-term outcomes accurately? Low risk – no long term outcomes were assessed **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – there is no indication that the effect of missing data was considered. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? Low risk – number of events is ≥10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk

Full citation	**Williams TK, Schneider EB, Black JH, et al. (2013) Disparities in outcomes for Hispanic patients undergoing endovascular and open abdominal aortic aneurysm repair. Ann Vasc Surg. 27(1):29-37.**
	Directness: directly applicable

E.3 Complex AAAs

Full citation	**de Guerre LE, Varkevisser RR, Swerdlow NJ, Liang P, Li C, Dansey K, van Herwaarden JA, Schermerhorn ML. Sex differences in perioperative outcomes after complex abdominal aortic aneurysm repair. Journal of vascular surgery. 2019 Jul 4.**
Study details	Study design: retrospective cohort Location(s): USA Study period: 2011 to 2017 Aim of the study: to evaluate the association of female sex and perioperative outcomes after endovascular and open complex AAA repair in a nationwide registry
Participants	Sample size: EVAR group, n=1,260; OSR group, n= 1,010 Inclusion criteria: patients undergoing endovascular or open repair of complex AAAs (defined as a proximal extent listed as juxtarenal, pararenal, or suprarenal or open procedures coded as repair of an AAA involving visceral vessels or EVAR using Cook Zenith Fenestrated Endovascular Graft) Exclusion criteria: open repair with an infrarenal proximal clamp position, emergency repair, patients with prior AAA repair, ruptured AAAs and thoracoabdominal aneurysms. Baseline characteristics: Median age (IQR): Women 75 (69–80), men 73 (67–79) Gender: EVAR group, 78.6% male; OSR group, 69.3% male Median aneurysm diameter: EVAR: women 5.5 (5.1–6), men 5.6 (5.1–6.2); OSR: women 5.7 (5.2–6.4), men 6 (5.5–6.75) Diabetes (insulin dependent): women 2.4%, men 2.5% Hypertension: women 80.2%; men 82.1 % COPD: women 22.2%; men 19.7% Congestive heart failure: women 1.7%; men 2.3%
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). Analysis: Authors examined independent associations between endovascular and open repair with the outcomes for female and male patients separately. Propensity scores calculated using logistic regression models and used to create inverse probability weights. No adjustment for anatomic complexity (to allow 'the inherent anatomic differences between female and male patients to persist').
Intervention	Complex EVAR

Full citation	**de Guerre LE, Varkevisser RR, Swerdlow NJ, Liang P, Li C, Dansey K, van Herwaarden JA, Schermerhorn ML. Sex differences in perioperative outcomes after complex abdominal aortic aneurysm repair. Journal of vascular surgery. 2019 Jul 4.**
Comparator	Complex OSR
Outcomes	Perioperative mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – recruitment over ≥5 yrs, unclear that year of operation was controlled for in analysis 1.2. Were cohorts from the same place? Low risk – cohorts were derived from the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – OSR excluded all infrarenal clamps (which is likely to include some cases that would be classified as juxtarenal if addressed with EVAR). **Confounding** 2.1. Does study control appropriately for demographics? Low risk – separate analyses for men and women, each controlling for demographics. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – good range of individual comorbidities. 2.3. Does study control appropriately for AAA characteristics? Moderate risk – just diameter. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – participants identified a detailed surgical registry with procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data obtained from a detailed surgical registry with procedure codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks appear to have been conducted on model specification/fit. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors highlight that the NSQIP database required hospitals to provide complete 30-day follow-up on at least 95% of patients. 4.3. Have different methods been compared within the study? High risk - different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models**

Full citation	**de Guerre LE, Varkevisser RR, Swerdlow NJ, Liang P, Li C, Dansey K, van Herwaarden JA, Schermerhorn ML. Sex differences in perioperative outcomes after complex abdominal aortic aneurysm repair. Journal of vascular surgery. 2019 Jul 4.**
	6.1 Is sample size adequate relative to number of covariates considered? High risk – 103 events; unknown number of covariates; probably >10 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions considered. **Overall risk of bias:** Moderate risk **Directness:** directly applicable

Full citation	**Fiorucci B, Speziale F, Kölbel T, Tsilimparis N, Sirignano P, Capoccia L, Simonte G, Verzini F. Short-and Midterm Outcomes of Open Repair and Fenestrated Endografting of Pararenal Aortic Aneurysms in a Concurrent Propensity-Adjusted Comparison. Journal of Endovascular Therapy. 2019 Feb;26(1):105-12.**
Study details	Study design: retrospective cohort Location(s): Germany and Italy Study period: 1998 to 2016 (OSR) / 2006 to 2015 (fEVAR) Aim of the study: to compare short- and midterm outcomes of patients treated with fEVAR and OSR for pararenal aortic aneurysms in patients from a group of high-volume centres in which both techniques were sufficiently well established to allow an up-to-date comparison of results
Participants	Sample size: EVAR group, n=41; OSR group, n= 102 Inclusion criteria: unclear (‘pararenal aortic aneurysms… consecutive patients electively treated with OSR or fEVAR’) Exclusion criteria: urgent or emergent treatment for symptomatic or ruptured aneurysms Baseline characteristics: Mean age: fEVAR 73, OSR 71 Gender: fEVAR group, 95.1% male; OSR group, 94.1% male Aneurysm diameter: NR Diabetes: fEVAR 4.9%, OSR 15.7% Hypertension: fEVAR 78.0%, OSR 89.2% COPD: fEVAR 36.6%, OSR 52.9% Renal failure: fEVAR 12.2%, OSR 28.4% Coronary artery disease: fEVAR 43.9%, OSR 32.4%
Methods	Data collection: Vascular databases containing prospectively collected data at 3 tertiary institutions were merged. Analysis: A propensity score according to type of treatment was constructed from a binary logistic regression using age, gender, CAD, and renal failure with a matching method that selected more than one participant from the OSR group for every patient in the fEVAR group.

Full citation	**Fiorucci B, Speziale F, Kölbel T, Tsilimparis N, Sirignano P, Capoccia L, Simonte G, Verzini F. Short-and Midterm Outcomes of Open Repair and Fenestrated Endografting of Pararenal Aortic Aneurysms in a Concurrent Propensity-Adjusted Comparison. Journal of Endovascular Therapy. 2019 Feb;26(1):105-12.**
Intervention	Fenestrated EVAR
Comparator	Complex OSR
Outcomes	Perioperative mortality, perioperative complications (all, cardiovascular, respiratory, renal)
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? High risk – cohorts drawn from different periods in time 1.2. Were cohorts from the same place? Low risk – cohorts were derived from the same centres. 1.3. Is the definition of AAA the same across cohorts? Moderate risk – no details of inclusion criteria. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – age and sex controlled for. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – only 2 individual comorbidities (coronary artery disease and renal failure). 2.3. Does study control appropriately for AAA characteristics? High risk – none. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – institution-specific datasets used but methods unclear. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – institution-specific datasets. 3.3. Is method of data collection likely to record long-term outcomes accurately? High risk – medical records (although 'missing data for those with an overdue follow-up >18 months were obtained when feasible through telephone interviews with patients, family, or general practitioners') **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks appear to have been conducted on model specification/fit. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – possible differential missingness – especially given asymmetrical recruitment periods – with no valid adjustment considered. 4.3. Have different methods been compared within the study? High risk - different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? High risk – method not reported 5.2. Was overlap / common support appropriately assessed? High risk – no assessment reported

Full citation	**Fiorucci B, Speziale F, Kölbel T, Tsilimparis N, Sirignano P, Capoccia L, Simonte G, Verzini F. Short-and Midterm Outcomes of Open Repair and Fenestrated Endografting of Pararenal Aortic Aneurysms in a Concurrent Propensity-Adjusted Comparison. Journal of Endovascular Therapy. 2019 Feb;26(1):105-12.**
	5.3. Has balancing of the covariates been demonstrated? High risk – not reported **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – not reported how many covariates were considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Gupta PK, Brahmbhatt R, Kempe K, et al. (2017) Thirty-day outcomes after fenestrated endovascular repair are superior to open repair of abdominal aortic aneurysms involving visceral vessels. J Vasc Surg. 66(6):1653-1658.**
Study details	Study design: retrospective cohort Location(s): USA Study period: 2008 to 2013 Aim of the study: to compare 30-day outcomes after FEVAR and OSR of AAAs involving visceral vessels.
Participants	Sample size: FEVAR group, n=535; OSR group, n= 1,207 Inclusion criteria: people who underwent EVAR or OSR of AAAs involving visceral vessels were included. Exclusion criteria: people who underwent thoracoabdominal aneurysm repair (CPT 33877 and ICD-9-CM 441.7) were excluded. Baseline characteristics: Mean age (range): FEVAR group, 75 (69-82) years; OSR group, 72 (66-77) years Gender: FEVAR group, 81.9% male; OSR group, 71.7% male Mean aneurysm diameter: not reported Diabetes: FEVAR group, 15.5%; OSR group, 10.8% Hypertension: FEVAR group, 79.4%; OSR group, 82.4% COPD: FEVAR group, 23.7%; OSR group, 20.6% Previous myocardial infarction, cardiac surgery or percutaneous cardiac intervention: FEVAR group, 26.4%; OSR group, 23.8%
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). The NSQIP database requires hospitals to provide complete 30-day follow-up on at least 95% of patients. Analysis: Forward stepwise multivariate logistic regression was performed. The inclusion criterion for multivariate analysis was a p value <0.1 on univariate analysis.

Full citation	**Gupta PK, Brahmbhatt R, Kempe K, et al. (2017) Thirty-day outcomes after fenestrated endovascular repair are superior to open repair of abdominal aortic aneurysms involving visceral vessels. J Vasc Surg. 66(6):1653-1658.**
Intervention	FEVAR
Comparator	OSR
Outcomes	30-day mortality, cardiac arrest, renal failure, and respiratory failure
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.4. Were cohorts from the same time period? Moderate risk – recruitment over ≥5 yrs, unclear that year of operation was controlled for in analysis 1.5. Were cohorts from the same place? Low risk – cohorts were derived from the same national surgical registry. 1.6. Is the definition of AAA the same across cohorts? Low risk – all participants had had complex aneurysms involving visceral vessels. **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – model for primary outcome measure (30-day mortality) only controlled for age. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate risk – model for primary outcome measure (30-day mortality) only controlled for history of COPD. 2.3. Does study control appropriately for AAA characteristics? High risk – there is no indication that AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – participants identified a detailed surgical registry with procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data obtained from a detailed surgical registry with procedure codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks appear to have been conducted on model specification/fit. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors highlight that the NSQIP database required hospitals to provide complete 30-day follow-up on at least 95% of patients. 4.3. Have different methods been compared within the study? High risk - different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A

Full citation	**Gupta PK, Brahmbhatt R, Kempe K, et al. (2017) Thirty-day outcomes after fenestrated endovascular repair are superior to open repair of abdominal aortic aneurysms involving visceral vessels. J Vasc Surg. 66(6):1653-1658.**
	5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk – not reported how many covariates were considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Locham S, Faateh M, Dakour-Aridi H et al. (2018) Octogenarians Undergoing Open Repair Have Higher Mortality Compared with Fenestrated Endovascular Repair of Intact Abdominal Aortic Aneurysms Involving the Visceral Vessels. Ann Vasc Surg. 51:192-199.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2006 to 2015 Aim of the study: to compare 30-day outcomes of FEVAR versus OSR in octogenarians undergoing repair of AAA involving the visceral vessels
Participants	Sample size: FEVAR group, n=242; OSR group, n=306 Inclusion criteria: people 80 ≥ years who underwent FEVAR or OSR for unruptured complex AAA involving visceral vessels were included. Exclusion criteria: concomitant open repairs, emergent cases, and patients <80 years or >90 years were excluded. Baseline characteristics: Median age (IQR): FEVAR group, 83 (82-86) years; OSR group,82 (81-85) years Gender: FEVAR group, 81.7% male; OSR group, 64.1% male Mean aneurysm diameter: not reported Diabetes: FEVAR group, 12.8%; OSR group, 5.6% COPD: FEVAR group, 18.2%; OSR group, 15.4% Congestive heart failure: FEVAR group, 1.2%; OSR group, 0.7% Hypertension: FEVAR group, 82.2%; OSR group, 83.3% Disseminated cancer: FEVAR group, 0.8%; OSR group, 0.7% Bleeding disorders: FEVAR group, 9.1%; OSR group, 10.1%

Full citation	**Locham S, Faateh M, Dakour-Aridi H et al. (2018) Octogenarians Undergoing Open Repair Have Higher Mortality Compared with Fenestrated Endovascular Repair of Intact Abdominal Aortic Aneurysms Involving the Visceral Vessels. Ann Vasc Surg. 51:192-199.**
Methods	Data collection: investigators identified participants and obtained data on their outcomes by querying the American College of Surgeons version of the National Surgical Quality Improvement Program (ACS-NSQIP) database using procedure codes. Authors highlighted that the ACS-NSQIP database routinely collects information based on patient's medical charts rather than billing data. Analysis: Multivariate logistic regression was performed. Covariates were chosen on the basis of clinical and statistical significance in univariate analysis (significance level not specified). All models were tested using variation inflation factor, Hosmer-Lemeshow goodness of fit test, and area under the receiver operating characteristic curve (C-statistic).
Intervention	FEVAR
Comparator	OSR
Outcomes	30-day mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA was similar across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – demographics, including age and gender, were controlled for. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a broad range of comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – the study did not control for AAA characteristics. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – a detailed surgical registry was used to identify participants. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed using a detailed surgical registry was used with procedure and diagnosis codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A– no long-term outcomes were assed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – checks were performed using the Hosmer–Lemeshow test and the C-statistic 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – there is no indication that the impact of missing data was considered. 4.3. Have different methods been compared within the study? High risk – different methods were not compared.

Full citation	**Locham S, Faateh M, Dakour-Aridi H et al. (2018) Octogenarians Undergoing Open Repair Have Higher Mortality Compared with Fenestrated Endovascular Repair of Intact Abdominal Aortic Aneurysms Involving the Visceral Vessels. Ann Vasc Surg. 51:192-199.**
	Analysis – matching 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk - number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** partially applicable

Full citation	**Locham S, Dakour-Aridi H, Bhela J, Nejim B, Bhavana Challa A, Malas M. Thirty-Day Outcomes of Fenestrated and Chimney Endovascular Repair and Open Repair of Juxtarenal, Pararenal, and Suprarenal Abdominal Aortic Aneurysms Using National Surgical Quality Initiative Program Database (2012-2016). Vascular and endovascular surgery. 2019 Apr;53(3):189-98.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2012 to 2016 Aim of the study: to compare short-term outcomes between endovascular (FEVAR and ChEVAR) and open repair of patients with suprarenal, juxtarenal, and pararenal AAAs using a large national surgical database
Participants	Sample size: fEVAR group, n=162; ChEVAR group, n=164, OSR group, n=865 Inclusion criteria: All patients undergoing endovascular (fEVAR or ChEVAR) and open repair of juxtarenal, pararenal, and suprarenal AAA. Exclusion criteria: emergent, outpatient, ruptured, TAA type IV, acute conversion to open repair, and not documented/infrarenal aneurysms. Baseline characteristics: Median age (IQR): fEVAR 74 (69–80) years; ChEVAR 75 (69–81); OSR 72 (66–77) years Gender: fEVAR 82.7% male; ChEVAR 70.7% male; OSR 71.3% male Median aneurysm diameter: fEVAR 5.8; ChEVAR 5.8; OSR 5.9 Diabetes: fEVAR 17.9%; ChEVAR 11.0%; OSR 11.9% COPD: fEVAR 29.0%; ChEVAR 23.8%; OSR 24.6% Congestive heart failure: fEVAR 3.7%; ChEVAR 3.7%; OSR 2.2%

Full citation	**Locham S, Dakour-Aridi H, Bhela J, Nejim B, Bhavana Challa A, Malas M. Thirty-Day Outcomes of Fenestrated and Chimney Endovascular Repair and Open Repair of Juxtarenal, Pararenal, and Suprarenal Abdominal Aortic Aneurysms Using National Surgical Quality Initiative Program Database (2012-2016). Vascular and endovascular surgery. 2019 Apr;53(3):189-98.**
	Hypertension: fEVAR 84.6%; ChEVAR 87.2%; OSR 83.0% Chronic renal failure: fEVAR 44.7%; ChEVAR 45.0%; OSR 40.0% Prior AAA surgery: fEVAR 29.0%; ChEVAR 34.9%; OSR 31.7%
Methods	Data collection: investigators identified participants and obtained data on their outcomes by querying the American College of Surgeons version of the National Surgical Quality Improvement Program (ACS-NSQIP) database using procedure codes. Analysis: Multivariate logistic regression was performed. Statistically significant and clinically relevant covariates based on univariate analysis and prior literature were included. All models were tested using Hosmer–Lemeshow goodness-of-fit test and area under receiver operative curve.
Intervention	fEVAR / ChEVAR
Comparator	OSR
Outcomes	30-day mortality; 30-day renal failure; 30-day cardiopulmonary failure
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA was similar across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – demographics, including age and gender, were controlled for. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a broad range of comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? Low risk – diameter and distal extent. 2.4. Could any adjustment variables have been affected by the intervention? High risk – controls for transfusion (appears to be perioperative). **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – a detailed surgical registry was used to identify participants. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed using a detailed surgical registry was used with procedure and diagnosis codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A– no long-term outcomes were assed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? Low risk – checks were performed using the Hosmer–Lemeshow test and the C-statistic

Full citation	**Locham S, Dakour-Aridi H, Bhela J, Nejim B, Bhavana Challa A, Malas M. Thirty-Day Outcomes of Fenestrated and Chimney Endovascular Repair and Open Repair of Juxtarenal, Pararenal, and Suprarenal Abdominal Aortic Aneurysms Using National Surgical Quality Initiative Program Database (2012-2016). Vascular and endovascular surgery. 2019 Apr;53(3):189-98.**
	4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – there is no indication that the impact of missing data was considered. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk - number of events is <10 times greater than number of variables considered. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Michel M, Becquemin J-P, Clément M-C, et al. (2015) Editor's choice – thirty day outcomes and costs of fenestrated and branched stent grafts versus open repair for complex aortic aneurysms. Eur J Vasc Endovasc Surg. 50:189-196**
Study details	Study design: retrospective cohort study Location(s): France Study period: 2010 - 2012 Aim of the study: to compare 30 day outcomes and costs of fenestrated and branched stent grafts (f/b EVAR) and open surgery (OSR) for the treatment of complex abdominal aortic aneurysms (AAA) and thoraco-abdominal aortic aneurysms (TAAA).
Participants	Sample size: EVAR group, n=268; OSR group, n=1,678 Inclusion criteria: high risk for open surgery; had an AAA >50 mm in men (45 mm in women), with or without thoracic aortic aneurysm >55 mm (50 mm in women), and with an infrarenal neck <10 mm in length or aneurysm extending to the suprarenal aorta Exclusion criteria: emergent and ruptured aneurysms as well as aortic dissections Baseline characteristics: Mean age (SD): EVAR group, 71.6 (8.5) years; OSR group, 69.2 (8.9) years Gender: EVAR group, 93.3% male; OSR group, 91.7% male

Full citation	**Michel M, Becquemin J-P, Clément M-C, et al. (2015) Editor's choice – thirty day outcomes and costs of fenestrated and branched stent grafts versus open repair for complex aortic aneurysms. Eur J Vasc Endovasc Surg. 50:189-196**
	Mean aneurysm diameter: not reported Para/juxtarenal AAA: EVAR group, 68.6%; OSR group, 82.4% Infradiaphragmatic TAAA: EVAR group, 15.7%; OSR group, 13.4% Supradiaphragmatic AAA: EVAR group, 15.7%; OSR group, 4.2% Hypertension: EVAR group, 61.5%; OSR group, 51.1% Hyperlipidemia: EVAR group, 42.4%; OSR group, 34.5% Diabetes: EVAR group, 14.5%; OSR group, 12.5% Coronary artery occlusive disease: EVAR group, 9.2%; OSR group, 8.2% Peripheral arterial disease: EVAR group, 8.0%; OSR group, 14.5% Cardiac insufficiency: EVAR group, 7.3%; OSR group, 3.2% Chronic pulmonary disease: EVAR group, 23.3%; OSR group, 14.4% Chronic renal disease: EVAR group, 8.8%; OSR group, 6.2%
Methods	Data collection: data for EVAR was collected using a multicentre prospective registry. Data for OSR was collected from the national hospital discharge database Analysis: multivariate analyses were performed on 30 day mortality using a Cox model
Intervention	EVAR
Comparator	OSR
Outcomes	30-day mortality
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk 1.2. Were cohorts from the same place? Low risk – same country 1.3. Is the definition of AAA the same across cohorts? High risk – EVAR cohort includes a greater proportion of thoracoabdominal AAAs **Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk 2.3. Does study control appropriately for AAA characteristics? Moderate risk – just extent 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection**

Full citation	Michel M, Becquemin J-P, Clément M-C, et al. (2015) Editor's choice – thirty day outcomes and costs of fenestrated and branched stent grafts versus open repair for complex aortic aneurysms. Eur J Vasc Endovasc Surg. 50:189-196
	3.1. Is method of data collection likely to have identified suitable participants accurately? High risk – different databases were used for each arm 3.2. Is method of data collection likely to record perioperative outcomes accurately? High risk – detailed surgical registries but different ones for each arm 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long-term outcomes were assed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks reported 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk 4.3. Have different methods been compared within the study? Moderate risk – different methods were compared, but all rely on assumption of selection on observables **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? High risk – not reported 5.2. Was overlap / common support appropriately assessed? High risk – not reported 5.3. Has balancing of the covariates been demonstrated? High risk – not reported **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** Partially applicable (includes some TAAAs that are likely to be out of scope)

Full citation	Orr NT, Davenport DL, Minion DJ et al. (2017) Comparison of perioperative outcomes in endovascular versus open repair for juxtarenal and pararenal aortic aneurysms: A propensity-matched analysis. Vascular. 25(4):339-345.
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2012 to 2015 Aim of the study: to compare 30-day outcomes of EVAR versus OSR of juxtarenal and pararenal aortic aneurysms
Participants	Sample size of unmatched cohort: complex EVAR group, n=395; OSR group, n=610 Sample size of matched cohort: complex EVAR group, n=263; OSR group, n=263

Full citation	**Orr NT, Davenport DL, Minion DJ et al. (2017) Comparison of perioperative outcomes in endovascular versus open repair for juxtarenal and pararenal aortic aneurysms: A propensity-matched analysis. Vascular. 25(4):339-345.**
	Inclusion criteria: all patients with juxtarenal or pararenal AAAs treated by EVAR or OSR between 2012 and 2015 were included Exclusion criteria: failed prior repairs, ruptured aneurysms or dissected aneurysms were excluded Baseline characteristics (of matched cohort): % <65 years: complex EVAR group, 15%; OSR group, 16% % >80 years: complex EVAR group, 21%; OSR group, 21% Gender: complex EVAR group, 74% male; OSR group, 76% male Mean aneurysm diameter: not reported Diabetes: complex EVAR group, 14%; OSR group, 10% Severe COPD: complex EVAR group, 21%; OSR group, 26% Coronary heart failure: complex EVAR group, 3.0%; OSR group, 3.4% Hypertension receiving treatment: complex EVAR group, 84%; OSR group, 83% Bleeding disorder: complex EVAR group, 12%; OSR group, 10%
Methods	Data collection: Data collection: investigators identified participants and obtained data on their outcomes by querying the American College of Surgeons version of the National Surgical Quality Improvement Program (ACS-NSQIP) database which contained information such as the proximal and distal extents of the aneurysm, specific operative characteristics, and 30-day postoperative vascular outcomes in both inpatient and outpatient settings. Analysis: Propensity score matching was performed to clinically match OSR and EVAR groups on preoperative risk and select perioperative factors that differed significantly in the unmatched groups (greedy nearest neighbour matching, caliper <0.15 standard deviations). Authors do not explicitly list what these factors were but examination of tables within the manuscript highlight that that the following factors were significantly different between groups: mean age, ASA class, % with acute renal failure, % smokers, % with bleeding disorders, % who had same day elective surgery, mean duration of operation, % juxtarenal/pararenal, distal extent of the aneurysm, and renal stent placement. For the purpose of this review it is assumed that all these factors were controlled for when deriving propensity scores. Group comparisons were then performed between the matched groups.
Intervention	Complex EVAR
Comparator	OSR
Outcomes	30-day mortality, length of stay, length of stay in ICU, discharge to home, reintervention (labelled return to OR), and adverse events (including Cardiac or respiratory failure, surgical site infection or dehiscence, renal insufficiency or failure, pneumonia, sepsis, DVT and pulmonary embolism)
Study Appraisal using NICE's	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants were selected from the same national surgical registry.

Full citation	**Orr NT, Davenport DL, Minion DJ et al. (2017) Comparison of perioperative outcomes in endovascular versus open repair for juxtarenal and pararenal aortic aneurysms: A propensity-matched analysis. Vascular. 25(4):339-345.**
bespoke risk of bias assessment tool	1.3. Is the definition of AAA the same across cohorts? Low risk – there was a significantly higher proportion of juxtarenal AAA in the OSR group in the unmatched cohort. Following propensity score matching this difference became non-significant. **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – as mentioned above (analysis), it is likely that only age was controlled for. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – although it was not explicitly stated, it is likely that a good range of comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? Low risk – although it was not explicitly stated, it is likely that AAA characteristics were controlled for. 2.4. Could any adjustment variables have been affected by the intervention? High risk – authors highlight that perioperative factors were controlled for. Some of which may have mediated the treatment effect. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – a detailed surgical registry was used to identify relevant participants. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed using a detailed surgical registry. 3.3. Is method of data collection likely to record long-term outcomes accurately?N/A – no long-term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – authors do not report how or if missing data was accounted for in their analyses. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Low risk - greedy nearest neighbour matching was performed; caliper <0.15 standard deviations 5.2. Was overlap / common support appropriately assessed? High risk – no assessment was reported. 5.3. Has balancing of the covariates been demonstrated? Moderate - conventional hypothesis tests were performed, with no evidence of significant differences. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A

Full citation	Orr NT, Davenport DL, Minion DJ et al. (2017) Comparison of perioperative outcomes in endovascular versus open repair for juxtarenal and pararenal aortic aneurysms: A propensity-matched analysis. Vascular. 25(4):339-345.
	Overall risk of bias: High risk **Directness:** directly applicable

Full citation	Raux M, Patel VI, Cochennec F, et al. (2014) A propensity-matched comparison of outcomes for fenestrated endovascular aneurysm repair and open surgical repair of complex abdominal aortic aneurysms. J Vasc Surg.60(4):858-63
Study details	Study design: retrospective cohort study Location(s): USA Study period: July 2001 to August 2012 Aim of the study: compare 30-day outcomes of FEVAR and OSR at 2 high-volume centres where FEVAR was undertaken for high-risk patients
Participants	Sample size on matched cohort: FEVAR group, n=42; OSR group, n=147 Inclusion criteria: people with complex aneurysms who underwent elective FEVAR or OSR were included. Only patients who would have required an actual or anticipated completely suprarenal or more proximal clamp position were included in the study. Exclusion criteria: people with type I-IV thoracoabdominal aneurysms, ruptured or symptomatic aneurysms, patients with a redo aortic surgery or a history of aortic intervention, and patients with actual or anticipated infrarenal clamp position were excluded. Baseline characteristics (of matched cohort): Mean age (SD): FEVAR group, 73 (10) years; OSR group, 73 (7.8) years Gender: FEVAR group, 88% male; OSR group, 82% male Mean aneurysm diameter: not reported Hypertension: FEVAR group, 74%; OSR group, 80% Myocardial infarction: FEVAR group, 26%; OSR group, 36% Chronic heart failure: FEVAR group, 14%; OSR group, 12% Coronary artery disease: FEVAR group, 43%; OSR group, 34% COPD: FEVAR group, 36%; OSR group, 25% Cerebrovascular accident: FEVAR group, 7.1%; OSR group, 7.5% Diabetes: FEVAR group, 19%; OSR group, 14%
Methods	Data collection: participants were identified by retrospective review of medical records from 2 high-volume medical centres: one centre only performed OSR and the other only performed FEVAR. Patients who received FEVAR were considered high-risk for OSR (High-risk criteria did not consider aneurysm morphology).

Full citation	**Raux M, Patel VI, Cochennec F, et al. (2014) A propensity-matched comparison of outcomes for fenestrated endovascular aneurysm repair and open surgical repair of complex abdominal aortic aneurysms. J Vasc Surg.60(4):858-63**
	Analysis: Propensity score matching was performed. Initially, multivariate regression was used to generate propensity scores by controlling for all variables that were found to be significantly associated (p values<0.05) with the odds of performing FEVAR in univariate analyses. These included demographic variables, multiple relevant comorbidities, as well as actual/ anticipated clamp location. Propensity score matching was then performed using the caliper method, matching each case (FEVAR) with four controls (OSR) ≤0.2 standard deviations of the propensity score. The propensity matched groups were then compared using univariate methods **as well as** multivariate analyses (using multivariate logistic regression).
Intervention	FEVAR
Comparator	OSR
Outcomes	30-day mortality, and adverse events (including procedural and graft complications, cardiac, renal, and respiratory complications) within 30 days of treatment
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation 1.2. Were cohorts from the same place? High risk – patients who underwent FEVAR and those who underwent OSR were treated at different hospitals. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA is the same across cohorts. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – the study controls for age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – a good range of relevant comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? Low risk – the study controlled for clamp location which is a proxy for aneurysm type/location. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – participants were identified by reviewing medical records. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – outcomes were assessed using medical records 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long term data were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were reported.

Full citation	**Raux M, Patel VI, Cochennec F, et al. (2014) A propensity-matched comparison of outcomes for fenestrated endovascular aneurysm repair and open surgical repair of complex abdominal aortic aneurysms. J Vasc Surg.60(4):858-63**
	4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – due to the nature in which data were collected (direct review of medical records for short-term outcomes) it is unlikely that there would be a high amount of missing data. 4.3. Have different methods been compared within the study? Moderate risk – different methods were compared but all relied on the same assumption about selection. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Low risk – matching was performed using the caliper method, matching each case with four controls ≤0.2 standard deviations of the propensity score. 5.2. Was overlap / common support appropriately assessed? High risk – no assessment was reported 5.3. Has balancing of the covariates been demonstrated? Moderate risk – conventional hypothesis tests were performed, with no evidence of significant differences. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Tinelli G, Crea MA, de Waure C, et al. (2018) A propensity matched comparison of fenestrated endovascular aneurysm repair and open surgical repair of pararenal and paravisceral aortic aneurysms. J Vasc Surg. March:1-10**
Study details	Study design: retrospective propensity-matched cohort study Location(s): Italy Study period: January 2010 to June 2016 Aim of the study: This study investigated the outcomes of a current series of patients treated with fenestrated and branched endovascular aneurysm repair (FEVAR) or open surgical repair (OSR) for pararenal abdominal aortic aneurysms (pr-AAAs), including juxtarenal, suprarenal, and type IV thoracoabdominal aneurysms. This study compares the outcomes of these procedures from two high-volume centers without the bias induced by a learning curve.
Participants	Sample size: FEVAR group, n=102; OSR group, n=102 Inclusion criteria: all patients with a pr-AAA requiring suprarenal or supravisceral proximal clamping were included in the study

Full citation	**Tinelli G, Crea MA, de Waure C, et al. (2018) A propensity matched comparison of fenestrated endovascular aneurysm repair and open surgical repair of pararenal and paravisceral aortic aneurysms. J Vasc Surg. March:1-10**
	Exclusion criteria: all F-BEVAR patients were deemed unsuitable for OSR after multidisciplinary evaluation because of high-risk comorbidities. The study excluded patients treated for extent I to III thoracoabdominal aneurysms, ruptured or symptomatic aneurysms, and dissections or connective tissue disorder aneurysms. Baseline characteristics: Mean age (SD): FEVAR group, 71.8 (8.0) years; OSR group, 71.7 (7.0) years Gender: FEVAR group, 95.1% male; OSR group, 92.2% male Mean aneurysm diameter(SD): FEVAR group, 59.8 (8.8) cm; OSR group, 60.6 (9.3) cm Coronary artery disease: FEVAR group, 42.2%; OSR group, 38.2% COPD: FEVAR group, 40.2%; OSR group, 38.2% Chronic kidney disease: FEVAR group, 24.5%; OSR group, 27.5% Diabetes: FEVAR group, 12.7%; OSR group, 11.8%
Methods	Data collection: This retrospective cohort study compared the outcomes of FEVAR and OSR for pr-AAA by analysing prospectively collected data from two centres: the Aortic Center (ACL; Lille, France) and the Vascular Unit of Fondazione Policlinico Universitario Gemelli (FPUG; Rome, Italy) Analysis: Propensity matching
Intervention	FEVAR
Comparator	OSR
Outcomes	30-day mortality, in-hospital mortality, end of study survival at a median follow-up of 38.9 months, theatre time, fluoroscopy time, adverse events, reintervention
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk - >5 years of recruitment without adjustment for year of operation 1.2. Were cohorts from the same place? High risk – 2 different centres 1.3. Is the definition of AAA the same across cohorts? Low risk **Confounding** 2.1. Does study control appropriately for demographics? Low risk 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk 2.3. Does study control appropriately for AAA characteristics? Low risk –diameter and (anticipated) clamp level (as a proxy of proximal extent) 2.4. Could any adjustment variables have been affected by the intervention? Low risk **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Low risk – medical records

Full citation	**Tinelli G, Crea MA, de Waure C, et al. (2018) A propensity matched comparison of fenestrated endovascular aneurysm repair and open surgical repair of pararenal and paravisceral aortic aneurysms. J Vasc Surg. March:1-10**
	3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – medical records 3.3. Is method of data collection likely to record long-term outcomes accurately? High risk – no details given; reliance on medical records alone would be high risk **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – no details provided 4.3. Have different methods been compared within the study? High risk – different methods were not compared **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? Moderate risk – caliper method – stated that threshold was 2 standard deviations of the propensity score (this is unusually high; it is possible the authors mean 0.2 SDs, which is common) 5.2. Was overlap / common support appropriately assessed? Low risk – balance assessment was made using various tests and checking quantile-quantile plots 5.3. Has balancing of the covariates been demonstrated? Moderate risk – conventional hypothesis tests were performed, with no evidence of significant differences. **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Tsilimparis N, Perez S, Dayama A, et al. (2013) Endovascular repair with fenestrated-branched stent grafts improves 30-day outcomes for complex aortic aneurysms compared with open repair. Ann Vasc Surg. 27(3): 267-73.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2005 to 2010 Aim of the study: to compare the real-world operative and perioperative outcomes of FEVAR and OSR for complex AAA,
Participants	Sample size: FEVAR group, n=264 group, n=1,091 Inclusion criteria: patients who underwent FEVAR or OSR for unruptured complex AAAs (juxtarenal and pararenal aneurysms and type IV thoracoabdominal aortic aneurysms) were included. Exclusion criteria: not reported. Baseline characteristics: Mean age (SD): FEVAR group, 74 (9) years; OSR group, 71 (9) years Gender: FEVAR group, 82.2% male; OSR group, 71.5% male Mean aneurysm diameter: not reported Diabetes: FEVAR group, 16%; OSR group, 11% Severe COPD: FEVAR group, 21.6%; OSR group, 20.5% Previous PCI: FEVAR group,19.3%; OSR group, 19.4% Previous cardiac surgery: FEVAR group, 19.7%; OSR group, 23.5% Previous cardiac surgery: FEVAR group, 75%; OSR group, 85.2% Cardiovascular accident/stroke with neurologic deficit: FEVAR group, 2.3%; OSR group, 4.6% Previous operation within 30 days: FEVAR group, 1.9%; OSR group, 1.6%
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). Patients who underwent repair procedures for complex AAAs were identified using diagnostic codes and procedure codes. Analysis: Multivariable logistic regression was performed. Authors state that confounders were identified through running regression models with type of repair and one additional preoperative risk factor or demographic variable at a time as predictors and observing how the results differed from running a logistic model using age alone. A change of more than 10% between the crude and adjusted odds ratio of age was used as evidence that the covariate was a possible confounder. A final logistic regression model was run using type of repair and all confounders found in this way.
Intervention	FEVAR
Comparator	OSR
Outcomes	30-day mortality

Full citation	Tsilimparis N, Perez S, Dayama A, et al. (2013) Endovascular repair with fenestrated-branched stent grafts improves 30-day outcomes for complex aortic aneurysms compared with open repair. Ann Vasc Surg. 27(3): 267-73.
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – ≥5-yr recruitment with no adjustment for year of operation. 1.2. Were cohorts from the same place? Low risk – cohorts were drawn from the same national surgical database. 1.3. Is the definition of AAA the same across cohorts? High risk – no details of extent of complexity, which is likely to differ between OSR and EVAR, especially over 6-year period during which complex EVAR evolved **Confounding** 2.1. Does study control appropriately for demographics? Moderate risk – study only controls for age. 2.2. Does study control appropriately for comorbidity and/or fitness? Moderate – a limited number of individual comorbidities were controlled for. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables that could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – data were collected from a detailed surgical registry. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data were collected from a detailed surgical registry. 3.3. Is method of data collection likely to record long-term outcomes accurately?N/A – no long term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no details about checks for model specification/fit were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – there is no indication that missing data were accounted for in the analyses. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? High risk - number of variables considered is not reported, and is likely to be ≥1/10 number of events.

Full citation	**Tsilimparis N, Perez S, Dayama A, et al. (2013) Endovascular repair with fenestrated-branched stent grafts improves 30-day outcomes for complex aortic aneurysms compared with open repair. Ann Vasc Surg. 27(3): 267-73.**
	6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Ultee KHJ, Zettervall SL, Soden PA, et al. (2017) Perioperative outcome of endovascular repair for complex abdominal aortic aneurysms. J Vasc Surg. 65(6):1567-1575.**
Study details	Study design: retrospective cohort study Location(s): USA Study period: 2011 to 2013 Aim of the study: to examine perioperative outcomes of patients undergoing complex EVAR, focusing on differences with complex OSR and standard infrarenal EVAR. Note: data on complex EVAR is not considered in this review
Participants	Sample size: complex EVAR group, n= 411; OSR group, n=395 Inclusion criteria: people who underwent EVAR or OSR for unruptured juxtarenal, pararenal suprarenal (proximal extent) AAAs were included. All aneurysms treated with fenestrated endografts were also included. Exclusion criteria: people with thoracoabdominal aneurysms and ruptured AAA were excluded. People who underwent OSR with infrarenal aortic clamping were also excluded. Baseline characteristics: Mean age (SD): complex EVAR group, 74.9 (8.1) years; OSR group, 72.2 (8.3) years Gender: complex EVAR group, 77.6% male; OSR group, 66.8% male Mean aneurysm diameter: not reported Hypertension: complex EVAR group, 83.5%; OSR group, 85.1% Diabetes: complex EVAR group, 15.3%; OSR group, 11.1% COPD: complex EVAR group, 19.5%; OSR group, 23.5% Heart failure: complex EVAR group, 4.1%; OSR group, 2.0% Renal insufficiency: complex EVAR group, 19.7%; OSR group, 15.1%
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). Patients who underwent repair procedures for complex AAAs were identified using diagnostic codes and procedure codes. Analysis: Multivariate logistic regression was performed to assess independent risks associated with treatment approaches. Baseline characteristics were univariately tested, and predictors with a p value < 0.01 were added to the regression model.

Full citation	**Ultee KHJ, Zettervall SL, Soden PA, et al. (2017) Perioperative outcome of endovascular repair for complex abdominal aortic aneurysms. J Vasc Surg. 65(6):1567-1575.**
Intervention	Complex EVARsc
Comparator	OSR
Outcomes	30-day mortality and adverse events
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Low risk – cohorts were drawn from the same time period. 1.2. Were cohorts from the same place? Low risk – all participants were identified using the same national registry. 1.3. Is the definition of AAA the same across cohorts? Low risk – the definition of AAA was consistent across treatment arms. **Confounding** 2.1. Does study control appropriately for demographics? Low risk – study controlled for age and gender. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk – multiple comorbidities were controlled for. These varied according to outcome measure assessed. 2.3. Does study control appropriately for AAA characteristics? High risk – AAA characteristics were not controlled for. 2.4. Could any adjustment variables have been affected by the intervention? Low risk – no post-intervention variables that could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – data were collected from a detailed surgical registry. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data were collected from a detailed surgical registry. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no details about checks for model specification/fit were reported. 4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? High risk – there is no indication that missing data were accounted for in the analyses. 4.3. Have different methods been compared within the study? High risk – different methods were not compared. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? N/A 5.3. Has balancing of the covariates been demonstrated? N/A

Full citation	**Ultee KHJ, Zettervall SL, Soden PA, et al. (2017) Perioperative outcome of endovascular repair for complex abdominal aortic aneurysms. J Vasc Surg. 65(6):1567-1575.**
	Analysis – simple multivariable models 6.1 Is sample size adequate relative to number of covariates considered? High risk - number of variables considered is not clearly reported, but is very likely to be <1/10 of the number of events. 6.2 Were interactions between treatment and other covariates considered? High risk – no interactions were considered. **Overall risk of bias:** High risk **Directness:** directly applicable

Full citation	**Varkevisser RR, O'Donnell TF, Swerdlow NJ, Liang P, Li C, Ultee KH, Pothof AB, De Guerre LE, Verhagen HJ, Schermerhorn ML. Fenestrated endovascular aneurysm repair is associated with lower perioperative morbidity and mortality compared with open repair for complex abdominal aortic aneurysms. Journal of vascular surgery. 2018 Dec 12.**
Study details	Study design: retrospective cohort Location(s): USA Study period: 2012 to 2016 Aim of the study: to compare perioperative outcomes using FEVAR with open complex AAA repair and infrarenal EVAR in a nationwide multicenter registry.
Participants	Sample size: FEVAR group, n=220; OSR group, n=181 Inclusion criteria: all patients undergoing EVAR or open complex AAA repairs within the targeted NSQIP registry, using Current Procedural Terminology (CPT) codes; FEVAR cases identified using a code for the Zenith fenestrated device Exclusion criteria: nonelective repairs, thoracoabdominal or thoracic aneurysms, EVAR devices for infrarenal repair that were used <100 times, infrarenal EVARs with a concurrent CPT code for visceral vessel repair or open repair were excluded. Baseline characteristics: Median age (IQR): FEVAR group, 75 (69.5-81) years; OSR group, 72 (67–77) years Gender: FEVAR group, 82.3% male; OSR group, 77.2% male Median aneurysm diameter (IQR): FEVAR group, 5.6cm (5.3–6.0cm); OSR group, 5.8cm (5.5–6.5cm) Hypertension: FEVAR group, 80%; OSR group, 80.1% Diabetes (insulin dependent): FEVAR group, 2.7%; OSR group, 2.2% COPD: FEVAR group, 22.3%; OSR group, 18.8% Heart failure: FEVAR group, 3.2%; OSR group, 0.6%

Full citation	**Varkevisser RR, O'Donnell TF, Swerdlow NJ, Liang P, Li C, Ultee KH, Pothof AB, De Guerre LE, Verhagen HJ, Schermerhorn ML. Fenestrated endovascular aneurysm repair is associated with lower perioperative morbidity and mortality compared with open repair for complex abdominal aortic aneurysms. Journal of vascular surgery. 2018 Dec 12.**
Methods	Data collection: data were extracted from a detailed national surgical registry (the National Surgical Quality Improvement Program; NSQIP). The NSQIP database requires hospitals to provide complete 30-day follow-up on at least 95% of patients. Analysis: inverse probability weighting using propensity scores based on logistic regression with a priori selection of variables (note >20 covariates in dataset with only around 200 events)
Intervention	FEVAR
Comparator	OSR
Outcomes	30-day mortality, cardiac arrest, renal failure, and respiratory failure
Study Appraisal using NICE's bespoke risk of bias assessment tool	**Selection** 1.1. Were cohorts from the same time period? Moderate risk – recruitment over ≥5 yrs, unclear that year of operation was controlled for in analysis 1.2. Were cohorts from the same place? Low risk – cohorts were derived from the same national surgical registry. 1.3. Is the definition of AAA the same across cohorts? High risk – FEVAR cases identified by device type; OSR cases identified using a code that includes all complex aneurysms (of which some would be unlikely to be amenable to FEVAR) – note that >30% of FEVAR group had AAAs classified as infrarenal, whereas OSR group had none, and FEVAR group was <20% suprarenal and OSR group >40%. **Confounding** 2.1. Does study control appropriately for demographics? Low risk. 2.2. Does study control appropriately for comorbidity and/or fitness? Low risk. 2.3. Does study control appropriately for AAA characteristics? Moderate risk – AAA diameter controlled for but not extent. 2.4. Could any adjustment variables have been affected by the intervention? Low risk - no post-intervention variables which could mediate the treatment effect were controlled for. **Data collection** 3.1. Is method of data collection likely to have identified suitable participants accurately? Moderate risk – participants identified a detailed surgical registry with procedure codes specified. 3.2. Is method of data collection likely to record perioperative outcomes accurately? Low risk – data obtained from a detailed surgical registry with procedure codes specified. 3.3. Is method of data collection likely to record long-term outcomes accurately? N/A – no long-term outcomes were assessed. **Analysis – general** 4.1. Were any checks conducted on model specification and/or fit? High risk – no checks appear to have been conducted on model specification/fit.

Full citation	**Varkevisser RR, O'Donnell TF, Swerdlow NJ, Liang P, Li C, Ultee KH, Pothof AB, De Guerre LE, Verhagen HJ, Schermerhorn ML. Fenestrated endovascular aneurysm repair is associated with lower perioperative morbidity and mortality compared with open repair for complex abdominal aortic aneurysms. Journal of vascular surgery. 2018 Dec 12.**
	4.2. Are missing outcome data and covariates reported and, if necessary, adjusted for? Low risk – authors highlight that missing data were equally distributed between repair modalities. 4.3. Have different methods been compared within the study? Moderate risk – different methods compared but all rely on the same assumption about selection. **Analysis – matching** 5.1. Is the matching algorithm reported and reasonable? N/A 5.2. Was overlap / common support appropriately assessed? Low risk – statement that the distribution of propensity scores in the treated and untreated groups were plotted. Trimming of extreme weights performed as a sensitivity analysis. 5.3. Has balancing of the covariates been demonstrated? Low risk – stated that, after weighting, standardised differences were all ≤10% (although no detail provided) **Analysis – simple multivariable models** 6.1 Is sample size adequate relative to number of covariates considered? N/A 6.2 Were interactions between treatment and other covariates considered? N/A **Overall risk of bias:** High risk **Directness:** directly applicable

Appendix F GRADE tables

F.1 Exclusively or predominantly infrarenal AAA

No. of studies	Quality assessment				Summary of findings					Quality
	Risk of bias	Indirectness	Inconsistency	Imprecision	No. of patients		Relative effect estimate (95% CI)	Absolute expectation		
					EVAR	OSR		OSR	EVAR (95% CI)	
Perioperative mortality – relative effects less than 1 favour EVAR										
21[a]	V serious[b]	Not serious	V serious[c]	No serious	219,938	166,528	OR = 0.31 (0.27 to 0.36)	29 per 1,000[d]	20 fewer per 1,000 (18 to 21 fewer)	VERY LOW[e]
Duration of procedure (minutes) – relative effects less than 0 favour EVAR										
1[f]	V serious[b]	Not serious	NA	No serious	157	157	MD = 86.6 shorter (71.9 to 101.3 shorter)	215 mins[g]	128 mins (114 to 143 mins)	VERY LOW[e]
Perioperative complications (all) – relative effects less than 1 favour EVAR										
6[h]	V serious[b]	Not serious	Not serious[i]	No serious	124,923	85,740	OR = 0.35 (0.26 to 0.48)	357 per 1,000[j]	194 fewer per 1,000 (147 to 231 fewer)	VERY LOW[e]
Perioperative complications (cardiovascular) – relative effects less than 1 favour EVAR										
7[k]	V serious[b]	Not serious	Not serious[i]	No serious	138,692	124,481	OR = 0.35 (0.27 to 0.45)	88 per 1,000[j]	55 fewer per 1,000 (46 to 63 fewer)	VERY LOW[e]
Perioperative complications (respiratory) – relative effects less than 1 favour EVAR										
5[l]	V serious[b]	Not serious	Not serious[i]	No serious	134,633	120,345	OR = 0.20 (0.14 to 0.29)	175 per 1,000[j]	134 fewer per 1,000 (117 to 146 fewer)	VERY LOW[e]
Perioperative complications (renal) – relative effects less than 1 favour EVAR										
6[m]	V serious[b]	Not serious	Not serious[i]	No serious	134,724	120,436	OR = 0.29 (0.22 to 0.38)	81 per 1,000[j]	55 fewer per 1,000 (47 to 62 fewer)	VERY LOW[e]
Length of critical care stay (days) – relative effects less than 0 favour EVAR										
2[n]	V serious[b]	Not serious	Not serious	No serious	1,446	6,131	MD = 1.61 shorter (1.38 to 1.84 shorter)	3.4 days[d]	1.8 days (1.5 to 2.0 days)	VERY LOW[e]

No. of studies	Quality assessment				Summary of findings					Quality
	Risk of bias	Indirectness	Inconsistency	Imprecision	No. of patients		Relative effect estimate (95% CI)	Absolute expectation		
					EVAR	OSR		OSR	EVAR (95% CI)	
Total length of hospital stay (days) – relative effects less than 0 favour EVAR										
4[o]	V serious[b]	Not serious	Not serious[i]	No serious	132,337	121,319	MD = 5.67 shorter (5.11 to 6.22 shorter)	10.5 days[d]	4.8 days (4.3 to 5.4 days)	VERY LOW[e]
Discharge to location other than home – relative effects less than 1 favour EVAR										
3[p]	V serious[b]	Not serious	V serious[c]	No serious	131,779	120,761	OR = 0.31 (0.25 to 0.38)	175 per 1,000[q]	113 fewer per 1,000 (100 to 125 fewer)	VERY LOW[e]
Post-perioperative survival – relative effects less than 1 favour EVAR										
10[r]	V serious[b]	Not serious	V serious[c]	No serious	55,523	54,104	HR = 1.24 (1.13 to 1.35)	441 surviving per 1,000 at 10 years[s]	79 fewer per 1,000 (45 to 110 fewer)	VERY LOW
Reinterventions (all or unspecified) – relative effects less than 1 favour EVAR										
8[t]	V serious[b]	Not serious	V serious[c]	No serious	59,746	62,408	HR = 1.54 (1.21 to 1.96)	206 per 1,000 at 8 years[u]	93 more per 1,000 (38 to 158 more)	VERY LOW
Reinterventions (vascular) – relative effects less than 1 favour EVAR										
3[v]	V serious[b]	Not serious	V serious[c]	No serious	43,190	47,875	HR = 2.90 (1.49 to 5.65)	37 per 1,000 at 8 years[u]	67 more per 1,000 (18 to 155 more)	VERY LOW[e]
Reinterventions (non-vascular) – relative effects less than 1 favour EVAR										
1[w]	Serious[x]	Not serious	N/A	No serious	39,966	39,966	HR = 0.38 (0.36 to 0.40)	177 per 1,000 at 8 years[u]	106 fewer per 1,000 (102 to 109 fewer)	LOW[y]

[a] *Chadi et al. (2012), Choke et al. (2012), Hicks et al. (2015), Hua et al. (2005), Huang et al. (2015), Jetty et al. (2010), Johnson et al. (2006), Karthikesalingam et al. (2016) [UK], Karthikesalingam et al. (2016) [USA], Lee et al. (2004), Liang et al. (2018), Lo et al. (2013), Malas et al. (2014), Mark et al. (2013), Quintana et al. (2019), Salata et al. (2019), Schermerhorn et al. (2015) [2001–2004 cohort], Schermerhorn et al. (2015) [2005–2008 cohort], Sugimoto et al. (2017), Symonides et al. (2018), Tarbunou et al. (2019)*

[b] *Majority of studies at high risk of bias*

[c] *I^2 > 66.7%*

[d] *Source: UK National Vascular Registry (2017)*

[e] *Eligible for uprating, owing to large effect size, but cancelled out by multiple other reasons for downgrading*

[f] *Sugimoto et al. (2017)*

[l] *Huang et al. (2015), Liang et al. (2018), Nguyen et al. (2013), Schermerhorn et al. (2015), Schwarze et al. (2009)*[m] *Huang et al. (2015), Liang et al. (2018), Nguyen et al. (2013), Schermerhorn et al. (2015), Schwarze et al. (2009), Zabrocki et al. (2018)*

[n] *Huang et al. (2015), Jetty et al. (2010)*

[o] *Huang et al. (2015), Jetty et al. (2010), Schermerhorn et al. (2015), Schwarze et al. (2009)*

[p] *Jetty et al. (2010), Schermerhorn et al. (2015), Schwarze et al. (2009)*

[q] *Derived from Karthikesalingam et al. (2016) [UK cohort]*

[r] *Behrendt et al. (2017), de la Motte et al. (2013), Huang et al. (2015) [2000–2005 cohort], Huang et al. (2015) [2006–2011 cohort], Mark et al. (2013), Salata et al. (2019), Schermerhorn et al. (2015) [2001–2004 cohort], Schermerhorn et al. (2015) [2005–2008 cohort], Sugimoto et al. (2017), Symonides et al. (2018)*

No. of studies	Quality assessment				Summary of findings					Quality
	Risk of bias	Indirectness	Inconsistency	Imprecision	No. of patients		Relative effect estimate (95% CI)	Absolute expectation		
					EVAR	OSR		OSR	EVAR (95% CI)	

[g] *Source: EVAR-1 (Brown et al., 2012)*
[h] *Hua et al. (2005), Huang et al. (2015), Johnson et al. (2006), Locham et al. (2018b), Raval et al. (2012), Schwarze et al. (2009)*
[i] *High I^2, but all estimates agree there is a large effect; hence, no downgrading because statistical inconsistency is not important for decision uncertainty*
[j] *Source: Schwarze et al. (2009)*
7[k] *Feringa et al. (2007), Huang et al. (2015), Liang et al. (2018), Nguyen et al. (2013), Salata et al. (2019), Schermerhorn et al. (2015), Schwarze et al. (2009)*
[s] *Derived from EVAR-1 – 10-year survival conditional on surviving 30 postoperative days*
8[t] *Chang et al. (2015), Huang et al. (2015), Jetty et al. (2010), Liang et al. (2018), Salata et al. (2019), Schermerhorn et al. (2015) [2001–2004 cohort], Schermerhorn et al. (2015) [2005–2008 cohort], Sugimoto et al. (2017)*
[u] *Source: Schermerhorn et al. (2015)*
[v] *Jetty et al. (2010), Schermerhorn et al. (2015), Symonides et al. (2018)*
[w] *Schermerhorn et al. (2015)*
[x] *Single study at moderate risk of bias*
[y] *Uprated for large effect size*

F.2 Complex AAA

No. of studies	Quality assessment				Summary of findings					Quality
	Risk of bias	Indirectness	Inconsistency	Imprecision	No. of patients		Relative effect estimate (95% CI)	Absolute expectation (95% CI)		
					EVAR	OSR		OSR	EVAR	
Perioperative mortality – relative effects less than 1 favour EVAR0										
6[a]	V serious[b]	Not serious	Serious[c]	V serious[d]	980	3,383	OR = 0.90 (0.41 to 1.98)	4 more per 1,000 (17 fewer to 46 more)	35 per 1,000[e,f]	VERY LOW
Duration of procedure (minutes) – relative effects less than 0 favour EVAR										
2[g]	V serious[b]	Not serious	Not serious[h]	Not serious	365	365	MD = 107.7 shorter (23.9 to 191.5 shorter)	279 mins[i]	171 mins (88 to 255 mins)	VERY LOW[j]
Perioperative complications (all) – relative effects less than 1 favour EVAR										
4[k]	V serious[b]	Not serious	V serious[l]	V serious[d]	448	614	OR = 0.64 (0.23 to 1.76)	96 more per 1,000 (96 fewer to 347 more)	270 per 1,000[f,m]	VERY LOW
Perioperative complications (cardiovascular) – relative effects less than 1 favour EVAR										
4[k]	V serious[b]	Not serious	Not serious	Not serious	448	614	OR = 0.34 (0.21 to 0.55)	138 more per 1,000 (64 to 233 more)	92 per 1,000[f,f,m]	VERY LOW[j]
Perioperative complications (respiratory) – relative effects less than 1 favour EVAR										
4[k]	V serious[b]	Not serious	Serious[c]	V serious[d]	448	614	OR = 0.50 (0.17 to 1.47)	47 more per 1,000 (16 fewer to 192 more)	52 per 1,000[f,f,m]	VERY LOW
Perioperative complications (renal) – relative effects less than 1 favour EVAR										
4[k]	V serious[b]	Not serious	V serious[l]	V serious[d]	448	614	OR = 0.54 (0.21 to 1.40)	32 more per 1,000 (11 fewer to 126 more)	40 per 1,000[f,f,m]	VERY LOW
Length of critical care stay (days) – relative effects less than 0 favour EVAR										
2[g]	V serious[b]	Not serious	V serious[c]	V serious[d]	365	365	MD = 1.32 shorter (3.9 shorter to 1.3 longer)	3.1 days (0.5 to 5.7 days)	1.8 days[e,f]	VERY LOW
Total length of hospital stay (days) – relative effects less than 0 favour EVAR										
1[n]	V serious[b]	Not serious	N/A	Not serious	263	263	MD = 5.65 shorter (5.2 to 6.1 shorter)	14.5 days (14.0 to 14.9 days)	8.8 days[e,f]	VERY LOW[j]
Discharge to location other than home – relative effects less than 1 favour EVAR										
1[n]	V serious[b]	Not serious	N/A	Not serious	263	263	OR = 0.23 (0.14 to 0.39)	125 per 1,000[o]	93 fewer per 1,000 (72 to 105 fewer)	VERY LOW[j]

No. of studies	Quality assessment				Summary of findings					Quality
	Risk of bias	Indirectness	Inconsistency	Imprecision	No. of patients		Relative effect estimate (95% CI)	Absolute expectation (95% CI)		
					EVAR	OSR		OSR	EVAR	
Post-perioperative survival – relative effects less than 1 favour EVAR										
2[p]	V serious[b]	Not serious	N/A	Not serious	143	204	HR = 1.94 (1.15 to 3.27)	441 per 1,000 surviving at 10 years[q]	237 fewer per 1,000 (51 to 372 fewer)	VERY LOW
Reinterventions (all or unspecified) – relative effects less than 1 favour EVAR										
1[r]	V serious[b]	Not serious	N/A	Serious[s]	102	102	HR = 3.02 (0.96 to 9.51)	206 per 1,000 at 8 years[t]	296 more per 1,000 (7 fewer to 682 more)	VERY LOW

[a] *Fiorucci et al. (2019), Michel et al. (2015), Orr et al. (2017), Raux et al. (2014), Tinelli et al. (2018), Tsilimparis et al. (2013)*
[b] *Majority of studies at high risk of bias*
[c] *33.3% < I^2 < 66.7%*
[d] *95% confidence interval encompasses meaningful benefit for either EVAR or OSR*
[e] *Source: National Vascular Registry (2017)*
[f] *EVAR preferred as baseline option as more reliable data available*
[g] *Orr et al. (2017), Tinelli et al. (2018)*
[h] *High I^2, but all estimates agree there is a large effect; hence, statistical inconsistency is not important for decision uncertainty*
[i] *Weighted average of OSR arms in included studies*
[j] *Eligible for uprating, owing to large effect size, but cancelled out by reasons for downgrading*
[k] *Fiorucci et al. (2019), Orr et al. (2017), Raux et al. (2014), Tinelli et al. (2018)*
[l] *I^2 > 66.7%*
[f,m] *Source: Roy et al. (2017)*
[n] *Orr et al. (2017)*
[o] *Derived from Karthikesalingam et al. (2016) [UK cohort] (in absence of long-term data specific to complex AAA)*
[p] *Fiorucci et al. (2019), Tinelli et al. (2018)*
[q] *Source: EVAR-1 – 10-year survival conditional on surviving 30 postoperative days (in absence of long-term data specific to complex AAA)*
[r] *Tinelli et al. (2018)*
[s] *95% confidence interval encompasses meaningful harm for EVAR and no difference*
[t] *Source: Schermerhorn et al. (2015) (in absence of long-term data specific to complex AAA)*