Strategies for Postmarketing Surveillance of Drugs and Devices in Patients with ESRD Undergoing Dialysis

Moshe Vardi,* Robert W. Yeh,‡ Charles A. Herzog,§ Wolfgang C. Winkelmayer,§ Soko Setoguchi,§ and David M. Charytan*¶

Summary
The lack of evidence on the effectiveness and safety of interventions in chronic dialysis patients has been a subject of continuing criticism. New technologies are often introduced into the market without having specifically studied or even included patients with advanced kidney disease. Therefore, the need to generate valid effectiveness and safety data in this vulnerable subpopulation is of utmost importance. The US Food and Drug Administration has recently placed an increased focus on safety surveillance, and sponsors must now meet this additional postmarketing commitment. In patients with ESRD, the unique data collection environment in the United States allows for creative and efficient study designs to meet the needs of patients, providers, and sponsors. The purpose of this manuscript is to review the methodological and practical aspects of the different options for postmarketing study design in this field, with critical appraisal of their advantages and disadvantages.


Introduction
Drug and device approval in the United States is largely based on clinical trials sponsored by manufacturers. Data are usually confined to selected patients in a controlled setting with limited follow-up, and thus, they may fail to represent real life use. Consequently, the full range of an intervention’s effects is often revealed only after approval is granted by the Food and Drug Administration (FDA) (1). Postmarketing assessment, particularly assessment of safety, is, therefore, essential to optimize patient outcomes (2).

The Institute of Medicine concluded in 2006 that the science of drug safety needed major improvements (3). Subsequently, the FDA Amendments Act of 2007 Title IX, entitled “Enhanced Authorities Regarding Postmarket Safety of Drugs” (4), was established to ensure the FDA’s oversight of postapproval drug safety. The FDA Adverse Event Reporting System database is a voluntary reporting system for health care professionals, consumers, and manufacturers. It is used to evaluate postmarket safety, but the lack of consistent reporting and systematic data collection precludes its appropriateness as a sole measure to detect safety signals. Supplemental strategies include surveys of large databases of drug reaction reports or device-related events collected by government and private organizations (such as medical centers and insurance providers) (5) and the FDA Sentinel Initiative of active safety surveillance through access to nationwide electronic health care records. These passive and active strategies will likely change the face of postmarket surveillance (6) but should be supplementary to well designed postmarketing studies. Although such studies can be mandated by the FDA only under limited circumstances (7), they remain essential to corroborate premarket evidence of safety and effectiveness (8).

Patients with ESRD on dialysis constitute a unique population. These patients experience high morbidity and mortality (9). Furthermore, because patients on dialysis are poorly represented in randomized trials and may have altered drug excretion (10), the optimal drug dosage in the setting of ESRD is frequently unknown at the time of approval. Dialysis-related care is also unique. Care is delivered through a small number of large dialysis organizations (LDOs), the majority of care is paid for by Medicare, and data are collected by a universal registry, the US Renal Data System (USRDS). These features provide unique opportunities for studying postmarketing safety of drugs (including peritoneal dialysis solutions) and devices in ESRD. The objective of this review is to discuss potential postmarketing study designs for medications and devices in this distinctive setting.

Randomized Clinical Trials
Randomized clinical trials (RCTs) are the gold standard in terms of internal validity. Well designed RCTs should control for confounding, and their outcomes should be representative of the efficacy and safety of the intervention compared with the control within the studied population (11). Conounding is reduced, because the study design distributes both measured and unmeasured factors across the treatment arms in a balanced fashion (12,13). Furthermore, because RCTs typically require the reporting of adverse events occurring during the trial, primary and secondary outcomes are collected using uniform definitions and data capture methods. This process allows for adjudication of important outcomes in a blinded and objective fashion,
interpreting the treatment effect. In addition, RCTs may take steps to maximize compliance with treatment and follow-up, thus ensuring as close as possible a causal interpretation of treatment effect.

However, these advantages are offset by the use of selective enrollment criteria, small sample sizes powered to detect differences in study end points (often composite), and short follow-up duration. As a result, RCTs rarely represent the diversity of patients seen in practice, and their results may not be generalizable (14,15). Informed consent, protocol-related restrictions on medical care (with respect to intervention, control, or concomitant therapies), and enhanced surveillance of outcomes may also result in subtle changes in care that alter outcomes compared with unstudied populations (16–18). Lastly, rare events or small increases in the relative risk of events with high background rates in ESRD (e.g., mortality or myocardial infarction [MI]) that potentially translate into large increases in absolute risk may not be captured because of limited sample sizes. In summary, although RCTs enhance internal validity, they may have limited generalizability to the more diverse and complex clinical populations of interest for assessment of real world safety (13).

Given the advanced age and high degree of comorbidity in the routine dialysis population, findings of safety within a highly selected (and potentially healthier) RCT population may be falsely reassuring. In addition, RCTs are burdensome in their execution and require high-level coordination between clinical staff and study personnel to gather data and administer assigned therapies. The high cost and relatively small market size of the dialysis population may provide disincentives to performing an RCT of adequate size and power to rule out low-frequency safety events not seen in the premarketing experience.

Use of RCTs for studying postmarketing safety in ESRD is relatively common. Recent examples include trials of sildenail, iron gluconate, and cinacalcet (19–21). Notably, each of these studies was small (<300 patients), with insufficient power to detect uncommon events. Large RCTs, with a primary goal of assessing postmarketing safety, are less common in ESRD. However, important postmarketing safety data have been generated in phase IV RCTs motivated by other considerations. The Normal Hematocrit trial (22), for example, identified the potential mortality risks of exogenous erythropoietin, whereas the Evaluation of Cinacalcet HCl Therapy to Lower Cardiovascular Events (EOLVE; cinacalcet) (23) and Dialysis Clinical Outcomes Revisited (DCOR; sevelamer) (24) trials each greatly expanded the controlled experience with the tested therapy compared with preapproval trials.

Cluster Randomization
Cluster randomization, in which the unit of randomization is the study site or group of subjects rather than an individual subject, is increasingly used in medical research (25,26). When both the control and the intervention of interest are within the standard of care, requirements for individual consent and adverse event reporting may be relaxed. Furthermore, the delivery of identical care within each cluster may simplify logistics (for example, by eliminating the need for storage and administration of both the control and experimental medications at a given site). When the correlation between patient characteristics and the cluster is low, the design should mirror the equalization of measured and unmeasured factors of standard RCTs (27–30). However, the effective sample size of a cluster RCT depends on the strength of intraclass correlation among patients, and the power of such a design is lower (31).

Because interventions and delivery of care are typically standardized within dialysis units and because the majority of units in the United States belong to just a few LDOs, cluster randomization is appealing. This design, for example, might be ideally suited for the comparison of vitamin D or intravenous iron preparations, because units typically stock a single agent; also, the choice of preparation is made at the unit level with minimal regard to patient characteristics or the treating physician’s preference. Cluster randomization in this setting might be feasible with little departure from the usual standard of care and minimal effect of trial participation on outcomes compared with an RCT.

Despite these advantages, experience with this approach seems to be limited. A MEDLINE search using the terms cluster and hemodialysis with a limit of clinical trials, for example, returned only 147 citations, and none of the citations were a postmarketing safety study. Several factors may explain the apparent lack of enthusiasm. First, because a cluster design requires randomization at the level of the individual unit, it may require a higher level of LDO cooperation than for a typical RCT. Second, patterns of practice (external to the tested intervention) and commonality of physician and patient characteristics may cluster at the LDO and unit levels. These factors can threaten the supposition that unit-level randomization equalizes distribution of measured and unmeasured confounders and undermines study validity. Stratifying the cluster randomization by factors known to be highly correlated at the unit level (e.g., a dialysis unit’s standardized mortality rate) can reduce these imbalances, but areas of correlation may not be obvious. A clustered approach may also be less desirable when one intervention is not within the standard of care or when there are significant safety concerns. In this case, the requirement for individual consent may mitigate some advantages of a clustered approach. Finally, because sample size is a driver of costs, cluster RCTs may represent an undesirable tradeoff between cost and logistical efficiency relative to standard RCTs.

Nonrandomized Designs
Although nephrology has been criticized for a lack of RCTs (32), when long-term safety is the primary focus, nonrandomized designs offer known advantages. Generalizability is enhanced, because data are generated from real world patients in routine clinical settings, and requirements for monitoring, adjudication, and reporting of adverse events are less stringent. Observational studies, thus, tend to provide cost savings compared with RCTs. Although nonrandomized approaches suffer from a greater likelihood of bias and require careful design and analytic methods to control for confounding, they have clear advantages in terms of statistical power and external validity—the ability to generalize their findings to the types of patients encountered in clinical practice.
Observational Studies with Primary Data Collection

Prospective cohorts have been used to study the safety of spironolactone in adult hemodialysis patients and iron gluconate in the pediatric dialysis population (33,34). Whether observational or interventional, a prospective approach can facilitate the collection of all confounding variables of interest. Prospective data collection also allows standardized adjudication of outcomes and uniform definition and measurement of variables across sites, thereby enhancing data consistency. This approach allows for studying factors outside of routine clinical parameters either in terms of the collection frequency or because they are not typically recorded in existing databases. Finally, a prospective approach allows for collection of biologic samples needed to assess safety (e.g., the development of antibodies to new biologics).

Prospective observational studies may be particularly suitable in ESRD. Delivery of care to large numbers of patients at a single site allows for time-efficient data capture. Because many parameters of interest (BP and potassium) are routinely recorded, new measurements taken solely for study purposes may be unnecessary. Additionally, dialysis unit staff may be leveraged to administer testing that would otherwise require dedicated research interactions. However, a prospective observational study is still likely to be costly and logistically complex. In particular, compliance with data collection schedules may be burdensome or require significant reimbursement to the LDO or dialysis units. Data collection by staff not specifically trained for or dedicated to the research also mandates investment in education, postcollection quality control, and edit-checking processes to ensure data accuracy. These factors may partly offset the advantages of this approach. In addition, a requirement for consent may affect generalizability if there are systemic differences between patients who consent and patients who refuse enrollment (35).

Observational Studies Using Existing Databases

Existing databases may capture large population-based cohorts that are representative of the actual usage in clinical practice. A database study may, thus, be the only means of ruling out small but clinically important differences in the safety of two treatments, and it may be the only means of evaluating rare outcomes or delayed effects. Several existing databases make this approach particularly feasible in dialysis. These databases include the USRDS, the proprietary databases of the LDOs, and on a smaller scale, the Dialysis Outcomes and Practice Patterns Study (DOPPS) or Employer Group Health Plans (EGHP) for privately insured patients.

The USRDS captures almost all US dialysis patients and records baseline patient characteristics and comorbidities as reported by clinicians at the initiation of dialysis. For patients with Medicare fee for service, baseline conditions of interest and subsequent outcomes can be extracted from diagnostic and procedural codes. Although dialysis-related intravenous medication use has been available from administrative claims, the advent of Part D Medicare in 2006 has made assessment of oral medications feasible.

A disadvantage of the USRDS has been a lag time of up to 2 years to the release of cleaned data for public use. This lag time might lead to unacceptable delays in studying the postmarketing safety of new drugs after their introduction. For patients who are not eligible for Medicare before the onset of ESRD (i.e., nondisabled individuals younger than 65 years) or patients enrolled in Medicare Advantage, data on outcomes in the USRDS are incomplete. Thus, if incident patients are of particular interest, the USRDS cannot be used to study safety. Because only limited laboratory data are available in the USRDS, it is also of limited use in studying laboratory markers of safety other than parameters of anemia and mineral and bone disease. However, the granularity and real-time availability of some data elements may be enhanced after Consolidated Renal Operations in a Web-Enabled Network—which enables electronic submission of patient and clinical performance measure data to the Centers for Medicare and Medicaid Services (CMS)—is fully functional (36). To our knowledge, no published data have used the USRDS to specifically study the postmarketing safety of drugs in ESRD. However, a recent study that used the Surveillance Epidemiology and End Results Medicare database to compare carboplatin and bevacizumab therapy in nonsmall cell lung cancer has nicely shown the power of combining Medicare and disease-specific registry data for postmarketing evaluation (37).

If incident patients or laboratory markers are of interest, EGHP data may be useful. EGHP data are available through Thomson Reuters MarketScan Commercial Claims and Encounters (38) and Invентiv Health clinical databases (formerly Ingenix i3) (39). The size of these databases may enable the capture of a significant sample of incident dialysis patients, with reliable data on comorbidities and outcomes.

LDO electronic records represent another resource. In addition to patient demographics and comorbidities, LDO data reflect the information required to manage clinical practice and justify billing. Thus, these data frequently contain information not collected by USRDS or EGHP sources, such as dialysis prescriptions, oral medications, details of the dialysis procedure, vital signs, and laboratory findings.

Although some variables within these datasets have been validated (40), definitions, consistency in recording, quality control, and data cleaning procedures may vary from unit to unit within an LDO and between LDOs. Also, LDO datasets reflect a single dialysis organization, which may not be representative of other organizations. Data from multiple LDOs may be required if LDO-specific practices could affect a product’s safety profile. Lastly, LDO data (like EGHP data) are proprietary and may not be readily available without a high level of cooperation, or they may only be available at considerable cost.

The DOPPS database offers an alternative existing resource. DOPPS offers several advantages, including capture of international experiences and detailed medication, laboratory, and dialysis prescription data for a subset of patients (41). In contrast to LDO data, DOPPS is gathered in a uniform fashion using standard definitions, whereas comorbidities are obtained from direct patient contact rather than diagnostic codes like in the USRDS. The smaller sample size of DOPPS relative to USRDS and LDO data, however, limits power for assessing rare events or small increases in risk.

The LDO, DOPPS, and USRDS offer several important advantages. In particular, the cost of data collection has already been incurred. Using these sources may be the only feasible means of studying rare outcomes or small clinically
important differences in safety and effectiveness, because they are relatively inexpensive to use; also, they facilitate the study of larger cohorts than may be financially or logistically feasible in an RCT or prospective study where a sponsor has to bear the full cost of data capture. For example, analysis of Fresenius data has highlighted potential increases in mortality with digoxin, warfarin, and calcitriol (42–44).

Data reliability is an important concern. Although baseline comorbidities are represented in the USRDS (taken from the Medical Evidence Report form CMS-2728), this form has significant deficiencies and inaccuracies (9,45,46). Furthermore, several variables, such as history of MI or coronary revascularization, that are critical for adequate control of confounding are not currently collected on form CMS-2728. In addition, baseline data are not routinely updated and reflect only those comorbidities present at the time of dialysis initiation. Less information is available on the specific data points collected by LDOs, but the collected data may share many of the deficiencies of the USRDS. It is possible to supplement USRDS and LDO data by collecting comorbidity data from Medicare or EGHP codes used during a specified baseline period. However, this method requires enrollment for an adequate length of time before study initiation to allow for the generation of health care claims on existing conditions.

Outcomes ascertainment in the USRDS is based on diagnostic and procedural codes from Medicare claims. Although diabetes mellitus-, hypertension-, and epilepsy-coded diagnoses have been validated (47–49), the sensitivity and specificity for many other outcomes are unknown. For example, algorithms for common conditions, such as heart failure, MI, and stroke, are imperfect (50,51).

LDO outcomes data are most frequently generated from hospital discharge summaries by the nursing or administrative staff at local dialysis units. Because training for this collection is not controlled and standard definitions are not used, LDO outcomes data are likely to be inconsistently recorded. Data quality may be further compromised by the failure to obtain discharge summaries or include data on all outcomes of interest within those summaries. Finally, outpatient and emergency visits not resulting in a missed dialysis session may not generate transmittal of information to the local dialysis unit or recording of events within LDO datasets. Given these issues, differential classification of events across study sites has the potential to bias LDO-based study results.

Table 1 summarizes data sources for specific ESRD subpopulations. If use of an existing database is adopted and institutional review board approval is obtained, retrospective chart analysis can be used to refine data reliability (52). However, retrospective adjudication is logistically challenging in hemodialysis patients because of the complex nature of their comorbidities and the large numbers of events. Linkage of USRDS and LDO data with each other or prospectively collected data may allow for supplementation of missing data elements at a lower cost than a fully prospective approach, but data privacy concerns and differences in data definitions across datasets present significant challenges.

Finally, an observational design remains more susceptible to confounding and selection bias than a clustered or standard RCT. It is critical to use rigorous statistical and epidemiologic techniques (e.g., stratification, propensity matching, and multivariable modeling) to assure comparability of the treatment groups and adjustment for important confounding factors. For example, in addition to patient-specific demographics and comorbidities, it may be important to adjust for or stratify by geographical location (rural versus urban), clinic type (hospital versus freestanding), and the standardized mortality rate.

Combination Approaches

The dialysis setting offers the unique possibility of combining any of the above study designs. In theory, LDO or USRDS databases could provide the baseline and outcome data in a trial that randomizes patients to treatment or control and prospectively collects a limited number of variables of interest, such as blood samples. This approach has been used successfully to simplify and extend outcomes assessment in the DCOR study (53). Although this method involves tradeoffs, it may offer cost efficiencies that facilitate a larger study with improved power. For low-risk or standard of care interventions, one might envision a process, where LDO staff or information systems are leveraged to simplify randomization or automate data collection.

Applications to Device Safety

The aforementioned data sources can be potentially used to study device safety. However, detailed information about devices (manufacturer, model, and even type) may not be available in many data sources. For example, USRDS data may differentiate the types of coronary stents (drug eluting versus bare metal) but not manufacturers of the stent or number of stents implanted. Likewise, data on dialysis-specific devices may be recorded in LDO data. Assessing safety of specific types of devices by manufacturer or model may be feasible only with prospective observational or randomized studies given the lack of the standardized system to uniquely identify devices. Linkage with the American College of Cardiology’s National Cardiovascular Data Registries, however, may allow for study of outcomes with specific cardiovascular devices in ESRD (e.g., implantable defibrillators). Alternatively, because some implantable devices are tracked by industry for product safety purposes, linkage of manufacturer’s proprietary databases with LDO or USRDS data, if feasible done despite proprietary and data privacy concerns, could enable more granular assessment of safety in specific devices. Although similar methodological issues apply to the assessment of device safety, additional considerations on the role of operator’s proficiency and perioperative management are needed to understand the safety of devices.

Summary

This review details the benefits and disadvantages associated with alternative approaches to postmarketing analysis of safety in ESRD patients on dialysis, which are summarized in Table 2. When considering the best methodological approach for postmarketing studies in the dialysis population, one must take into consideration the patient population (hemodialysis or peritoneal dialysis
and incident or prevalent patients), intervention (oral or intravenous medications or devices), and relevant outcomes of interest (mortality, cardiovascular outcomes, and rare events). Studies assessing an intervention in an incident pediatric population in which the outcome of interest is known to be weakly represented in the coding
systems, for example, require different approaches than studies assessing an intervention in elderly patients in which outcomes are well captured. Although a well-designed RCT or prospective cohort study is likely to yield valid results, the unique nature of dialysis care with the centralization of care, the availability of data from existing and generalizable databases, and the ability to compensate for deficiencies through complementary data sources makes database approaches attractive. A hybrid approach using more than one methodology, thereby compensating for their weaknesses, can also be useful.

The recent collaboration between the American Society of Nephrology and the FDA—the Kidney Health Initiative—emerged with the purpose of enhancing the study of kidney health and patient safety (http://www.asn-online.org/khi/) (54). Its objectives include the development of approaches to the systematic collection of retrospective or prospective data, the establishment of data standards, and the development of innovative and efficient trial designs. These initiatives and others are likely to enable additional opportunities for novel approaches in trial design and conduct in this unique clinical environment.

Acknowledgments

We acknowledge help from Sandra Tong, Dan Cooper, and Krishna Polu of Affymax Ltd. (Palo Alto, CA) in preparation of this manuscript.

Disclosures

The authors consulted for the Harvard Clinical Research Institute and/or the Chronic Disease Research Group on the design of a postmarketing study for Affymax (Palo Alto, CA). In addition, the following individual disclosures are listed. M.V. and R.W.Y. have no disclosures. C.A.H. has consulted for Abbott, Affymax, Amgen, Fibrogen, Johnson & Johnson, Medtronic, and Merck; provided research support for Affymax, Amgen, and Zoll; received royalties from UpToDate; has had equity interest Boston Scientific, Cambridge Heart, Johnson & Johnson, and Merck; served on the safety advisory board for Keryx; was a member of the Board of Trustees for Roche Foundation for Anemia Research; and provided research support for the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases. W.C.W. reports having served as a scientific advisor or member of adverse event adjudication committees for Amgen, Bayer, GlaxoSmithKline, Keryx, and Medtronic. S.S. reports receiving research support from Johnson & Johnson and honoraria for consulting from Affymax and Novartis. D.M.C has consulted for or contributed clinical trial design to Medtronic, Tengion, Keryx, and Affymax and has been on clinical events committees for PLC Medical.

References


Table 2. Summary table

<table>
<thead>
<tr>
<th>Variable</th>
<th>RCT</th>
<th>Observational with Existing Database*</th>
<th>Observational with Primary Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness of data(^b)</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>Generalizability</td>
<td>+</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>Data quality</td>
<td>+++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Potential for bias</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Logistical complexity</td>
<td>+++</td>
<td>$</td>
<td>$$</td>
</tr>
<tr>
<td>Costs</td>
<td>$$$</td>
<td>$</td>
<td>$$</td>
</tr>
</tbody>
</table>

A scale of + to ++++, where + is the lowest and ++++ is the highest, or a scale of $ to $$$$, where $ is the cheapest and $$$ is the most expensive, is used. RCT, randomized clinical trial.

*Dialysis Outcomes, Practice and Patterns Survey (DOPPS) data may be considered an existing database in most respects.

\(^b\)Timeliness of data may depend on when a postmarketing study is commenced. If a study is initiated immediately after drug/intervention reaches the market, it may take several years until a sufficient experience with that intervention is captured within large registries, such as the US Renal Data System, and available for analysis. In this case, dedicated data capture using a prospective approach (e.g., an RCT) may be quicker.


49. Quan H, Khan N, Hemmelgarn BR, Tu K, Chen G, Campbell N, Hill MD, Ghali WA, McAlister FA; Hypertension Outcome and Surveillance Team of the Canadian Hypertension Education Programs: Validation of a case definition to define hypertension using administrative data. Hypertension 54: 1423–1428, 2009


Published online ahead of print. Publication date available at www.cjasn.org.