Considerations in Using Registry Data to Support Pre-Market Applications of Medical Devices

Nelson Lu, Lilly Yue, Yunling Xu
Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD 20993

Abstract

In some situations, data from national/international registries may be utilized to support pre-market application of medical devices. In this paper, statistical and regulatory challenges are discussed on obtaining valid statistical inference utilizing registry data. Examples for potential use of registry data are presented.

Key Words: Registry, observational comparative study, propensity score, data quality

1. Introduction

Device registries have been used for many purposes such as short- and long-term surveillance, fulfillment of post-market observational study commitments for regulatory bodies, and comparative safety and effectiveness assessments [1]. Examples include Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS), Transcatheter Valve Therapy (TVT) Registry, Extracorporeal Life Support Organization (ELSO) Registry, and International Consortium of Orthopedic Registries (ICOR). Device registries, along with other types of medical registries, could also potentially play important roles in device pre-market evaluations. The matter that the registry data can be viewed as scientific evidence to support regulatory approvals of medical devices is solidified in the 21st Century Cures Act, which has passed the House and been received in the Senate at this time [2][3].

For the existing approved products, registry data could provide evidence for labeling update or labeling extension, such as expanded indications and expanded patient population. In addition, the registry data may be used to derive a performance goal or objective performance criteria for a certain type of devices. The information obtained from registry may generate hypothesis to be tested in future clinical confirmatory studies.

Another usage of registry data in the premarket setting is to form a control group (with patient level data available) from registries to serve as a compactor to the investigational device. Another potential usage is to embed trials within registries.

As there would be great opportunities to utilize registry data to support pre-market applications of medical devices, challenges do arise and careful considerations are needed. In section 2, we present general considerations. In section 3, we outline special considerations when control is formed from registry data. Several examples are provided in section 4. The article is concluded with summary in Section 5.

2. General Considerations

Device registries reflect the product use in the real world. A wider patient population is generally expected in a registry as opposed to a clinical study where inclusion/exclusion criteria are well defined. In addition, the skill levels and experience of the physicians are more varied in the real world setting rather than in a clinical trial one, as well as the levels of ancillary care. Patients are treated under less controlled environment in the real world as there are usually no clinical protocols and practices may vary among sites/physicians.

When it is intended to compare results of an IDE clinical study with results of control groups formed from a registry database, careful considerations are needed. An important question needs to be answered: How may clinical outcomes be affected due to the differences in natures between a registry and a well conducted clinical study? In some cases, the performance exhibited in a registry tends to be poorer than that exhibited in a clinical trial. Such situations may arise if, for examples, patients in clinical studies receive better care and medical attention. This would have important ramifications when the registry data are used as a comparative control, as the treatment effect of the investigational device relative to the control is likely to be overestimated. In some situations, it is possible that the clinical events of subjects in a registry are less likely to be detected than subjects in a clinical trial. Note that oftentimes these features may not even be captured or difficult to be captured in data sets, and therefore such a bias issue cannot be properly addressed via any statistical methodology.

A registry should possess some features in order to be utilized in the pre-market setting. Our discussion does not focus on regulatory issues such as informed consent, ethics, IRB approval, etc., but rather from the statistical and study design perspectives.

To obtain sound scientific evidence from a study, it is essential to have data with good quality. It is certainly true for registry data as well. The data quality of registries vary widely due to factors such as the purpose of the registry, carefulness in designing phase, various practical constraints, adequacy of site training, etc. To serve as a data source used to support pre-market applications, the data should have high quality.

Features to ensure good data quality in a well conducted clinical study should be applied in registry as well. Data need to be accurately collected and recorded. The registry study should be adequately monitored to maintain the data integrity. Relevant information, including all key baseline covariates and clinical outcomes, has to be collected in the registry. The missing data in baseline covariates and number of patients not completing the endpoint evaluation should be minimized. To reduce bias and variability of the study, mechanisms such as central clinical event committees and core labs should be utilized just like the way they are utilized in a clinical study [4].

Another important feature is the data accessibility to the FDA. The patient level data for baseline covariates and clinical outcomes should be available to the agency to perform the review, and audit system may be available to the agency.

3. Special Considerations

A major use of registry data in the pre-market regulatory setting is to form a control group, which serves as a comparator to the investigational device, which is the subject of an IDE clinical study. This section outlines some consideration for such cases.

A comparative conclusion could be reached with appropriate registry selection, careful study design and planning, and pertinent statistical methodology. In selection of a registry, points raised in section 2 need to be considered. The comparability between treatment groups is critical. In some cases, the treatment effect is believed to be confounded with time due to evolvement of medical technology, learning effect of device use, change in patient population, etc. In such a case, there should be sufficient overlap of timing between the enrollments of the registry and IDE study. As treatment effects may be expected to be different among regions, we recommend that the regions of a registry are matched with where the IDE study is conducted.

Once an appropriate registry is selected, good practice is needed to improve the objectivity of study design and interpretability of study results. Yue et. al. (2013) [4] advocates a two-stage study design process to address this issue. The practice outlined in that article is based on the principle of separation of design and analysis, which is proposed by Rubin [5][6][7]. With such a good practice, in conjunction with statistical methods such as propensity score methodology [8][9], the objective to conduct a less biased comparative study could be accomplished.

Among many points that need to be considered during the two-stage design process, two points are highlighted below.

The first regards the subject selection for the control group from the registry. In the selection process, it is critical that it should be done without accessing any clinical outcome data. Subjects may be selected based on the inclusion/exclusion criteria specified in investigational study, estimated propensity scores that are comparable to those of subjects in treatment group, and/or sufficient information on baseline covariates. It is possible that, in some cases, many registry subjects are poor matches to subjects in treatment group and thus not selected into the control group.

The second point is related to the sample size determination. Unlike a randomized controlled clinical study, the sample size estimation for the IDE study is not straightforward in the first design state due to uncertainties regarding the degree of comparability and sample size in control group. Our recommendation is to take a conservative approach to safeguard against unevenly distribution poor comparability and allows for greater flexibility in second stage design. This point is illustrated in Example 4 in the next section.

4. Examples

In this section, four examples are provided to illustrate important considerations presented in the previous sections. These examples are based on our review experience.

Example 1

Transcatheter aortic valve replacement (TAVR) is a minimally invasive catheter-based treatment for aortic stenosis patients who are deemed inoperable or at high risk during conventional surgical therapy. It was originally approved for only the transfemoral (via the leg) approach to TAVR for inoperable patients. To study alternative access for inoperable patients, clinical societies submitted a TVT registry-based IDE which was later approved by the FDA.

TVT registry has been created by the clinical societies, The Society of Thoracic Surgeons (STS) and the American College of Cardiology (ACC), in close collaboration with the Food and Drug Administration (FDA), the Center for Medicare and Medicaid Services (CMS), and the Duke Clinical Research Institute. The intent is to provide a data repository and reporting infrastructure to monitor the safety and effectiveness of TVT devices.

TVT registry possesses many nice features so that quality data can be collected to support the study objective and thus warrant the approval of the IDE. One of the purposes of the TVT registry was to allow gathering data to develop scientific evidence to support the expansion of the indications. Some elements that are essential for this example are described below. In the designing phase, particular attention was given to data elements and definitions that are harmonized with TAVR trials. All important baseline variables were identified and planned to be collected. As the data quality is a top priority, the means to monitor, audit, and adjudicate have been outlined. Extensive data quality program is in place to monitor data completeness and accuracy. Audit strategies were planned and can be executed by the FDA.

Detailed information regarding TVT registry can be found in [10].

Example 2

Results from a one-arm study with subjects using a new device were proposed to be compared with control group data formed from a registry. The primary endpoint is the 12-month event rate, and it was proposed to be evaluated using Kaplan-Meier method. The sponsor's original proposal to select the control subjects from the registry is described in the following. A subject is selected if (1) s/he meets the same inclusion/exclusion criteria of the investigational study, and (2) a primary endpoint event occurs to the subject before 12 months or the subject completes the 12-month evaluation. Based on this proposal, a lost-to-follow-up (LTF) registry patient is not a candidate for the control group.

The proposal of excluding LTF subjects from the registry data appears to be problematic as it may introduce bias. The control subjects should be selected solely based on their baseline characteristics, not the clinical outcome or follow-up. Registry subjects with comparable baseline characteristics to the subjects in the treatment group should be included even if no complete clinical outcome is available. The missing clinical outcome data for such subjects should be addressed by missing data analysis method, but not by excluding them from the data analysis.

Example 3

The performance of a new device was evaluated by comparing the clinical outcomes with a control group, which was selected from a registry database. The primary endpoint is 30-

day adverse event rate. Based on the protocol submitted in the IDE, a single arm study of 250 subjects treated with the investigational device was proposed.

A total of 500 control subjects from the registry were selected based on the same inclusion/exclusion criteria used in the IDE study. After the single-arm IDE study was finished, the 250 subjects were matched with 500 control subjects using the 1:1 matching method (without replacement) based on the logit of estimated propensity scores. The matching was performed with different caliper sizes. The sponsor proposed to use a caliper size of a 0.4 of the standard deviation of the logit of the estimated propensity scores. It resulted in 200 matched pairs. By doing so, a total of 50 treated subjects and 300 control subjects were discarded based on this proposal.

It appears that the control group does not provide good matches. In addition, two major concerns were raised by the FDA. First, there were no plans to establish a firewall to mask outcome, and no independent statistician was identified. As it was unknown whether the matching was performed with outcome data in sight, the objectivity of the study was in doubt. Second, it is problematic to discard subjects in the treatment group. This is because that it would be difficult to identify the intended population by leaving out some subjects in the treatment group based on their estimated propensity scores. Note that, however, it generally raises little concern to leave out some control subjects as such a practice does not alter the intended treated population.

Example 4

A sponsor intended to conduct a one-arm IDE study to assess the non-inferiority regarding the primary endpoint of a cardiovascular device to a control where the control group was extracted from a national registry. The primary endpoint is the treatment success rate at 12 months.

In the first design stage, the sponsor anticipated that 400 control subjects would be available. The sponsor expected that the success rate for both treatments would be 80%. With the 10% non-inferiority margin using a one-sided α of 0.025 of the Wald test, the sponsor proposed 300 subjects for the IDE study. This was derived by treating the study design as if it was a randomized controlled trial with greater power of 90% in attempt to compensate the potential imbalance in sample size distribution among strata.

After the enrollment of the IDE study was completed and all baseline covariates were collected, an independent statistician who had no access to the outcome data performed the propensity score modeling and design. The final agreed study design is based on sub-classification method. The sample size distribution of the design is displayed in Table 1. However, unfortunately, the re-calculated power based on this distribution became 74%, which was less than the desired level. This relatively awkward situation may be prevented if the sample size was calculated more conservatively in the first design stage.

Table 1: Sample Size Distribution Among Strata

Stratum	1	2	3	4	5	Total
N_1	20	38	55	78	109	300
N_0	120	102	85	62	31	400
Total	140	140	140	140	140	700

5. Summary

Although well designed and well conducted clinical trials generally generate good scientific evidence to support pre-market applications for medical devices, information gathering from international and national registries may provide same opportunities in some cases. Quality of the registry is a deciding factor for the suitability of the pre-market purpose.

For a registry being used to support a pre-market application, data must be demonstrated as accurate, robust, reliable, and complete. Too much missing data in clinical outcomes or baseline covariates would hinder such use. Ideally, the FDA should have the access to patient level data of a registry. For any future registry that is intended to support pre-market application, the FDA should get involved during the design stage of the registry.

Registry data can be used as a comparator to an investigational device, which may be assessed in an IDE clinical study. In this case, it is crucial that good practice is implemented. The objective is to select similar subjects in terms of patient characteristics from the registry to the subjects in the IDE study. Two stage design process should be utilized. The clinical outcome should be blinded during the process, in which typically propensity score methodology is utilized. Note that, although the subjects from registry may be discarded from the control group, the subjects in the treatment group (IDE study) should not be as this would make it difficult to identify the intended population. Also note that the sample size estimation in first stage design may not be straightforward due to various uncertainties.

In using registry data to support pre-market applications, there are certain limitations that may not be overcome. The data quality would not be as good as that of clinical study. Poor data quality hardly generates any sound scientific evidence. In the comparative study, it is possible that no appropriate control group can be formed as a comparator. Also, extensive design effort would be needed. Generally speaking, the evidence based on registries may not be as strong as that based on well designed and well conducted clinical studies.

References

[1] Registries for evaluating patient outcomes: a user's guide, third edition. Agency for Healthcare Research and Quality

[2] H.R.6-114th Congress (2015-106) – 21st Century Cures Act, https://www.congress.gov/bill/114th-congress/house-bill/6/text (accessed September 24, 2015).

[3] Gaffney, A. Regulatory explainer: The (updated) 21st Century Cures Act. http://www.raps.org/Regulatory-Focus/21st-Century-Cures-Act/ (accessed 24 September, 2015).

[4] Yue, Q.L., Lu, N., and Xu, Y. (2014). Designing premarket observational comparative studies using existing data as controls: challenges and opportunities. Journal of Biopharmaceutical Statistics 24:994-1010.

[5] Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to thetobacco litigation. Health Services & Outcomes Research Methodology 2:169–188.

[6] Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallel with the design of randomized trials. Statistics in Medicine 26:20–36.

[7] Rubin, D. B. (2008). For objective causal inference, design trumps analysis. Annals of Applied Statistics 2(3):808–840.

[8] Rosenbaum, P. R., Rubin D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 70(1):41–55.

[9] Rosenbaum, P. R., Rubin D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association 79:516–524.

[10] Carroll, J., Edwards, F., Marinac-Dabic, D., Brindis , R., Grover, F., Peterson, E., Tuzcu, E. M., Shahian, D., Rumsfeld, J, Shewan, C., Hewitt, K., Holmes, D. and Mack, M. (2013) The STS-ACC Transcatheter Valve Therapy National Registry: A New Partnership and Infrastructure for the Introduction and Surveillance of Medical Devices and Therapies. Journal of the American College of Cardiology. Vol. 62 No. 11 1026-34