



What type of clinical evidence is needed to assess medical devices?

Rosanna Tarricone^{1,2}, Paola Roberta Boscolo¹ and Patrizio Armeni¹

Affiliations: ¹CERGAS, Centre for Research on Social and Healthcare Management, Bocconi University, Milan, Italy. ²Dept of Policy Analysis and Public Management, Bocconi University, Milan, Italy.

Correspondence: Rosanna Tarricone, CERGAS, Bocconi University, Via Roentgen, 1, 20136 Milan, Italy.
E-mail: rosanna.tarricone@unibocconi.it

ABSTRACT The objective of this mini-review is to discuss the role of real-world studies as a source of clinical evidence when experimental studies, such as randomised controlled trials (RCTs), are not available. Waiting for RCT evidence when the technology is diffusing could be anti-economical, inefficient from the policy perspective and methodologically questionable.

We explain how real-world studies could provide relevant evidence to decision makers. Matching techniques are discussed as a viable solution for bias reduction.

We describe a case study concerning a cost-effectiveness analysis based on real-world data of a technology already in use: Mitraclip combined with medical therapy *versus* medical therapy alone in patients with moderate-to-severe mitral regurgitation. The CEA has encountered the scepticism of most reviewers, due not to the statistical methodology but to the fact that the study was observational and not experimental. Editors and reviewers converged in considering real-world economic evaluations premature in the absence of a RCT, even if in the meantime the technology had been implanted >30 000 times. We believe there is a need to acknowledge the importance of real-world studies, and engage the scientific community in the promotion and use of clinical evidence produced through observational studies.



@ERSpublications

Real-world data are a valid complement and/or alternative to RCTs to support policy decisions on medical devices <http://ow.ly/SHsZ300pfCB>

Introduction

Limited resources and limitless health needs require policy makers to choose amongst several treatment options and to prioritise technological innovations. Economic evaluation analysis in healthcare are consolidated practices in most countries (www.inahta.org), and serve to support the decision-making process to formulate evidence-based guidelines, and adopt or reject recommendations.

In the last few decades, the faster escalation of technologies other than drugs (*e.g.* diagnostics and medical devices), representing >10 000 patent applications filed in 2012 to the European Patent Office *versus* ~5400 applications filed for drugs and ~5300 for biotechnologies [1], convinced health analysts that economic evaluation methods and techniques, originally developed with pharmaceuticals in mind, may not always be appropriate to assess technologies with several different characteristics. Six reasons that make the assessment of medical devices more challenging than that of drugs have been identified [2]. Further work has focussed mainly on one of these six reasons, *i.e.* the source and role of clinical evidence [3]. In some cases, experimental clinical studies for medical devices (*i.e.* randomised controlled studies) are difficult, impossible or unethical. Randomising patients to invasive surgeries is always a challenge when the alternative is noninvasive treatments. For the same reason, blinding might not be possible. Randomisation

Editorial comment in *Eur Respir Rev* 2016; 25: 221–222 and 223–226.

Received: Feb 26 2016 | Accepted after revision: May 07 2016

Conflict of interest: Disclosures can be found alongside this article at err.ersjournals.com

Provenance: Submitted article, peer reviewed.

Copyright ©ERS 2016. ERR articles are open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

is not ethical when large differences have already emerged between the new device and the standard of care. This may occur because current regulation of medical devices (European directives 90/385/EEC, 93/42/EEC, 98/79/EC and 2007/47/EC) allows their use in clinical practice well before, or in the absence of, any experimental clinical study. Moreover, health outcomes of procedures with medical devices (e.g. surgical procedures and implantable medical devices) highly depend upon end-users' experience, which makes the time of assessment crucial for medical devices, as eloquently explained by Buxton's law: "it is always too early [...] until, unfortunately, it is suddenly too late" [4].

There is increasing recognition that, in the case of medical devices, experimental clinical studies might be less relevant than real-world data to making policy decisions on, for example, reimbursement and coverage [5].

Since the early diffusion of medical devices generally precedes formal health technology assessment (HTA) recommendations and, if at all, experimental evidence, decision making cannot ignore that some physicians' and healthcare organisations are already familiar with the technology. Given this framework, the questions therefore become, can we decide on coverage and reimbursement of new devices in the absence of randomised controlled trial (RCT) data? Are real-world data accumulated on the new device usable to inform decision making? If so, under which circumstances? The answer to these questions is yes, provided real-world data are robustly analysed and elaborated, because there is no other option unless we prevent end-users from using new devices until experimental evidence makes itself available, if ever.

The objective of this mini-review is to discuss the role and use of experimental and real-world evidence to support the healthcare-related decision-making process as to the allocation of scarce resources (*i.e.* in the HTA of medical devices).

Under certain conditions, real-world studies can provide relevant evidence to decision makers, even in absence of RCTs, therefore becoming not only a complementary source of evidence but also a low-cost, rapid and valuable substitute for technologies whose diffusion process has already started. To this purpose, we use a case study concerning a cost-effectiveness analysis (CEA) based on real-world data of a technology already in use. The case study is about the CEA of Mitraclip (Abbott Vascular, Menlo Park, CA, USA) *versus* optimal medical therapy (OMT) in patients with moderate-to-severe mitral regurgitation. Although some RCTs on the Mitraclip procedure are available, none published so far compared it with medical therapy alone, which resulted as the only appropriate comparator. In order to reach its objective, this mini-review is organised as follows. In the next section, we explain under what circumstances real-world studies can be helpful in addition to or instead of experimental studies. We then illustrate current available methods to control for selection bias, *i.e.* the most important weakness of real-world studies. The case of Mitraclip is used to exemplify one of our suggested methods, propensity score-based regression adjustment.

The role of real-world studies as a source of clinical evidence

The gold-standard research designs, such as systematic reviews and meta-analysis, double-blind controlled trials, and cohort prospective studies, often cannot offer prompt information to decision makers. Often, medical devices are under assessment by governments and HTA organisations while diffusing in medical practice. A common challenge for most countries is therefore finding the appropriate balance between assessing the efficacy and safety on the one hand, and allowing rapid access to patients on the other [3].

The European Forum "Relative Effectiveness" Working group has defined real-world studies as "a measure in understanding health care data collected under real life practice circumstances" [6] and at the payer level, a real-life study has been defined as "anything that is not interventional". Real-world studies are increasingly recognised as a valuable source of clinical evidence [5, 7, 8], especially when clinical trials are either not available, difficult to realise, unethical or available but not providing a clear aid for decision making, lacking a deep understanding of a drug or device implications on current management in a specific real-life setting [5, 7–10].

Clinical evidence for devices can vary notably due to differences in providers' training, and users' skills and experiences [2]. Real-world studies can monitor variations in clinical evidence over time, accounting for a learning curve, and across different users, with multicentre research designs.

From this, real-world studies are likely to provide prompt and more relevant information for decision makers than RCTs. In other words, if the design of RCTs remains the gold standard for identifying the causal relationship between the use of a technology and a variation in the clinical conditions of patients (something that a real-world study will never replace), real-world studies might be preferred when: 1) the effectiveness of a technology is likely to differ largely from its efficacy due to use-specific effects, like the presence of a learning curve that dramatically impacts patient selection and health outcomes; 2) the realisation of an experimental study does not grant the elimination of biases from selection or awareness

(lack of blinding); 3) the regulation of the technology makes its diffusion in clinical practice not binding to experimental studies, which compels decision-makers to decide upon other evidence; 4) the RCT does not provide an acceptable representation of the economic consequences of use (e.g. protocol-driven consumption of resources for implantations and follow-ups hugely differ from actual clinical practice); and 5) the technology is constantly improving. Innovation of medical devices is often incremental, which makes repeated RCTs simply unbearable.

The possible sources of real-world data (databases, surveys, patient chart reviews, pragmatic clinical trials, observational data from cohort studies, and registries prospectively collecting information and then retrospectively analysing patients' data per specific centre and given conditions) and the phases when real-world studies can support decision-making process of multiple stakeholders are different before the launch at the early development stage, when demonstrating the value, and at and after the launch to support clinical practice.

Real-world studies and selection bias: is there any chance of escaping it?

Often, real-world studies have been accused of suffering of different types of bias leading to erroneous results. One of the main reasons for selection bias in real-world studies is the allocation of patients to either the treatment or the control group that is not random, and can therefore be dependent on patients' observable baseline characteristics and on unobservable factors like tacit component of clinical knowledge, *i.e.* a mix of holistic evaluation, medical experience and the intuition of physicians. As a consequence, the observed outcome can be endogenous with respect to the selection of the treatment and the control group. Nevertheless, several techniques have been applied to reduce the impact of selection bias like multivariate regression or nonparametric techniques based on the propensity score (the *ex ante* probability that a patient is assigned to treatment given their baseline characteristics). The goal of all propensity-score methods is to achieve a balance in observed covariates for the treated and control subjects [11, 12]. Among the main methods, two are particularly important: propensity score matching and propensity score-based regression adjustment. The former is a nonparametric method where the propensity score (logit or probit post-estimation of the *ex ante* probability of being treated) is used to group individuals into small groups and compare their outcome. After having compared the outcome of small groups (difference between the outcome of treated and untreated patients), the average difference is calculated and tested for significance [11]. As ROSENBAUM and RUBIN [13] stated in 1985, matched sampling is a method for selecting units from a large reservoir of potential controls to produce a control group of modest size that is similar to a treated group with respect to the distribution of observed covariates. Matching on the propensity score is a parsimonious way to reduce selection bias deriving from observational factors, only leaving space for selection bias based on unobservable factors. Some studies have discussed the contingent convenience of the specific matching technique [14–18]. The latter uses the propensity score as a covariate in the standard regression: in this way, a covariate in the regression adjusts the estimation for potential selection bias. Matching is not the only way of using propensity scores. In fact, regression adjustment and weighting are other techniques used to reduce selection bias [19]. This mini-review will focus on the general characteristics of bias reduction techniques that allow real-world studies to provide sound evidence that is qualitatively different from that of RCTs but often more useful and timely [20–22].

Table 1 describes in details the main differences between RCTs and real-world studies as to pros and cons of different sources of clinical evidence. In synthesis, real-world studies can substantially reduce the bias due to the lack of randomisation through matching techniques and can provide additional benefits to or in absence of RCTs, *i.e.* when RCTs are unethical, not possible or difficult to conduct [2], accounting for the learning curve, adhering to real-world clinical practice, implying cheaper and easier data collection methods, often relying on large existing samples, raising fewer ethical issues, and with research protocols potentially able to provide timely evidence to decision makers.

Policy makers are in fact supporting the use of real-world studies as a way to manage uncertainty and take locally grounded reimbursement decisions. Many countries have expressed commitment to considering the real value provided by new treatments so that effectiveness assessments can notably contribute to a new drug or device's market success through coverage and reimbursement decisions. Finally, members of the scientific community have also started to compare the evidence generated through real-world studies with RCTs, finding that “observational studies using [propensity score] methods produce treatment effect estimates that are of more extreme magnitude compared with those from RCTs, although the differences are rarely statistically significant” [24].

The case of Mitraclip

Mitraclip is an implantable device (“clip”) launched in 2008 allowing a fully percutaneous approach to mitral regurgitation. Four trials exist: COAPT, ISAR-CLIP, RESHAPE-HF and EVEREST II. The latter is

TABLE 1 Randomised clinical trials (RCTs) and real-world (RW) studies: pros and cons

Key element	RCTs	RW studies
Learning curve	Not considered Impacts can be biased if the protocol enhances the standardisation of the intervention	Can be controlled if the initial experience of the physician is known Could be useful to start data collection from the first intervention
Selection bias	Not applicable by design Nevertheless, some open-label studies allow patients to self-select In this case, it is possible to generate selection bias However, in the latter case, selection bias is seldom discussed in the study	The most relevant issue Appropriate data collection and statistical procedures can significantly reduce it Bias reduction can be measured and reported
Adherence to the real clinical practice	No: if the RCT's protocol serves to determine if a causal relationship between the technological innovation use and outcomes exists, it is often too distant from actual practice, especially in terms of patients' selection criteria and treatment protocol	Yes
Comparator	Head-to-head trials are not mandatory and placebo is often used This has increased the use of ex-post indirect comparisons (<i>i.e.</i> network meta-analysis) that often lead to opposite recommendations [23]	Always a direct comparator, used in real clinical practice (normally standard of care)
Costs	Generally high It may increase when replications are needed due to versioning	Low if the study is retrospective Comparable to an RCT if the study is prospective
Ethical issues	Requires approval by ethical committees because it alters clinical practice	Clinical practice is not altered and ethical issues mostly regard patients' privacy
Economic evaluation	Feasible ex-post Seldom, costs are also collected during the experimental phase but often irrelevant since protocol driven	Feasible in parallel for prospective studies For retrospective studies, costs are collected ex-post (however, the costs reflect clinical practice closer than in the case of RCTs)
Sample	Minimum required by the expected magnitude of the effect	Large Depending on the statistical requirements, a RW analysis may need many observations (because many observations might be lost if common support is imposed by PSM) or data regarding the period prior to the intervention (difference-in-differences)
Time to results	It depends on the protocol	Depends on the protocol (prospective studies) Short for retrospective studies Ideal for follow-up studies
Internal validity	High	Always inferior to RCTs However, appropriate study design and statistical methods can substantially increase RW study internal validity
External validity	Low The conditions of the RCT do not reflect those where the policy decision would apply	High The conditions of RW studies reflect routine practice

PSM: propensity score matching.

the first RCT assessing Mitraclip's efficacy *versus* conventional mitral valve surgery. Only one RCT (ISAR-CLIP) has reached its conclusion while the others have only released preliminary data. To keep collecting safety and effectiveness data, two registries were created and linked to the EVEREST trials: EVEREST II High Risk Registry and EVEREST II REALISM-High Risk. The technology started being implanted as early as 2008 (well before the publication of the first trials in 2012). In 2013, 8000

patients were treated with Mitraclip in 40 countries [25, 26]. In the same year, Italy accounted for >12% of all implants, with a growth rate from 2009 to 2013 of 1260% [26]. In 2016, the total number of implants worldwide has more than doubled to 30000 (Abbott Vascular; personal communication).

The high-risk study [27], based on EVEREST II, suggested that the technology was most effective when used for high-risk patients not eligible for surgery who would otherwise have been treated with medical care only. Therefore, the end-user should consider Mitraclip as an option when surgery is impossible. To the clinical efficacy, the policy maker needs to add the economic efficiency by estimating the cost per unit of effectiveness of such therapeutic option. A CEA [28] based on the high-risk study was the first attempt to help policy makers decide about Mitraclip reimbursement. However, a trial-based economic evaluation carried out when the technology is already in a post-adoption phase of diffusion fails to provide the policymaker with a credible representation of the reality: first, the high-risk trial compared Mitraclip (n=76) with a mix of medical management (n=33) and mitral valve surgery (n=5); second, the sample was considerably small and the obtained effect (76% 12-month survival in the treatment arm compared to 55% in the control arm) is not univocally referred to any of the comparators; third, the economic evaluation necessarily reflected an experimental design, thus failing to exploit the existing amount of observable evidence generated by >5 years of use in the clinical practice, with particular reference to the actual side-therapies and to the variation of effectiveness due to use-specific aspects. Despite the impressive use of the technology, as of today, only a few European governments and HTA bodies have explicitly issued a recommendation on coverage and reimbursement of Mitraclip: in the UK, the National Institute for Health and Care Excellence [29] advised not to fund Mitraclip routinely because of a lack of quality evidence; the Ludwig Boltzmann Institute in Austria [30] also suggested a negative recommendation due to lack of evidence; Osteba in Spain, despite underlining a low level of evidence, instead made a positive recommendation to include Mitraclip in the national health basket for high surgical risk patients currently undergoing medical therapy and presenting defined characteristics [31].

For this reason, a study group composed of health economists, cardiologists and cardiothoracic surgeons has carried out an economic evaluation of Mitraclip combined with medical therapy *versus* medical therapy alone in patients with moderate/severe mitral regurgitation and advanced heart failure [32]. The research team accepted the challenge of a real-world CEA. Since no data were collected with the intention of conducting an economic evaluation, a database had to be built putting together the set of clinical records of patients who received Mitraclip in two Italian centres at the time the study was initiated (n=232) and a sample of patients eligible for Mitraclip but who did not receive it (n=151) from the Italian Network on Acute and Chronic Heart Failure registry, a database promoted by the ANMCO (Italian Association of Hospital Cardiologists). Records of both the treatment and control groups were complete and detailed, and assumptions were made only for utility weights that were extracted from the literature for the inclusion of general age-related mortality in the Markov model and for some minor details regarding the causes of re-hospitalisations during the follow-up period. The two groups showed important differences at baseline, thus supporting the existence of selection bias. In a previous version of the manuscript, we adopted an unusual yet plausible approach, where the propensity score was estimated through a probit regression and used to match patients based on a 2/1 ratio. The idea was to exploit the data of the investigational group, similar to the randomisation approach of EVEREST II trial, using the same ratio. However, we recognised that using a propensity score for matching may have limited the power of the comparison. Indeed, when the sample size is medium, as in our case, using a propensity score for adjustment rather than for matching purposes may have been more instrumental in preserving the power of the comparison while addressing the impact of identified confounders [33]. For these reasons, we decided to use a different adjustment approach for our comparison. In particular, we re-calculated the propensity score and used it as a covariate in a Cox proportional hazard regression. In this way, we maintained the number of patients in the original groups, reducing the possibility of Type II error (*i.e.* false-negative results). Notably, our results for the main clinical outcome of interest (survival) did not change compared with the previous version of the manuscript and thereby did not influence significantly the outcomes of the CEA where survival is just one piece of the puzzle. Since the propensity score showed a good ability to control confounding (c-statistic >0.8), the regression adjustment was able to significantly reduce the component of selection bias relative to observable factors. The selection bias relative to unobservable factors, instead, remained in the data as a noneliminable bias. Several journals rejected the paper based on this limitation (*i.e.* the inherent difference between an RCT and an observational study), arguing that in the absence of experimental evidence, an economic evaluation was premature.

In brief, the scientific community showed a negative position towards the acceptance of a real-world study. Cost-effectiveness, according to reviewers, is reliable if based upon experimental evidence, although the latter has not been claimed to implant the device. No one agreed to consider observational evidence different but not inferior to experimental evidence with respect to the objective of the evaluation. For

policy decisions (e.g. coverage and reimbursement), external validity, adherence to real clinical practice and prompt evidence (real-world studies) are as valuable as internal validity (RCTs), and certainly more relevant than experimental study designs.

Discussion and conclusions

Innovation of medical devices is different enough from other technologies, namely drugs, to require different approaches to be assessed. Large investments in research are being made by several stakeholders, including the European Union (www.medtehta.eu/wps/wcm/connect/Site/MedtechHTA/Home, www.adhophta.eu and www.advance-hta.eu), to deepen our knowledge on current methods for assessing medical devices, especially dealing with clinical evidence. The debate on regulation of medical devices has very often focussed on the type and role of clinical evidence [3, 34–36]. Clinical evidence for medical devices is often more challenging to collect through traditional experimental studies such as RCTs and, although RCTs rank high in terms of internal validity, other sources of data can be relevant and reliable to take policy decisions [37], especially when the adoption and diffusion of new devices largely precede any RCTs. Real-world data represent a valid complement and/or alternative to RCTs when the latter are unable to measure the intrinsic characteristics (e.g. learning curve, incremental innovation, organisational impact and price dynamics) of medical devices and to estimate their impact on the final decisions in terms of coverage and reimbursement. Nevertheless, real-world data are not problem-free and selection bias can invalidate the final results if not correctly handled. There are several statistical techniques proven to be effective to reduce selection bias and extensively used in this field [19–22].

Our CEA analysis of Mitraclip plus OMT *versus* OMT alone made use of propensity score-based regression adjustment to inform decision-makers, including clinicians, as to whether its incremental costs outweigh incremental effectiveness [31].

Clinical evidence drawn from real-world studies will never be as reliable as that drawn from a perfect RCT in identifying the causal relationship between intervention and changes in clinical conditions. However, when RCTs are not possible or there is the potential to alter the randomisation because blinding is not possible, their internal validity is lower and there is no reason to argue that the selection bias reintroduced in this way is less severe than that inherent in real-world studies.

Real-world studies do not aim to show whether a technology works but, instead, if it can work in actual clinical practice. From a policy perspective, the message of a real-world study is different but not inferior to the message of a RCT. Policy makers should assess whether the technology is worth investing (incremental clinical value coupled with sustainability) and should, therefore, be more concerned with evidence produced in real settings rather than experimental ones.

Comparative RCTs for medical devices often follow the product's launch and may be available after a considerable amount of time. However, this does not delay the diffusion process, which means that real-world data can be collected and promptly assessed in the meantime so to support policy makers to provide recommendations, although not necessarily conclusive at first instance (e.g. coverage with evidence development), to control diffusion while waiting for additional evidence. If, conversely, HTA and policy recommendations are not timely, the risk is that they might be ineffective or less incisive since when they become available, the technology has already been integrated in clinical practice. Finally, we believe there is a need to acknowledge the importance of real-world studies and engage the scientific community, especially those physicians keen to adopt innovative technologies, in the promotion and use of clinical evidence produced through observational studies.

If, in the absence of RCTs, we do not believe in methodologically sound real-world data to produce clinical and economic recommendations on technological innovations, how can we dare to implant even one of these innovations in real patients?

References

- 1 EUCOMED. The European Medical Technology Industry in Figures. Brussels, MedTech Europe, 2012.
- 2 Drummond M, Griffin A, Tarricone R. Economic evaluation for devices and drugs – same or different? *Value Health* 2009; 12: 402–404.
- 3 Tarricone R, Torbica A, Ferré F, *et al.* Generating appropriate clinical data for value assessment of medical devices: what role does regulation play? *Expert Rev Pharmacoecon Outcomes Res* 2014; 14: 707–718.
- 4 Buxton MJ. Problems in the economic appraisal of new health technology: the evaluation of heart transplants in the UK. In: Drummond MF. *Economic appraisal of health technology in the European community*. Oxford, Oxford Medical Publications, 1987; pp. 103–118.
- 5 Dreyer NA, Tunis SR, Berger M, *et al.* Why observational studies should be among the tools used in comparative effectiveness research. *Health Affairs* 2010; 29: 1818–1825.
- 6 Annemans L, Aristides M, Kubin M. Real-Life Data: A Growing Need. www.ispor.org/news/articles/Oct07/RLD.asp
- 7 Dreyer NA, Schneeweiss S, McNeil B, *et al.* GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010; 16: 467–471.

- 8 Johnson ML, Crown W, Martin BC, *et al.* Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report-Part III. *Value Health* 2009; 12: 1062–1073.
- 9 Garrison LP Jr, Neumann PJ, Erickson P, *et al.* Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health* 2007; 10: 326–335.
- 10 Berger ML, Mamdani M, Atkins D, *et al.* Good research practices for comparative effectiveness research: defining, reporting, and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force report-part I. *Value Health* 2009; 12: 1044–1061.
- 11 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41–55.
- 12 Peikes DN, Moreno L, Orzol SM. Propensity score matching. *Am Stat* 2008; 62: 3.
- 13 Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 1985; 39: 33–38.
- 14 Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *J Thorac Cardiovasc Surg* 2007; 134: 1128–1135.
- 15 Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27: 2037–2049.
- 16 Hill J. Discussion of research using propensity-score matching: comments on ‘A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003’ by Peter Austin. *Stat Med* 2008; 27: 2055–2061.
- 17 Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. *Biom J* 2009; 51: 171–184.
- 18 Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health* 2006; 9: 377–385.
- 19 Kurth T, Walker AM, Glynn RJ, *et al.* Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006; 163.3: 262–270.
- 20 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41–55.
- 21 Dehejia RH, Wahba S. Propensity score-matching methods for Nonexperimental causal studies. *Rev Econ Stat* 2002; 84: 151–161.
- 22 Peikes DN, Moreno L, Orzol SM. Propensity score matching: a note of caution for evaluators of social programs. *Am Stat* 2008; 62: 222–231.
- 23 Spinner DS, Birt J, Walter JW, *et al.* Do different clinical evidence bases lead to discordant health-technology assessment decisions? An in-depth case series across three jurisdictions. *Clinicoecon Outcomes Res* 2013; 5: 69–85.
- 24 Dahabreh IJ, Sheldrick RC, Paulus JK, *et al.* Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. *Eur Heart J* 2012; 33: 1893–1901.
- 25 Minha S, Torguson R, Waksman R. Overview of the 2013 Food and Drug Administration Circulatory System Devices Panel meeting on the MitraClip Delivery System. *Circulation* 2013; 128: 864–868.
- 26 Abbot Vascular. Percutaneous mitral valve repair with the Mitraclip clip delivery system in high surgical risk patients. www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/CirculatorySystemDevicesPanel/UCM343684.pdf Date last accessed: May 7, 2016.
- 27 Whitlow PL, Feldman T, Pedersen WR, *et al.* Acute and 12-month results with catheter-based mitral valve leaflet repair: the EVEREST II (Endovascular Valve Edge-to-Edge Repair) High Risk study. *J Am Coll Cardiol* 2012; 59: 130–139.
- 28 Mealing S, Feldman T, Eaton J, *et al.* EVEREST II high risk study based UK cost-effectiveness analysis of MitraClip in patients with severe mitral regurgitation ineligible for conventional repair/replacement surgery. *J Med Econ* 2013; 16: 1317–1326.
- 29 NHS Commissioning Board. Clinical Commissioning Policy Statement: Percutaneous mitral valve leaflet repair for mitral regurgitation. www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2013/09/a09-ps-b.pdf Date last accessed: May 29, 2016.
- 30 Ludwig Boltzmann Institute. Perkutane Mitralklappenintervention mittels Mitralclip bei Mitralklappeninsuffizienz [Percutaneous mitral valve intervention device Mitraclip for mitral valve insufficiency]. http://eprints.hta.lbg.ac.at/967/1/DSD_41_Update2012.pdf Date last accessed: May 7, 2016.
- 31 Osteba. MitraClip® para la reparación del reflujo de la válvula mitral: informes de evaluación de tecnologías sanitarias [MitraClip® for repair of mitral valve regurgitation: health technology assessment information]. www.osakidetza.euskadi.eus/contenidos/informacion/2014_osteba_publicacion/es_def/adjuntos/MitraClip.pdf Date last accessed: May 7, 2016.
- 32 Armeni P, Boscolo PR, Tarricone R, *et al.* Real-world cost effectiveness of MitraClip combined with medical therapy *versus* medical therapy alone in patients with moderate or severe mitral regurgitation. *Int J Cardiol* 2016; 209: 153–160.
- 33 Biondi-Zoccai G, *et al.* Are propensity scores really superior to standard multivariable analysis? *Contemp Clin Trials* 2011; 32: 731–740.
- 34 Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. *N Engl J Med* 2012; 366: 848–855.
- 35 Boulton AJM. Registration and regulation of medical devices used in diabetes in Europe: need for radical reform. *Lancet Diabetes Endocrinol* 2013; 1: 270–272.
- 36 Zuckerman D, Brown P, Das A. Lack of publicly available scientific evidence on the safety and effectiveness of implanted medical devices. *JAMA Intern Med* 2014; 174: 1781–1787.
- 37 Ciani O, Taylor RS. A more evidence based approach to the use of surrogate end points in policy making. *BMJ* 2011; 343: d6498.