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