Practice of Epidemiology

Methodological Considerations in Observational Comparative Effectiveness Research for Implantable Medical Devices: An Epidemiologic Perspective

Jessica J. Jalbert, Mary Elizabeth Ritchey, Xiaojuan Mi, Chih-Ying Chen, Bradley G. Hammill, Lesley H. Curtis, and Soko Setoguchi*

* Correspondence to Dr. Soko Setoguchi, Duke Clinical Research Institute, P.O. Box 17969, Durham, NC 27715 (e-mail: soko.setoguchi@duke.edu).

Initially submitted June 19, 2013; accepted for publication July 10, 2014.

Medical devices play a vital role in diagnosing, treating, and preventing diseases and are an integral part of the health-care system. Many devices, including implantable medical devices, enter the market through a regulatory pathway that was not designed to assure safety and effectiveness. Several recent studies and high-profile device recalls have demonstrated the need for well-designed, valid postmarketing studies of medical devices. Medical device epidemiology is a relatively new field compared with pharmacoepidemiology, which for decades has been developed to assess the safety and effectiveness of medications. Many methodological considerations in pharmacoepidemiology apply to medical device epidemiology. Fundamental differences in mechanisms of action and use and in how exposure data are captured mean that comparative effectiveness studies of medical devices often necessitate additional and different considerations. In this paper, we discuss some of the most salient issues encountered in conducting comparative effectiveness research on implantable devices. We discuss special methodological considerations regarding the use of data sources, exposure and outcome definitions, timing of exposure, and sources of bias.

comparative effectiveness; epidemiologic methods; medical device epidemiology; pharmacoepidemiology; prostheses and implants; United States Food and Drug Administration

Implantable medical devices are used to restore body functions, prevent life-threatening events, prolong life, and improve quality of life. Use of implantable devices has increased steadily (1), and with their use being concentrated in older populations, the trend is likely to continue. The Food and Drug Administration (FDA) classifies medical devices into 3 classes according to the level of control necessary to assure safety and effectiveness (2). Class III devices include most implantable devices (e.g., hip implants, pacemakers, stents), pose the highest risks, and are intended to undergo premarket approval, analogous to a new drug application. However, the dominant mechanism used for market entry is Section 510(k) of the Food, Drug and Cosmetic Act, which requires that the device be “substantially equivalent” to one already being marketed (3, 4). Devices entering the market through this process are “cleared” rather than “approved” by the FDA.

Prompted by several high-profile recalls of medical devices, the FDA recently asked the Institute of Medicine to conduct an independent review of the clearance process (4). The Institute of Medicine concluded that the process was insufficient to assure safety and effectiveness and also detailed weaknesses in the FDA’s postmarketing surveillance system. Thus, not only are there gaps in the understanding of device safety and effectiveness at the time of market entry, but there is also a need for high-quality data to fill those gaps in the postmarketing setting.

Although the objectives of pharmacoepidemiology and medical device epidemiology are similar, differences between devices and medications have important implications for how epidemiologic methods are developed and applied. In this article, we discuss special methodological considerations in device comparative effectiveness research with regard to data sources, exposures, outcomes, time, and bias.
DATA SOURCES FOR EVALUATING DEVICE EFFECTIVENESS

Administrative claims databases

Administrative claims databases are derived from claims submitted for payment for clinical services and treatments. Within these databases, there exist neither unique device identifiers (UDIs) nor a claims-based system permitting identification of medical devices with sufficient granularity. In claims data, implantable devices are captured using Current Procedural Terminology or *International Classification of Diseases* procedure codes. Different health-care settings may use different coding systems or a mix of coding systems, as is the case with Medicare data. Implantation of a device in an inpatient setting will be coded using both *International Classification of Diseases* codes (inpatient file) and Current Procedural Terminology codes (physician/carrier file), whereas an outpatient procedure will be coded using Current Procedural Terminology codes. Although Current Procedural Terminology codes tend to be more specific than *International Classification of Diseases* codes, both types of codes relate to the implantation rather than the device itself and often convey little information about the actual device.

Sampled national data sources

National surveys such as the Nationwide Inpatient Sample (5) and the Nationwide Emergency Department Sample (6) provide data on representative samples of inpatient and emergency department visits, respectively, and can be used to derive national estimates. However, data on implantable medical devices are also captured using procedure codes and thus are subject to the same limitations as those detailed above for administrative data. Moreover, device performance can only be assessed during the encounter, because data are not captured longitudinally.

Device registries

Device registries can be used to obtain detailed information on devices and clinical conditions (see Table 1 for a list of example registries). In our experience, a major limitation of device registries for comparative effectiveness research is the limited follow-up information. Registries may only collect data during the encounter for the implantation up to the time of discharge or may have significant loss to follow-up. For example, the Centers for Medicare & Medicaid Services created an implantable cardioverter-defibrillator (ICD) registry and a carotid artery stenting database in response to changes in the National Coverage Determinations for these devices. Neither data source, however, contains follow-up data beyond the index hospitalization. In the Society for Vascular Surgery’s Vascular Registry (2005–2008), follow-up information was available for 59.0% of carotid artery stenting and 54.3% of carotid endarterectomy patients, while those who were followed had mean and median follow-up durations of less than 1 year. For identification of long-term outcomes, device registries have been linked to administrative databases such as Medicare (7, 8).

Many device registries collect information on patients receiving a device but not on patients receiving “standard-of-care” or alternative treatments. In this situation, other disease, drug, or device registries have been used to identify comparison groups (8, 9). Because registries are often focused on a specific procedure or device, comparative effectiveness analyses using multiple registries and/or data sources may need to be repeated, especially if more than 1 comparator is possible. Published literature may also be used to provide aggregate data for a historical or objective performance criterion for comparison. The aggregate comparison afforded by published literature, however, may be insufficient for comparative effectiveness analyses, because information supplied within a published article may not be granular to the level of the individual, and thus accounting for differences between the intervention and comparator groups may not be possible.

Data quality and granularity can be heterogeneous across and within registries and can change over time as data collection forms are updated or reporting requirements change. Whether the data are complete and accurate depends highly on the procedures in place to ensure that the highest-quality data are recorded at the time of collection (e.g., clear data element definitions and training of data collectors) and to identify and correct sources of error after collection (e.g., data checks and imputation). Within a device registry, data quality may vary highly depending on differences in regulatory structures. For example, the Improving Pediatric and Adult Congenital Treatment (IMPACT) Registry is a voluntary national registry, but the contribution of states like Massachusetts and New York is mandatory with data adjudication. Systematic ascertainment biases across individual centers, states, or systems contributing to a registry could result in treatment effect heterogeneity across centers, states, or systems. Both data quality and granularity can limit the analyses that can be performed to certain time periods, certain data contributors, or certain groupings of devices.

Medical records

The pros and cons of using medical records for research purposes in pharmacoepidemiology are largely the same as those for device comparative effectiveness research. Adoption of electronic health records (EHRs) by US hospitals has been slow, with fewer than 10% having a basic EHR system by 2008 (10). In facilities with EHRs, much of the information may be in the form of unstructured free text, which is difficult to search, summarize, and analyze. Text mining may be used to “numericize” and extract the information but can be labor-intensive. Many EHRs are also institution-specific, such that services rendered outside the institution are not captured, and the lack of standardization also hinders efforts to combine EHRs across multiple institutions.

IDENTIFYING DEVICES

Unique device identifiers

Devices and their components are often not consistently identifiable with great specificity, such as by brand, model,
version, or design. The FDA Amendments Act of 2007 charged the agency with developing and instituting a comprehensive UDI system for medical devices. The final rule was released in September 2013 and required that labels and packages for class III medical devices include UDIs by September 2014 and that the remainder of implantable devices include UDIs by September 2015 (11). The UDI system is a requirement for manufacturers and is expected to assist with the identification of medical devices for adverse event reporting and with the reduction of errors at the point of

Table 1. Characteristics of Cardiovascular and Orthopedic Device Registries in the United States

<table>
<thead>
<tr>
<th>Data Source and Registry</th>
<th>Exposure(s)</th>
<th>Type of Information Collected</th>
<th>Participation</th>
<th>Catchment Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Cardiovascular Data Registry&lt;sup&gt;a&lt;/sup&gt; (45)</td>
<td>Carotid artery stenting, carotid endarterectomy</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Voluntary</td>
<td>National</td>
</tr>
<tr>
<td>CARE Registry</td>
<td>Implantable cardioverter-defibrillators</td>
<td>Clinical, procedural, implant, and in-hospital outcomes</td>
<td>Mandatory</td>
<td>National</td>
</tr>
<tr>
<td>ICD Registry</td>
<td>Cardiac catheterization</td>
<td>Clinical, procedural, implant, and in-hospital outcomes</td>
<td>Voluntary</td>
<td>National</td>
</tr>
<tr>
<td>IMPACT Registry</td>
<td>Transcatheter aortic valve replacement</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Voluntary</td>
<td>National</td>
</tr>
<tr>
<td>TVT Registry</td>
<td>Percutaneous coronary intervention with or without stenting</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Voluntary</td>
<td>National</td>
</tr>
<tr>
<td>Veterans Administration data sources</td>
<td>Implantable cardioverter-defibrillators</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Voluntary</td>
<td>National</td>
</tr>
<tr>
<td>CART-CL</td>
<td>Pacemaker</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Voluntary</td>
<td>National</td>
</tr>
<tr>
<td>VHA National ICD Surveillance Center</td>
<td>Total hip arthroplasty, total knee arthroplasty</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Voluntary</td>
<td>National</td>
</tr>
<tr>
<td>Pacemaker Surveillance Program</td>
<td>Total hip surgery, total knee surgery</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Massachusets General Hospital (Boston, Massachusetts)</td>
<td></td>
</tr>
<tr>
<td>IMPACT Registry</td>
<td>Hip arthroplasty, knee arthroplasty, spine arthroplasty, shoulder arthroplasty</td>
<td>Clinical, procedural, implant, and in-hospital outcomes</td>
<td>HealthEast Care System Hospitals (St. Paul, Minnesota)</td>
<td></td>
</tr>
<tr>
<td>IMPACT Registry</td>
<td>Hip arthroplasty, hip revision, knee arthroplasty, shoulder arthroplasty, shoulder revision</td>
<td>Clinical, procedural, implant, and in-hospital outcomes</td>
<td>Voluntary</td>
<td>1 US hospital</td>
</tr>
<tr>
<td>IMPACT Registry</td>
<td>Total hip arthroplasty, total knee arthroplasty</td>
<td>Clinical, procedural, implant, and in-hospital outcomes; follow-up data for revision only</td>
<td>Kaiser Permanente health-care plan</td>
<td></td>
</tr>
<tr>
<td>IMPACT Registry</td>
<td>Total hip arthroplasty, hip revision, total knee arthroplasty, knee revision</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>Mayo Clinic (Rochester, Minnesota)</td>
<td></td>
</tr>
<tr>
<td>IMPACT Registry</td>
<td>Hip replacement, knee replacement</td>
<td>Clinical, procedural, implant, in-hospital, and follow-up outcomes</td>
<td>National, 165 hospitals</td>
<td></td>
</tr>
<tr>
<td>IMPACT Registry</td>
<td>Hip arthroplasty, knee arthroplasty</td>
<td>Clinical, procedural, implant, and in-hospital outcomes</td>
<td>Virginia</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CARE, Carotid Artery Revascularization and Endarterectomy; CART-CL, Cardiac Assessment, Reporting, and Tracking System for Cardiac Catheterization Laboratories; CERT, Centers for Education & Research on Therapeutics; GLORY, Global Orthopaedic Registry; HSS, Hospital for Special Surgery; ICD, implantable cardioverter-defibrillator; IMPACT, Improving Pediatric and Adult Congenital Treatments; TVT, Transcatheter Valve Therapy; VHA, Veterans Health Administration.

<sup>a</sup> American College of Cardiology, Washington, DC.

Am J Epidemiol. 2014;180(9):949–958
care. The UDI requirements do not extend to administrative billing data or EHRs. However, capture of UDI data would provide detailed information about a device which would be useful for payers and researchers (12). As the UDI is implemented, it will probably be incorporated differentially by data source and region, and even across facilities. Researchers should treat UDI-based variables with the same rigor and skepticism as they would other variables that are subject to differential reporting and missingness. Given the limitations involved in assessing the performance of specific devices and models, we have limited our discussion to methodology issues associated with assessing device types or categories.

Device life cycles

Medical devices tend to have shorter life cycles, or time on the market before being replaced, than medications. Incremental changes to devices are common, and if the modifications are not considered to significantly impact safety or effectiveness, the sponsor may receive clearance for newer models without submitting data to the FDA. This regulatory environment allows marketed devices to evolve continually. Over time, a device may become substantially different from the one that was originally approved or cleared (13, 14). Minor changes in design could also impact the device’s safety or effectiveness profile, as was the case for the femoral component of a hip replacement device which was found to fracture prematurely on the left side when the location of the etched icon of the company was changed to the left side (15). Short life cycles complicate the study of the long-term safety and effectiveness of a specific device design or feature. Even in data sets with detailed device information, too few people may have received a given device to meaningfully assess performance. On the other hand, when available information fails to distinguish between device types or characteristics, estimates may consist of uniform associations between all device subtypes and the outcome or these associations may be nonuniform, stemming from differences in performance or preferential use of certain subtypes of the device. Stratifying by model, brand, or design may be particularly informative in cases where device mechanics (e.g., pulsatile vs. continuous-flow ventricular assist devices), components (e.g., synthetically derived vs. non-synthetically derived urogynecological mesh), or design (e.g., tapered vs. untapered stents) differ radically within a class.

Distinguishing the anatomical side of a device

Some devices may be implanted in 1 anatomical structure on either side of the body. The same device-related procedure codes could signify a device revision, an implantation on the opposite side of the body, or a bilateral or staged procedure. Distinguishing between revisions and new implantations can be accomplished if laterality is known and if information about previous procedures and/or indication is available.

In claims databases, modifier codes can indicate implantation side and bilateral or staged procedures. The degree of missingness in claims is procedure-dependent, as reimbursement is not affected by whether or not information on laterality is provided (16). For example, among Medicare beneficiaries undergoing total knee replacement between 2000 and 2009, modifier codes indicating the procedural side were complete in 52% of nonbilateral procedures (17), while information on laterality was available for fewer than 10% of Medicare beneficiaries undergoing carotid artery stenting (J. J. Jalbert, Brigham and Women’s Hospital, personal communication, 2012) but for close to 90% for age-related macular degeneration procedures performed between 2006 and 2008 (L. H. Curtis, Duke University, personal communication, 2014) (18). In medical records, the quality of data on laterality is likely to vary across institutions, but 1 study found that it was incomplete or inaccurate in 34% of operative notes (19). Registries for devices implanted in body parts with bilaterality generally collect information on the implantation side, with few or no missing data. For example, the National Cardiovascular Data Registry’s Carotid Artery Revascularization and Endarterectomy (CARE) Registry (2005–2009) and Kaiser Permanente’s National Total Joint Replacement Registry (2001–2005) have minimal missing data on laterality (20, 21). Complete information on laterality may stem from quality control practices, imputation, or selective release of records meeting completeness thresholds. Researchers should be aware of quality control standards used by the keepers of the data repository.

Identifying “true” exposure

Medical devices regulating biological processes may have more than 1 exposure type. For example, ICDs and pacemakers deliver electrical impulses to maintain adequate heart rhythm or rate. “True” exposure can be defined as either permanent implantation of the device or the intermittent shocks it administers. Data may be collected and stored in devices like ICDs (analogous to having a medication event monitoring system) and transmitted through networks to treating physicians (22). When such data are available, the number and strength of shock therapy treatments can be quantified, the appropriateness of shock therapy can be determined, and different exposure categories can be created. While, at a minimum, it is generally possible to identify medical device class, information on the frequency or strength of electrical impulses is not collected in most observational data sources. When data on the activity of such devices are lacking, “true” exposure is unknown. This is analogous to the situation in pharmacoepidemiology, where it is not possible to infer from dispensing dates in claims data precisely how patients took their medications. In medical device epidemiology, as in pharmacoepidemiology, it may be necessary to make assumptions about exposure based on available information and to conduct sensitivity analyses.

IDENTIFYING COMPARATORS

At least 2 potential issues need to be considered when using comparison groups from different databases. First, bias due to noncomparability is possible, because exposed and unexposed patients may have different characteristics as a result of being sampled differently and/or because they come from different source populations (23). In a study
linking a device and a disease registry to Medicare data, it is possible that patients from the registry will differ with respect to demographic characteristics, comorbidity burden, disease severity, and/or geographical location even though they are all Medicare beneficiaries. Since information on some of these factors may be unmeasured, comparisons across different databases could be biased due to noncomparability of exposed and unexposed populations. Bias may be minimized through consideration of hospital clusters and controlling for proxies of the potential confounders. Second, even if a comparable group is identified, different data collection forms, definitions, and instructions may be used. This non-comparability of available information for confounder adjustment may result in residual confounding.

IDENTIFYING OUTCOMES OF INTEREST

Periprocedural outcomes

Although most medications are administered and studied in outpatient settings, device implantations frequently require a brief hospitalization. Establishing when an outcome occurs relative to device implantation during this hospitalization can be challenging in administrative databases where procedure dates are recorded but dates of diagnosis are not, particularly when the outcome is also a risk factor. For instance, carotid artery stenting is indicated for stroke prevention, but stroke is both a risk factor for the procedure and a potential complication resulting from the procedure (Figure 1). Variables collected in administrative data may be used to develop algorithms to disentangle preprocedural events from postprocedural events. The admitting diagnosis may be used to identify conditions preceding the implantation but is rarely validated. The Centers for Medicare & Medicaid Services recently implemented a "present on admission" variable that applies to all listed diagnoses and indicates whether the condition was present at the time of inpatient admission. Discharge diagnostic codes exist for some postoperative complications such as stroke but may have low sensitivity, particularly for patients with multiple comorbid conditions, and they are also rarely validated.

Length of hospital stay may offer some additional insights, but attributing a lengthy hospitalization to a certain complication should be done on a case-by-case basis and by examining all diagnoses and procedures listed for that hospitalization, as well as patterns of care. While registries may not suffer from this problem, lack of follow-up data may limit analyses of periprocedural outcomes to in-hospital complications.

Long-term outcomes

Medications consist of chemicals that must be administered repeatedly or released in a sustained fashion to maintain their therapeutic effects. While medical devices such as dermal fillers may require repeat administration, many devices maintain their therapeutic effects through an unchanging vehicle consisting of mechanical and electronic components. Thus, questions about the long-term safety or effectiveness of implantable devices mainly concern the durability of the devices. The definition of “long-term” depends on the expected life span of the device, with some intended to last for the duration of the patient’s life.

Assessing long-term outcomes using registries is difficult, since many registries include 30 or fewer days of follow-up. In the United States, no single data source covers patients throughout their lives, such that all observations are left- and/or right-censored. Once patients reach 65 years of age, most can be followed in Medicare data for the remainder of their lives. However, the issues of device durability and long-term outcome assessment are more relevant for younger subjects with longer life expectancies. A hybrid approach combining exposure and baseline covariate assessment in a registry or database with prospective collection of data on the outcome variable and time-varying covariates from another database could provide a better understanding of long-term outcomes, especially in younger subjects. As was noted above, when using multiple databases, differential ascertainment of variables between the 2 data sources is possible. Thus, when using a hybrid approach, assessment of the differences in ascertainment is needed.

Figure 1. Possible timing of stroke occurrence relative to a carotid artery stenting (CAS) procedure when a Medicare patient has a discharge diagnosis of stroke.

*Am J Epidemiol.* 2014;180(9):949–958
BIAS-RELATED CONSIDERATIONS
Confounding by indication or severity

Depending on the condition, medications may be used for patients with milder disease and device implantation may be reserved for those with more severe disease (or vice versa). A stepwise approach to treatment might also be recommended; for instance, a medical device may be considered if drug treatment has failed. Although knowledge of the guidelines and standards of practice is necessary to assess the direction, magnitude, and potential for confounding by indication or severity, this is not different from within-drug or within-procedure comparisons.

With sufficiently detailed information on disease severity or indication, confounding by indication or severity may be handled through study design considerations or analytically. In a recent study of the comparative effectiveness of ICDs using Medicare-linked registry data (2005–2008), we compared baseline characteristics of patients receiving ICDs with the baseline characteristics of those who did not, before and after restricting the population to patients who were eligible for ICDs for primary prevention (see Tables 2 and 3). The difference in ejection fraction between the 2 groups (48% in non-ICD patients vs. 25% in ICD patients) was significantly reduced after restricting the data to patients who were eligible for an ICD for primary prevention (29% in non-ICD patients vs. 25% in ICD patients). Adjustment for confounding by severity or indication may also be done analytically. High-dimensional propensity score modeling, which uses codes as proxies for confounding adjustment, may minimize confounding by severity or indication when administrative data alone are available (24), but, depending on the comparison, this propensity score technique may not be sufficient to remove all confounding factors (25).

Healthy-user bias

Surgical and other procedures pose short-term risks in exchange for long-term benefits. Patients with severe target disease or comorbid conditions are less likely to be considered for invasive procedures. The selection of healthier patients into more invasive treatments for the same indication can be problematic when comparing different treatment modalities, such as devices versus medical management. Healthy-user bias is not only a function of disease severity but also of better overall physiological health, functional status, psychological well-being, and more health-seeking behaviors (26, 27)—factors that are either difficult to quantify or not routinely measured in administrative, registry, or EHR data. Controlling for healthy-user bias may only be achieved in observational studies when information on health behaviors or their surrogates is available or a good instrument for performing an instrumental-variable analysis exists. Sensitivity analyses assessing the impact of healthy-user bias are necessary, and more research is needed to understand factors associated with the selection of patients to understand the magnitude of healthy-user bias in comparative effectiveness research on medical devices.

Provider characteristics

Characteristics such as physician specialty (28, 29), hospital specialty (30–32), hospital ownership (30), hospital size

### Table 2. Characteristics of Non-ICD Patients Versus ICD Patients Before Application of Primary Prevention ICD Indications, 2005–2008

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Non-ICD Patients (n = 535,885)</th>
<th>ICD Patients (n = 4,990)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years*</td>
<td>69 (20)</td>
<td>66 (19)</td>
</tr>
<tr>
<td>Male sex</td>
<td>278,344 52</td>
<td>3,469 70</td>
</tr>
<tr>
<td>White race</td>
<td>354,414 66</td>
<td>3,528 71</td>
</tr>
<tr>
<td>Ejection fraction, %</td>
<td>48 (33–59)</td>
<td>25 (20–30)</td>
</tr>
<tr>
<td>Admission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>138 (119–158)</td>
<td>127 (112–144)</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>75 (64–88)</td>
<td>73 (64–84)</td>
</tr>
<tr>
<td>Serum sodium, mEq/L</td>
<td>138 (136–141)</td>
<td>139 (136–141)</td>
</tr>
<tr>
<td>Serum B-type natriuretic peptide, pg/mL</td>
<td>761 (375–1,546)</td>
<td>822 (330–1,707)</td>
</tr>
<tr>
<td>Serum creatinine, mg/dL</td>
<td>1.2 (1.0–1.7)</td>
<td>1.2 (1.0–1.5)</td>
</tr>
<tr>
<td>Discharge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>121 (109–137)</td>
<td>118 (105–131)</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>67 (59–75)</td>
<td>68 (60–76)</td>
</tr>
<tr>
<td>Serum sodium, mEq/L</td>
<td>138 (136–140)</td>
<td>137 (135–139)</td>
</tr>
<tr>
<td>Serum B-type natriuretic peptide, pg/mL</td>
<td>524 (250–1,114)</td>
<td>561 (227–1,300)</td>
</tr>
<tr>
<td>Serum creatinine, mg/dL</td>
<td>1.3 (1.0–1.9)</td>
<td>1.2 (1.0–1.6)</td>
</tr>
</tbody>
</table>

Abbreviations: ICD, implantable cardioverter-defibrillator; IQR, interquartile range.

* Values are expressed as mean (standard deviation).
and teaching status (32) affect patient outcomes during the periprocedural period. Both the lead-in phase of the Carotid Revascularization Endarterectomy vs. Stenting Trial (CREST) (33) and the trial itself (34) suggested that physician specialty may affect outcomes following carotid artery stenting. Physician specialty or provider preference may exert a strong influence on therapeutic modality selection (devices vs. drugs) and on what types of devices the patient ultimately receives. A volume-outcome relationship between higher procedural volume (35, 38–42) and hospital procedural volume (35, 38–42) and lower periprocedural outcome risks have also been documented for a number of devices and procedures. A recent study found a strong inverse association between past-year physician and hospital carotid artery stenting volume and 30-day mortality risk among Medicare beneficiaries (2005–2009), after adjustment for patient and provider characteristics (43). More technically complex device deliveries and implantations will have significant learning curves associated with their use that are likely to change over time, particularly for first-in-class devices or for those devices with significant changes in their use.

Exploring provider characteristics to account for them appropriately is essential to conducting valid comparative effectiveness research on medical devices. Provider characteristics which may affect exposure selection and outcomes should be evaluated as potential confounders and effect-measure modifiers. In the presence of a learning curve, device performance and, concomitantly, comparative effectiveness estimates may change over time. To obtain valid inferences, clustering of patients should also be taken into account by introducing random or fixed effects or by using the sandwich variance estimator.

**Device failure or removal (sick stopper bias)**

Patients taking medications could have various degrees of adherence, from completely stopping the medication to skipping doses to taking the medication as prescribed. In most data sources, measuring medication adherence requires making strong assumptions. While adherence issues do not generally apply to implantable medical devices, devices can fail, be turned off, and/or be removed. Whether or not to take this into account depends on the study question. If the goal is to assess real-world effectiveness, where some degree of device failure or removal is expected, comparing the different modalities without adjusting for device failure or removal rates but reporting device failure or removal rates is appropriate. If the goal is to compare conditional effectiveness assuming no device failure or removal, this must be taken into account in the analysis.

**Bias related to treatment assignment: immortal time bias and time-varying confounding by severity**

Comparative effectiveness studies of devices are potentially susceptible to immortal person-time, a period of follow-up time during which, by design, the outcome(s) of interest cannot occur (44, 45). It arises when survival time (i.e., time from the beginning of a defined follow-up period to the development of the study outcome) is a study endpoint and when the start of the follow-up period varies for different treatment groups. This arises from the fact that treatment assignment is not decided at the time that observation of the subjects begins. In comparing devices with medications, immortal person-time may arise when treatments are used in a
stepwise manner (e.g., drug first and then the device if the drug was not effective). For the device user, time spent on medication may be a period during which, by design, the outcomes of interest cannot occur. Either misclassifying or omitting immortal time from the analysis can result in immortal person-time bias (46, 47). This bias is described extensively in pharmacoepidemiology but less so in studies of medical devices (45, 47–50).

Immortal person-time bias was initially described in the context of assessing the survival benefit of heart transplants in the 1970s (51). To illustrate this type of bias, consider a study comparing the effectiveness of left ventricular assist devices (LVADs) with heart transplantation in the era of the Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure (REMATCH) Study (52), which demonstrated the efficacy of long-term LVAD as destination therapy. Most patients awaiting heart transplantation receive an LVAD as bridge therapy. However, given that patients must survive long enough to undergo transplantation, comparing survival after LVAD placement with survival after transplantation will introduce immortal person-time bias if the treatment sequence is not considered and the immortal person-time for heart transplant patients while on LVAD is not appropriately allocated.

Immortal time bias can be corrected by using appropriate study designs and statistical analyses. The performances of different methods for handling immortal time bias were recently compared (53, 54). However, immortal person-time bias can be further complicated if treatment groups are not comparable in terms of disease severity and comorbidity burden. Therefore, how immortal time is handled can have a complex and significant impact on the estimates. Careful attention to the time from initial diagnosis to the sequence of treatments and their modalities, as well as use of an appropriate method for handling immortal time, is needed to avoid introducing this type of bias.

CONCLUSION

In this article, we have discussed some of the most pertinent methodological issues facing researchers conducting comparative effectiveness studies of medical devices; we have discussed special considerations regarding the use of data sources, exposure and outcome identification, exposure timing, and sources of bias. For the continued advancement of the field, we envision a UDI system implemented in administrative databases, registries, and EHRs enabling the assessment of product-specific performance in the real world. In addition, the development of methods for medical device epidemiology, built on current understanding of the similarities and differences between pharmacoepidemiology and medical device epidemiology, would greatly facilitate comparative effectiveness research for implantable medical devices.

ACKNOWLEDGMENTS

Author affiliations: Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts (Jessica J. Jalbert, Chih-Ying Chen); LA-SER Analytica, New York, New York (Jessica J. Jalbert); Procter & Gamble, Mason, Ohio (Mary Elizabeth Ritchey); Duke Clinical Research Institute, Durham, North Carolina (Bradley G. Hammill, Lesley H. Curtis, Xiaojuan Mi, Soko Setoguchi); and Department of Medicine, Duke University School of Medicine, Durham, North Carolina (Lesley H. Curtis, Soko Setoguchi).

S.S. was supported by midcareer development award K02HS017731 from the Agency for Healthcare Research and Quality, US Department of Health and Human Services. The project was also supported by contract HHS29020050016I from the Agency for Healthcare Research and Quality as part of the Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) Network and by contract HHSM50020100001I from the Centers for Medicare & Medicaid Services, US Department of Health and Human Services.

We thank Damon M. Seils for editorial assistance.

Dr. Mary Elizabeth Ritchey is a full-time employee of Procter & Gamble (Cincinnati, Ohio). Dr. Soko Setoguchi reported receiving research support from Johnson & Johnson (New Brunswick, New Jersey) and has made available online a detailed listing of financial disclosures (http://www.dcri.duke.edu/about-us/conflict-of-interest/). Dr. Lesley H. Curtis reported receiving grants from Medtronic, Inc. (Minneapolis, Minnesota) and Boston Scientific Corporation (Natick, Massachusetts) and receiving research and salary support from Johnson & Johnson, GlaxoSmithKline (Brentford, United Kingdom), the Agency for Healthcare Research and Quality (Rockville, Maryland), and the National Heart, Lung, and Blood Institute (Baltimore, Maryland). Dr. Curtis also has made a detailed listing of financial disclosures available online (http://www.dcri.org/research/coi.jsp).

REFERENCES


5. Healthcare Cost and Utilization Project (HCUP). Overview of the National (Nationwide) Inpatient Sample (NIS).


