Long-term active surveillance of implantable medical devices: an analysis of factors determining whether current registries are adequate to expose safety and efficacy problems

Samprit Banerjee,[1] Bruce Campbell,[2] Josh Rising,[3] Allan Coukell,[3] Art Sedrakyan[1]

¹Healthcare Policy and Research, Weill Cornell Medical College, New York City, New York, USA
²University of Exeter Medical School, Exeter, UK
³Health Care Programs, The Pew Charitable Trusts, Washington, DC, USA

Correspondence to
Dr Samprit Banerjee;
sab2028@med.cornell.edu

INTRODUCTION

Ensuring the long-term safety and effectiveness of medical devices is critical to public health. Collecting outcomes through registries has exposed some high-profile failures of implanted devices, such as metal-on-metal hip prostheses.[1] Well-established registries exist in many countries and more are being developed. Recently, the US Food and Drug Administration (FDA) began supporting registry-based efforts to expand the current infrastructure for recording the real-world performance of medical devices.[2–5] However, the utility of registry-based active surveillance for detecting safety signals and efficacy concerns is by no means certain. Do they collect the key outcomes most likely to identify problems? Do they accrue sufficient numbers of patients to detect underperforming devices? Do they receive and review patient outcomes at appropriate intervals after implantation of devices? They need to produce relevant information in a timely way to help regulators and clinicians recognize devices with safety problems quickly. The data they provide can also help device manufacturers develop better next-generation products.

In this Analysis, we focus on four high-risk and widely used implanted medical devices: total hip replacement (THR) and total knee replacement (TKR) devices, stent grafts for endovascular aneurysm repair (EVAR), and surgical mesh implanted for pelvic organ prolapse (POP). We review evidence on device performance and analyze the likely numbers of patients whose data need to be recorded to detect a device which is performing significantly worse than expected. We review the current capacity of registries' active surveillance and the supporting infrastructure to conduct these analyses.

LONG-TERM ACTIVE DEVICE SURVEILLANCE AND THE ROLE OF REGISTRIES

Many questions about the long-term safety and effectiveness of medical devices remain unanswered at the time of a device's approval by a regulatory agency, because of the time required to gather sufficient data (especially data on long-term performance) and the high costs associated with data collection. An agreed plan for real-world long-term active surveillance can reduce preapproval data collection burden and give patients earlier access to new and safe technologies. The balance between preapproval and postapproval data collection required by regulatory agencies (eg, FDA) is gradually shifting.[6] Traditionally, long-term active surveillance of device safety has relied on adverse event reports that physicians, healthcare institutions, manufacturers, and patients submit to the FDA and to regulators in other countries, and on regulator-required postapproval studies that manufacturers conduct. However, previous research has demonstrated the shortcomings of postapproval studies[7] and adverse event reports: these include under-reporting and lack of denominator data to conduct population-level adverse event estimates.[7 8]

Registries offer a promising alternative or adjunct: they have exposed serious device-related problems in the past and recent reports have highlighted their potential.[9–11] Linkage of registries and creation of coordinated registry networks (CRN) are likely to become an important method for tracking patient outcomes and assessing the performance of devices.

FOCUS ON PARTICULAR HIGH-RISK IMPLANTED DEVICES

We chose four types of implanted devices as test cases, which are in frequent use and for which registries and CRNs are being developed, to track their performance:

▶ *THR and TKR—as two separate device categories:* Total joint replacement (TJR), both hip and knee, is the fastest growing elective device-based surgery worldwide. In the USA, over 400 000 hip and 610 000 knee replacements are performed annually, with combined numbers projected to reach 6 million by 2030.[12–14]

▶ *EVAR:* Abdominal aortic aneurysm (AAA) repair occurs more frequently in the USA (278 921) compared with the UK (29 300) during 2005–2012 yet aneurysm-related death is three times more likely in the UK compared with the USA.[15] Aneurysms were traditionally treated by inserting a synthetic graft in a major open surgery, but now the most common approach is via a significantly less invasive endovascular procedure with a stent graft: this is EVAR. In 2012, over 30 000 EVAR procedures were performed in the USA.[16]

▶ *Surgical mesh for POP:* Surgical mesh is often used in POP repair. In 2010, an estimated 300 000 POP repairs were completed in the USA.[17] There are current concerns about the safety of these procedures: 1 in 11 women experienced problems with vaginal mesh implants in the UK according to National Health Service (NHS) data on 92 000 women from Hospital Episodes Statistics.[18]

For each type of device, we selected three primary outcome measures to assess the performance of each device. The choice of outcomes was based on the most commonly collected key data items, related to safety, effectiveness and/or patient-reported outcomes and are summarized in table 1.

DO REGISTRIES ENROLL SUFFICIENT NUMBER OF PATIENTS?

The number needed to follow (NNF) is the number of patients a registry needs for a specific brand/type/class of device, to detect statistically significantly (p<0.05) worse performance compared with a prespecified threshold.

Table 1 Summary and background for three measures chosen for each of the four devices

Measure chosen	Description
Total hip/knee replacement	
Harris Hip Score/Knee Society Score	▶ Harris Hip Score (HHS) is a clinician-administered survey based on a 100-point scale to assess hip pain and function after surgery. ▶ Knee Society Score (KSS) is a validated 200-point scale to assess knee condition after surgical intervention.
QoL	▶ Patient-reported physical, social and mental functioning using the Short Form 12 (SF-12) and Short Form 36 (SF-36).
Revision	▶ A reoperation due to mechanical or biological failure. ▶ Used as safety indicator registries.[36–39] ▶ Widely accepted as a patient-centered outcome.
Endovascular aneurysm repair (EVAR)	
Endoleak	▶ Continued perfusion and pressurization. ▶ Considered as an effectiveness outcome.[40 41] ▶ 20%–25% of EVAR procedures are complicated by endoleaks.[42 43] ▶ Excluding type II, the cumulative endoleak rate is 5.67% at 2 years.[26]
30-day all-cause mortality	▶ Most frequently used primary outcome in major clinical trials such as EVAR 1, EVAR 2 and ACE.[44–47] ▶ Mortality varies from 0.5% to 3.6%[48] based on different studies and different EVAR grafts.[40 41 49]
Secondary vascular intervention	▶ Major indications: endoleak, graft migration, kinking of limb grafts, stenosis and occlusions, and landing site enlargement.[50] ▶ Reintervention rates are 17% for intermediate follow-up (up to 4 years) and 23% for a longer term follow-up (beyond 4 years) with a median of 6 years.[51]
Surgical mesh for POP	
QoL	▶ King's College Pelvic Organ Prolapse Quality of Life (P-QoL)[27–29] is the most commonly used validated questionnaire with nine domains.
Total reoperation rate	▶ Reoperation due to postoperative complications and prolapse recurrence. ▶ Total 1-year reoperation rate is estimated to be 3%–4%.[30–32] ▶ Estimates for longer term (1.5–5 years) vary between 8.5% and 13%.[52 53]
Mesh erosion	▶ One of the most frequently experienced safety issues related to vaginal mesh is the risk of vaginal-urethral erosion/extrusion. ▶ Factors contributing to erosion include operative technique, implant size, and the specific properties of the material (size, stiffness, elasticity and basic tissue compatibility, and so on).[54] ▶ 110 studies based on safety warning, issued by FDA in 2011: approximately 10% of women undergoing POP repair with mesh experienced mesh erosion within 1 year.[17 55]

FDA, Food and Drug Administration; POP, pelvic organ prolapse.

Appropriate choice of the threshold and of follow-up time for each device and outcome combination, used in our analysis, is described in more detail below. Operationally, the threshold for binary outcomes was a gold standard event rate and for continuous outcomes was a mean baseline or preoperative score. Then, the probability of detecting a significant departure from this gold standard or statistical power was calculated. For binary outcomes, this was a one-side test of proportions using the formula:

$$Power = 1 - \beta = 1 - \Phi\left(\frac{\sqrt{n}(p_0 - p_1)}{\sqrt{p_1(1-p_1)}} + z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}}\right)$$

where n = NNF, p_1 is the device performance rate and p_0 is the gold standard performance rate, $\Phi(.)$ is the standard normal cumulative distribution function, $z_a = a^{th}$ quantile of a standard normal distribution, α = significance level and $1 - \beta$ = power. Power for a generic binary outcome (table 2) was calculated for 1.5 times increased odds of underperforming compared with a gold standard performance rate per recommendations.[19] For continuous outcomes, the power was computed for a one-sided alternative in a generalized estimating equations framework with a quasis-core-based test statistic[20] assuming only one group, a linear effect of time and a range (0.05–0.50) of intraclass correlation coefficient, a parameter that accounts for correlations among repeated measures. Power for a generic continuous outcome (table 3) was calculated by considering small to moderate departures (0.3–0.5 SD) in the scale of Cohen's d which is recommended for quality and patient-reported outcomes.[21 22]

Of note, these computations do not reflect missed follow-up measures, the rates of which vary widely with respect to procedures and outcomes. Our recommendation is to inflate NNF by appropriate context-dependent expected missing rates in the planning stage and in the analysis stage using statistical methods for ascertainment bias correction due to missing data such as robust or doubly robust methods (eg, inverse probability weighting or augmented inverse probability weighting) or use multiple imputation.[23] Another important caveat is that our calculations do not account for subgroup analysis requirements (eg, hip implants performance in men and women separately) and the NNF needs to be adjusted if such subgroup comparisons are planned in advance.

AT WHAT TIME POINTS SHOULD REGISTRIES EVALUATE DEVICE PERFORMANCE?

The number of times a device is evaluated is conceptualized in two ways—'one look', when performance evaluation is performed at the end of a specified time period (eg, 30 days for 30-day mortality); and 'multiple looks' when performance evaluation is performed repeatedly at several time intervals (eg, 1, 5 and 10 years). While the choice of time intervals is guided by narrative review of literature, they are presented for illustrative purposes only and our methodology can be adopted for any choice of time point. Critically, however, the overall type I error

Table 2 Power for a generic binary outcome.

Year	2			5			10		
Yearly rate (%)	0.50	0.75	1.00	0.50	0.75	1.00	0.50	0.75	1.00
Cumulative rate (%)	1.00	1.50	2.00	2.50	3.75	5.00	5.00	7.50	10.00
n=500	0.201	0.261	0.317	0.369	0.486	0.583	0.583	0.726	0.8197
n=1000	0.322	0.428	0.521	0.601	0.753	0.849	0.849	0.945	0.9797
n=1500	0.431	0.568	0.677	0.761	0.891	0.952	0.952	0.991	0.9982
n=2000	0.528	0.68	0.788	0.862	0.955	0.986	0.986	0.999	0.9999
n=2500	0.612	0.767	0.865	0.923	0.982	0.996	0.996	1	1
n=3000	0.684	0.833	0.916	0.958	0.993	0.999	0.999	1	1
n=3500	0.744	0.882	0.948	0.978	0.998	1	1	1	1
n=4000	0.795	0.917	0.969	0.988	0.999	1	1	1	1

Rows correspond to various sample size for each device brand. Columns correspond to various yearly event rates of a generic binary event, resulting in cumulate rate at years 2, 5 and 10. Each cell denotes the power to detect an OR of 1.5. Cells colored red, white, yellow and green denote <50%, 50%–70%, 70%–90% and >90% power, respectively.

Table 3 Sample size table for generic continuous outcome.

			One look	Long-term follow-up with multiple looks		
				1 year	5 years	10 years
ICC	Power	Cohen's *d*	n	n	n	n
0.05	0.8	0.3	131	187	167	151
0.25	0.8	0.3	104	147	132	119
0.5	0.8	0.3	69	98	88	80
0.05	0.9	0.3	181	246	220	198
0.25	0.9	0.3	143	194	173	157
0.5	0.9	0.3	96	130	116	105
0.05	0.8	0.5	47	68	60	55
0.25	0.8	0.5	38	53	48	43
0.5	0.8	0.5	25	36	32	29
0.05	0.9	0.5	66	89	79	72
0.25	0.9	0.5	52	70	63	57
0.5	0.9	0.5	35	47	42	38

Rows correspond to various levels of power, intraclass correlation coefficient (ICC) and effect size (measured by Cohen's d). Columns provide sample size for one look or assessment and continual monitoring with multiple looks or assessments at 1, 5 and 10 years. The type I error for 'one look' is controlled at 5% and for 'multiple looks' is jointly controlled at 5% using Bonferroni's method.

or alpha, in every situation, should be controlled at 5% and adjusted for multiple looks, at certain time intervals, using a simple and conservative method such as Bonferroni's method (as we use here) or other alpha-spending methods such as O'Brien-Fleming, Pocock or Lan-DeMets.

WHAT INDICATES THAT A DEVICE IS NOT PERFORMING AS EXPECTED?

The expected performance of devices at each follow-up time point was guided by a narrative review of literature (see online supplementary tables 1-4). When available (eg, endoleak endpoint after AAA) we gave priority to estimated thresholds from studies that conducted meta-analysis or meta-regression. In other instances, our estimated performance thresholds are conservative and a formal systematic review or meta-analysis is unlikely to change the results substantially.

Hip and knee replacement

Key outcomes were scores of physical function, quality of life (QoL) measures and revision rates. Our assessment of Harris Hip Score (HHS) showed that following a cohort of 341 patients will provide >90% power to detect an underperforming device with a clinically meaningful change of 4.5 HHS points (or d=0.3) after 5 and 10 years (table 4). Similarly, our assessment for Knee Society Score showed that a cohort of 341 patients is required to provide >90% power to detect an outlier device with a clinically meaningful change of 4.5 points (or d=0.3) at the end of 5 and 10 years.

Postoperative Short Form 12 (SF-12) is the most commonly used measure of general QoL. Using SF-12

we found that a cohort of 181 (66) patients is needed to identify a small (moderate) change of 1.4 (2.4) points for THR or 1.6 (2.7) points for TKR, for 'one look' at the end of 1 year with 90% statistical power.

Based on National Institute for Health and Care Excellence guidelines,[24] rates of revision surgery for joint replacements should be <5% at 10 years as a metric of good performance and is commonly assessed at years 2, 5, and 10.[25] Following up a cohort of 2500 patients for 5 years would have >90% power to detect 1.5 times increased odds of revision compared with the annualized revision rate of 0.5%.

Endovascular aneurysm repair

Three key outcomes are 30-day mortality, occurrence of endoleaks, and the need for secondary interventions (for endoleaks and other complications associated with the stent graft). Thirty-day all-cause mortality varies between 0.5% and 3.6% (see online supplementary table 3) with a best estimate around 2%. Following up a cohort of 2200 patients would provide >90% power to detect 1.5 times increased odds of 30-day all-cause mortality compared with the performance goal of 2% mortality.

A recent meta-regression[26] estimated the cumulative rate of endoleak (excluding type II) occurrence to be 5.67% at 2 years or an annualized rate of 2.84%. A cohort of 1400 participants would provide >90% power to detect 1.5 times higher odds of endoleak occurrence at the end of 2 and 5 years. However, our NNF estimate is 525 to detect 1.75 times higher odds with 80% power that is similar to that estimated in Kent *et al*'s study.[26]

Secondary vascular interventions are estimated to occur at an annual rate of 4% (online supplementary table 3).

Table 4 Number needed to follow (NNF) of three measures of four devices and a power of >90%

Measure chosen	Gold standard	Acceptable change from gold standard	Time point(s) to assess at	Maximum NNF*
Total hip/knee replacement				
Harris Hip Score/Knee Society Score	Postoperative score at year 2	4 points or $d=0.3$	5 and 10 years	341
QoL	Preoperative score	THR: 1.4, 2.4 points or $d=0.3, 0.5$ TKR: 1.6, 2.7 points or $d=0.3, 0.5$	1 year	181
Revision	Annual rate of 0.5%	OR=1.5	5 and 10 years	2500
EVAR				
Endoleak	Annual rate of 2.8%	OR=1.5	2 and 5 years	1400
30-day all-cause mortality	2%	OR=1.5	30 days	2200
Secondary vascular intervention	Annual rate of 4%	OR=1.5	4 and 6 years	450
Surgical mesh for POP				
Erosion	6%	OR=1.5	1 year	800
QoL	Preoperative score	8.6, 14.4 points or $d=0.3, 0.5$	1 year	181
Total reoperation rate	3%–4%	OR=1.5	1 year	1500

The gold standard for continuous outcomes is based on a baseline measure (before surgery or 2 years after surgery) and that for binary outcomes is based on gold standard rates from literature or expert opinion.
*The maximum is over multiple scenarios of gold standards and time points.
EVAR, endovascular aneurysm repair; POP, pelvic organ prolapse; QoL, quality of life; THR, total hip replacement; TKR, total knee replacement; d, Cohen's d.

With this performance goal, the NNF is 450 for 4 and 6 years (see table 1) to have >90% power to detect 1.5 times increased odds of secondary vascular intervention.

Surgical mesh for POP

Key outcomes are QoL measures, mesh erosion and need for reoperation. Effectiveness outcomes in published reports on mesh repair are typically based on QoL scores (P-QoL) before and after the procedure.[27–29] Following up a cohort of 181 patients has >90% power to identify a clinically meaningful 8.6 point change in P-QoL for 'one look' at 1 year postoperatively.

Mesh erosion is an important safety endpoint, and a cohort of 800 women followed up for 1 year would provide >94% power to detect 1.5 times higher odds for erosion compared with an overall erosion rate of 6% (see online supplementary table 4).

The risk of reoperation following POP repair with mesh due to postoperative complication or prolapse recurrence was estimated to be 3%–4% within 1 year.[30–32] A cohort of 1500 patients would provide >90% power to detect 1.5 times increased odds of reoperation.

ARE CURRENT REGISTRIES SUFFICIENT?

Our analyses suggest that registries do not need to follow unrealistically large cohorts to identify outlier performance of types of devices (eg, metal-on-metal hip implants). We evaluated current registry infrastructure to identifying deviations from expected performance of the devices considered using three objective criteria—whether the registry contains enough patient records, captures the relevant measures, and conducts enough longitudinal follow-up—and found most registries either contain sufficient numbers of patients already or are expected to be so in future.

Specifically for hip and knee replacements, the American Joint Replacement Registry, Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement registry, the Kaiser-Permanente National Total Joint Replacement Registry, and the Michigan Arthroplasty Registry together have data on over 1 million TJRs annually but lack robust capture of functional and quality measures (eg, HHS, QoL). The US Vascular Quality Initiative (VQI), launched by the Society for Vascular Surgery in 2011, has over 350 participating centers across 46 US states and Ontario, and has collected data on over 32 200 AAA repairs and adequately captures key performance measures. The Pelvic Floor Disorders Registry, founded to support US FDA recommendations for increased monitoring of transvaginal mesh use, began collecting data in late 2015 and is currently too new to conduct robust analyses.[33]

Of note, it is unclear if the current registries are sufficiently large for each device brand (eg, DePuy metal-on-metal implant) because such details are not always recorded and future efforts to harmonize definitions need to be undertaken. In such scenarios, device classes could be studied by combining multiple brands and types (eg, compare metal-on-metal with metal-on-ceramic hip replacement devices).

WHAT ARE THE OBSTACLES TO PROGRESS?

Conducting long-term active surveillance is not easy and incomplete follow-up can limit the usefulness of registry

data. This problem could be addressed by linking registries with other data sources (eg, insurance claims data in private health systems) that collect patient outcomes over extended periods of time (eg, efforts are underway to link the VQI with claims data resources,[34] providing longer follow-up to evaluate secondary vascular intervention and possibly endoleak occurrence after device use). In the USA, the FDA has proposed a new National Evaluation System for Health Technology,[3] which will integrate registries with claims data to provide long-term follow-up. This method is better suited for clinically measured safety or effectiveness outcomes (eg, revision surgery for joint replacement) than for patient-reported outcomes (eg, QoL) which requires direct data collection (eg, telephone interviews). In the UK, linkages with routinely collected NHS statistics, national mortality data and with the Clinical Practice Research Dataset[35] are important linkages for long-term data acquisition.

Incorporation of patient-reported outcomes into long-term active device surveillance will remain a challenge, because registries, claims data, or electronic health records do not currently capture such outcomes. Registry organizers, manufacturers, and healthcare providers will all have a part to play in developing mechanisms to collect patient-reported outcomes.

THE FUTURE

Advancing technology will enable increasing amounts of useful data to be collected by device registries. Electronic health records, smartphone apps and wearable devices will enable a range of information, including patient-reported outcome measures, to be captured. Increasing capacity for data linkage offers huge potential, but this will need careful attention to its governance. International linkages will increase the power of information gathering, but will require attention to differing national data legislative requirements. Manufacturer's registries straddle national boundaries, but these need assurances of transparency and unbiased oversight to provide confidence in their data. Clear plans for the use of registries when new devices are presented for regulatory approval will help shift the balance that is required between premarket and postmarket evidence. It will allow earlier access to new products while providing assurance of a mechanism to monitor performance. The data from registries will help manufacturers to develop ever safer and more effective devices for future use.

REFERENCES

1. Sedrakyan A, Campbell B, Graves S, *et al*. Surgical registries for advancing quality and device surveillance. *Lancet* 2016;388:1358–60.
2. Food drug administration center for devices and radiological health. *Guidance for industry and food and drug administration staff: use of real-world evidence to support regulatory decision-making for medical devices*. FDA Maryland, 2017.
3. Shuren J, Califf RM. Need for a national evaluation system for health technology. *JAMA* 2016;316:1153–4.
4. Faris O, Shuren J. An FDA viewpoint on unique considerations for medical-device clinical trials. *N Engl J Med* 2017;376:1350–7.
5. Sherman RE, Anderson SA, Dal Pan GJ, *et al*. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med* 2016;375:2293–7.
6. Food, Administration D. *Balancing premarket and postmarket data collection for devices subject to premarket approval. CDRH report*, 2015.
7. Reynolds IS, Rising JP, Coukell AJ, *et al*. Assessing the safety and effectiveness of devices after US Food and drug administration approval: FDA-mandated postapproval studies. *JAMA Intern Med* 2014;174:1773–9.
8. Normand SL, Marinac-Dabic D, Sedrakyan A, *et al*. Rethinking analytical strategies for surveillance of medical devices: the case of hip arthroplasty. *Med Care* 2010;48(6 Suppl):S58–67.
9. Krucoff MW, Sedrakyan A, Normand S-LT. Bridging unmet medical device ecosystem needs with strategically coordinated registries networks. *JAMA* 2015;314:1691–2.
10. The Brookings Institution. *Strengthening patient care: building an effective national medical device surveillance system*, 2015.
11. Medical Device Epidemiology Network. *Recommendations for a national medical device evaluation system*, 2015.
12. Steiner C, Andrews R, Barrett M. *HCUP projections: Mobility/Orthopedic procedures 2011 to 2012*. U.S. Agency for Healthcare Research and Quality, 2012.
13. Prevention CfDCa. NCHS, National Hospital Discharge Survey. 2005 annual summary with detailed diagnosis and procedure data, December 2007 centers for disease control and prevention, 2014. Available: http://www.cdc.gov/nchs/about/major/hdasd/listpubs.htm
14. Kurtz S, Ong K, Lau E, *et al*. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *The Journal of bone and joint surgery American volume* 2007;89:780–5.
15. Karthikesalingam A, Vidal-Diez A, Holt PJ, *et al*. Thresholds for abdominal aortic aneurysm repair in England and the United States. *N Engl J Med* 2016;375:2051–9.
16. Quality AfHRa. HCUPnet: agency for healthcare research and quality, 2015. Available: http://hcupnet.ahrq.gov/HCUPnet.app/
17. Administration USFaD. *Urogynecologic surgical mesh: update on the safety and effectiveness of transvaginal placement for pelvic organ prolapse*, 2011.
18. News BBC. Vaginal mesh operations should be banned, says NICE, 2017. Available: http://www.bbc.com/news/health-42110076
19. Mello MM, Goodman SN, Faden RR. Ethical considerations in studying drug safety—the Institute of Medicine report. *Mass Medical Soc* 2012.
20. Liu G, Liang K-Y. Sample size calculations for studies with correlated observations. *Biometrics* 1997;53:937–47.
21. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582–92.
22. Revicki D, Hays RD, Cella D, *et al*. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.
23. Little RJ, Rubin DB. *Statistical analysis with missing data*. John Wiley & Sons, 2014.
24. Excellence NIfHaC. *Total hip replacement and resurfacing arthroplasty for end-stage arthritis of the hip: technology appraisal guidance*, 2014.

25. Singh JA, Schleck C, Harmsen S, *et al*. Clinically important improvement thresholds for Harris hip score and its ability to predict revision risk after primary total hip arthroplasty. *BMC Musculoskelet Disord* 2016;17.
26. Kent F, Ambler GK, Bosanquet DC, *et al*. The safety of device registries for endovascular abdominal aortic aneurysm repair: systematic review and meta-regression. *Eur J Vasc Endovasc Surg* 2018;55:177–83.
27. Feldner PC, Castro RA, Cipolotti LA, *et al*. Anterior vaginal wall prolapse: a randomized controlled trial of SIS graft versus traditional colporrhaphy. *Int Urogynecol J* 2010;21:1057–63.
28. Maher CF, Feiner B, DeCuyper EM, *et al*. Laparoscopic sacral colpopexy versus total vaginal mesh for vaginal vault prolapse: a randomized trial. *Am J Obstet Gynecol* 2011;204:360.e1–7.
29. Hviid U, Hviid TVF, Rudnicki M. Porcine skin collagen implants for anterior vaginal wall prolapse: a randomised prospective controlled study. *Int Urogynecol J* 2010;21:529–34.
30. Chughtai B, Mao J, Buck J, *et al*. Use and risks of surgical mesh for pelvic organ prolapse surgery in women in New York state: population based cohort study. *BMJ* 2015;350:h2685.
31. Kelly EC, Winick-Ng J, Welk B. Surgeon experience and complications of transvaginal prolapse mesh. *Obstet Gynecol* 2016;128:65–72.
32. Chughtai B, Barber MD, Mao J, *et al*. Association between the amount of vaginal mesh used with mesh erosions and repeated surgery after repairing pelvic organ prolapse and stress urinary incontinence. *JAMA surgery* 2016.
33. Bradley CS, Visco AG, Weber LeBrun EE, *et al*. The pelvic floor disorders Registry: purpose and development. *Female Pelvic Med Reconstr Surg* 2016;22:77–82.
34. Hoel AW, Faerber AE, Moore KO, *et al*. A pilot study for long-term outcome assessment after aortic aneurysm repair using Vascular Quality Initiative data matched to Medicare claims. *J Vasc Surg*. In Press 2017;66:751–9.
35. Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data resource profile: clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology* 2015;44:827–36.
36. U.S. Food and Drug Administration. Summary of Safety and Effectiveness Data (SSED)-Ceramic Total Hip System: U.S. Food and Drug Administration, 2009. Available: http://www.accessdata.fda.gov/cdrh_docs/pdf3/P030042b.pdf
37. D'Antonio J, Capello W, Manley M, *et al*. Alumina ceramic bearings for total hip arthroplasty: five-year results of a prospective randomized study. *Clinical orthopaedics and related research* 2005;436:164–71.
38. MacDonald SJ, McCalden RW, Chess DG, *et al*. Metal-on-metal versus polyethylene in hip arthroplasty: a randomized clinical trial. *Clin Orthop Relat Res* 2003;406:282–96.
39. Hooper GJ, Rothwell AG, Stringer M, *et al*. Revision following cemented and uncemented primary total hip replacement: a seven-year analysis from the New Zealand joint registry. *J Bone Joint Surg Br* 2009;91:451–8.
40. Administration USFaD. Summary of Safety and Effectiveness Data (SSED) - Talent Abdominal Stent Graft System: U.S. Food and Drug Administration, 2008. Available: http://www.accessdata.fda.gov/cdrh_docs/pdf7/p070027b.pdf
41. U.S. Food and Drug Administration. Summary of Safety and Effectiveness Data - Powerlink System: U.S. Food and Drug Administration, 2004. Available: http://www.accessdata.fda.gov/cdrh_docs/pdf4/p040002b.pdf
42. Hellinger JC. Endovascular repair of thoracic and abdominal aortic aneurysms: pre- and postprocedural imaging. *Tech Vasc Interv Radiol* 2005;8:2–15.
43. Veith FJ, Baum RA, Ohki T, *et al*. Nature and significance of endoleaks and endotension: summary of opinions expressed at an international Conference. *J Vasc Surg* 2002;35:1029–35.
44. Greenhalgh RM, Brown LC, Kwong GP. Comparison of endovascular aneurysm repair with open repair in patients with abdominal aortic aneurysm (EVAR trial 1), 30-day operative mortality results: randomised controlled trial. *Lancet* 2004;364:843–8.
45. Endovascular aneurysm repair and outcome in patients unfit for open repair of abdominal aortic aneurysm (EVAR trial 2): randomised controlled trial. *The Lancet* 2005;365:2187–92.
46. Lederle FA, Freischlag JA, Kyriakides TC, *et al*. Outcomes following endovascular vs open repair of abdominal aortic aneurysm: a randomized trial. *JAMA* 2009;302:1535–42.
47. Becquemin J-P. The ACE trial: a randomized comparison of open versus endovascular repair in good risk patients with abdominal aortic aneurysm. *J Vasc Surg* 2009;50:222–4. discussion 4.
48. Stather PW, Sidloff D, Dattani N, *et al*. Systematic review and meta-analysis of the early and late outcomes of open and endovascular repair of abdominal aortic aneurysm. *Br J Surg* 2013;100:863–72.
49. U.S. Food and Drug Administration. Summary of safety and probable Benefit-Ovation abdominal stent graft system: U.S. Food and drug administration, 2011. Available: http://www.accessdata.fda.gov/cdrh_docs/pdf10/h100008b.pdf
50. Becquemin J-P, Kelley L, Zubilewicz T, *et al*. Outcomes of secondary interventions after abdominal aortic aneurysm endovascular repair. *J Vasc Surg* 2004;39:298–305.
51. Paravastu SCV, Jayarajasingam R, Cottam R, *et al*. Endovascular repair of abdominal aortic aneurysm. *Cochrane Database Syst Rev* 2014;50.
52. Diwadkar GB, Barber MD, Feiner B, *et al*. Complication and reoperation rates after apical vaginal prolapse surgical repair. *Obstetrics & Gynecology* 2009;113:367–73.
53. Morling JR, McAllister DA, Agur W, *et al*. Adverse events after first, single, mesh and non-mesh surgical procedures for stress urinary incontinence and pelvic organ prolapse in Scotland, 1997-2016: a population-based cohort study. *Lancet* 2016.
54. Mistrangelo E, Mancuso S, Nadalini C, *et al*. Rising use of synthetic mesh in transvaginal pelvic reconstructive surgery: a review of the risk of vaginal erosion. *J Minim Invasive Gynecol* 2007;14:564–9.
55. Abed H, Rahn DD, Lowenstein L, *et al*. Incidence and management of graft erosion, wound granulation, and dyspareunia following vaginal prolapse repair with graft materials: a systematic review. *Int Urogynecol J* 2011;22:789–98.