In Data We Trust: The Role and Utility of Dialysis Provider Databases in the Policy Process

Mahesh Krishnan,[†] Helen M. Wilfehrt,* and Eduardo Lacson Jr.*[‡]*

Clin J Am Soc Nephrol 7: 1891–1896, 2012. doi: 10.2215/CJN.03220312

*Medical Informatics and [†]Medical Office, DaVita Clinical Research, Minneapolis, Minnesota; and [‡]Fresenius Medical Care, Waltham, Massachusetts

Correspondence: Dr. Mahesh Krishnan, DaVita Clinical Research, 825 South 8th Street, Suite 300, Minneapolis, MN 55404-1208. Email: mahesh.krishnan@davita.com

Introduction

The American dialysis setting represents one of the most data-dense disease settings in the world (1). With the hope of better understanding ESRD, its treatment, and its associated costs, data are collected through public databases, registries, and proprietary databases. Although no one data source provides a complete picture, data help inform health policy, with different types of data providing different insights. In this review, we discuss the different data sources in the ESRD treatment setting, with a particular focus on the data coming from large dialysis organizations (LDOs) and their dialysis information systems (DISs). We discuss the benefits and limitations of these DISs and how they help inform healthcare policy.

Sources of Data

Various sources of data in the nephrology setting are described below and in Table 1. The major public source of data comes from Medicare claims and data collection forms under the auspices of the Centers for Medicare and Medicaid Services (CMS). In the United States, CMS administers the payment for ESRD treatment under authority of an amendment to Public Law 92–603 enacted in 1972. The US Renal Data System (USRDS), an organization funded by the National Institute of Diabetes and Digestive and Kidney Diseases, uses CMS claims data to maintain a national data registry of the ESRD population (2). The USRDS, in turn, conducts insightful analytics on disease trends, treatments, and outcomes, reporting them in an annual data report (2). The USRDS Coordinating Center is operated by the Minneapolis Medical Research Foundation in Minneapolis, Minnesota. Because of the nature of claims data, there is generally a 2-year lag, challenging the ability to track the impact of treatment or policy changes in real time. Also, CMS contracts with other entities to process its data and produce reports. The University of Michigan Kidney Epidemiology and Cost Center (3) is one of these entities, and it produces Dialysis Facility Reports that describe summary process and outcome data at the dialysis facility level (4).

Other data sources also describe or report on ESRD patients and conditions. These sources include different registries, such as registries for transplantation (*i.e.*, the United Network for Organ Sharing registry) (5) or the quotidian dialysis regimen (6). Proprietary databases also exist, such as the databases assembled by Arbor Research Collaborative for Health to inform analyses for the Dialysis Outcomes and Practice Patterns Study (7). Although hospitals, suppliers, physicians, and other providers contribute their share, much of the data represented in these sources is collected from providers of dialysis care. With consolidation of individual dialysis facilities into LDOs, large-provider databases become important sources of analytics that can inform scientific and policy-related inquiries.

Dialysis Provider Information Systems and Databases

Many large dialysis providers use DISs that apply a unifying set of business rules and procedures to facilitate standardized data collection across a large number of dialysis facilities, setting the stage for a system of dialysis care (Figure 1 and Table 1). This standardization is further enhanced by the use of a limited number of central reference laboratories, resulting in relatively small, manageable differences in most measured biochemical values within a given system of dialysis care, thereby making analytics between facilities possible. Within a system of dialysis care, the DISs operate across various theaters of treatment, such as vascular access centers, oral medication pharmacies, and others. Data for a given patient are tracked by a unique identifier that allows an individual patient's care to be followed across multiple dialysis units within the same system and across the theaters of treatment within the system. Together, these features can potentially allow for a longitudinal picture of an individual patient's clinical course to be studied more comprehensively.

Data collected from the DIS are stored in central databases with an update lag of 1–2 days for most variables, allowing for those individuals tasked with ensuring and improving quality the ability to perform real-time analytics. This ability, combined with a robust continuous quality improvement process that allows for both active surveillance and outlier identification, is currently being used by providers (albeit still at a rudimentary level) as an

Table 1. Summary of select data sources in the nephrology setting

Data Source	Description	Limitations
Centers for Medicare and Medicaid Services (CMS)	Claims data collected under amendment to Public Law 92–603; additional data collection limited (*e.g.*, Form CMS-2728)	Limited clinical data (actual treatments) and delay in availability (currently 2 years behind)
US Renal Data System (2) (http://www.usrds.org/)	Funded by the National Institute of Diabetes and Digestive and Kidney Diseases; laims, epidemiologic, and clinical-based, stand-alone database (diagnoses, demography, biochemical data, claims, treatment and payor histories, hospitalization events, physician/supplier services, and providers)	Delay in availability (currently 2 years behind)
University of Michigan Kidney Epidemiology and Cost Center (3)	This group produces Dialysis Facility Reports under contract with CMS and provides them to individual facilities; each facility is provided information on their patients, treatment patterns, transplantation patterns, hospitalization, and mortality rates compared with local and national averages	Predominantly based on claims data, limited clinical data (actual treatments), and delay in availability (currently ~1 year behind)
Individual insurers	Claims data collected based on insurance claims	Limited clinical data (actual treatments) and limited availability as patients transfer between insurers
United Network for Organ Sharing Registry (5) (http://www.unos.org/)	Registry of patients awaiting solid organ transplants, with limited data available after transplantation	Limited clinical data (*i.e.*, actual treatments, cormidities, and outcomes)
Quotidien Dialysis Registry (6) (http://www.quotidien.org)	Registry originally designed to collect data about outcomes on intensive hemodialysis prescriptions (>5 times/wk); now, it also includes data on other alternative hemodialysis regimens	Limited to select forms of dialysis
Dialysis information systems from large dialysis organizations	Databases created using facility-based clinical systems that collect detailed data about what orders and treatments are given within a dialysis facility, including detailed laboratory test data and prescriptions	Limited details on comorbidities, limited information about what happens outside the clinic (*e.g.*, hospitals), limited information about nuances of care, and living datasets subject to availability at the time of data pull

opportunity to improve the lives of patients that we serve. In addition, the databases often function as *de facto* registries that are able to provide information not only on individual patients but also status and operational reporting for aggregate patient populations.

Access to full data in these provider databases is generally limited to internal investigators who perform analytics related to quality of care and performance improvement activities. In some instances, external academic collaborators, industry researchers, or clinical research organizations are contractually provided access to deidentified granular data.

Comparative Advantages of DISs over Other Sources

General

DISs from LDOs with their granular electronic medical records have significant advantages over claims-based data sources. First, real-time collection and subsequent analysis of electronic medical record data allows the data to be available more rapidly compared with the traditional 2-year lag associated with claims-based datasets, such as USRDS. Second, data from the DISs include data from all payor sources, allowing for analytics in patients beyond the Medicare beneficiary population, the latter of which forms the basis of the USRDS dataset. Third, DISs from the LDOs provide the capacity to analyze and evaluate incident patients who present to the dialysis clinic. In contrast, the USRDS dataset contains limited information about incident patients because of the coordination of benefits period associated with the transition between private and Medicare coverage. The availability of data during this period is particularly important in light of recent findings, which have focused on the increased mortality and morbidity associated with the first 90–120 days of initiating dialysis (8).

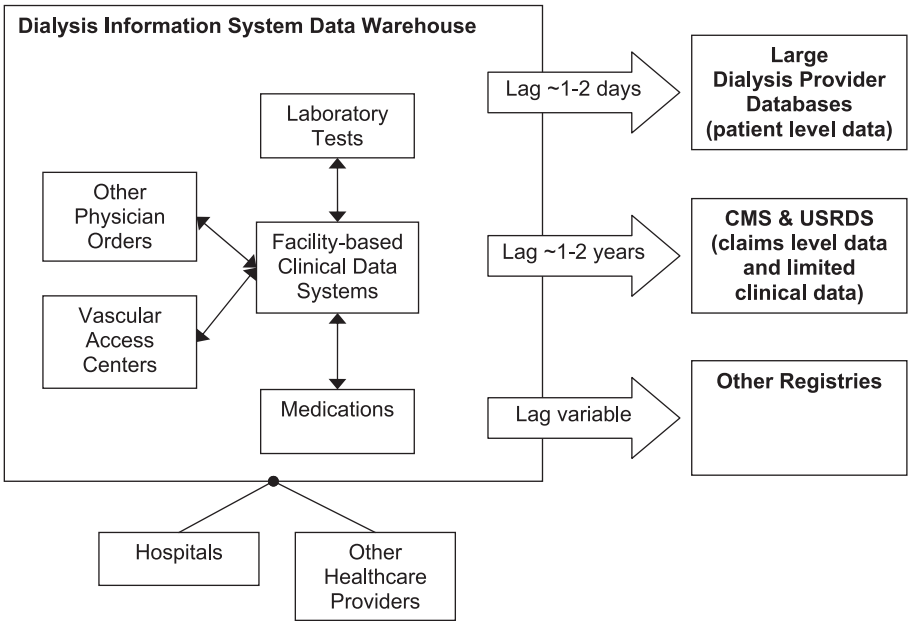

Figure 1. | **Data flow in the nephrology setting.** High-level description of the dialysis information system within large dialysis organizations and the interactions with external databases.

Laboratory Biomarkers

Traditionally, claims data have provided limited information on patient-level biochemical laboratory values—only monthly hematocrit and urea reduction ratio values are collected as part of the claims process. Therefore, claims data provide limited insight into the patient-level clinical findings that drive the treatment decisions of the individual nephrologist and healthcare team. To remedy this shortcoming, *V* codes recently have been added to the claims form, allowing the collection of ancillary data such as vascular access types and Kt/V but only as categorical values defined by numerical ranges. Although this addition addresses the needs for surveillance and quality improvement efforts, it limits the use of these fields for research analytics. In contrast, provider DIS data has nearly all patient-level biochemical laboratory values—values obtained from blood samples drawn or submitted by the dialysis facility from a patient.

Clinical Biomarkers and Therapies

In these provider databases, clinical indices (*e.g.*, weight or blood pressure), corresponding treatment prescriptions, and intravenous medication orders are tracked. Plus, these databases include quantitative data on actual use of agents contained within the composite rate, such as heparin administration during dialysis. Combining granular laboratory data and concomitant medication information is particularly useful in certain situations (*e.g.*, assessing the use of different vaccine preparations associated with specific immunologic response results) (9). The DISs also allow quality control efforts to relate prescribed with delivered treatments, the understanding of which is critically important to determine and evaluate factors related to patient compliance, such as treatment time.

Vascular Access and Quality of Life

Other data elements that can be tracked with great fidelity in the DISs include vascular access use and conversion as well as some data from several quality of life instruments. The SF-36 data have been evaluated previously (10), although as a result of the recent changes to the conditions of coverage, the Kidney Disease Quality of Life-36 survey is now commonly being offered to dialysis patients, with their responses being entered into the clinical systems. Therefore, the DISs provide the capacity to draw associations between patient-reported quality of life outcomes and biochemical laboratory values or other medical events of interest.

Utility and Impact on Policy

With recent changes in conditions for coverage along with future changes when more oral agents are added to the bundle beginning in 2014, provider databases continue to improve their ability to track some oral medications. ESRD patients take many medications. Being able to track them becomes extremely important from both a medication adherence and management perspective. The accuracy of such medication data is being enhanced by the incorporation of dispensing and fill data normally associated with pharmacy benefits management entities. Such data potentially allow for the calculation of measures of patient compliance to medication regimen—albeit more work is needed to ensure that these programs are fully validated.

The benefits of having granular patient data across so many settings of care for so many patients are many. First, these longitudinal data captures allow for granular patient-level analytics. Second, the groupings of dialysis patients into discrete treating entities allows for facility-level analytics. Analytical strategies may provide insights or define clinically relevant problems or areas for potential interventions, with granular data informing changing

conditions, treatment patterns, and treatment responses over time. They can inform hypotheses and clinical trial design strategies for etiological and therapeutic research, highlighting confounding factors. They may also be used to determine incidence or prevalence of certain conditions or perform surveillance, such as in blood culture results for increased rates of resistant organisms like methicillin-resistant *Staphylococcus aureus* or emergence of Vancomycin-resistant strains. They may also be used to inform national projections, models, and estimates for outcomes, health services use, and health economics research (11).

Facility-level analytics, along with the ability of using the extensive data to create propensity score-matched patient cohorts and available longitudinal data, provide individuals in dialytic nephrology research with the unique ability to address certain biases, such as confounding-by-indication bias or immortal-time bias, both of which are prevalent in retrospective data analyses. Thus, the collective information creates the ability to link national/local policies or practices and procedures to short- and long-term changes in defined process and outcome measures. These analytics may also provide useful information for determining quality metrics and estimating the potential impact of system-level interventions (12).

Given the national distribution of the data and the large sample sizes in provider databases, analytics performed on them often reflect a representative sample of the US dialysis population. Therefore, analytics aimed at answering critical questions that are of national relevance, when performed using these large-provider databases, reflect real world practice—they provide generalizability that is not always obtainable from analytics using unique academic center experiences, claims-based analyses, and/or narrowly focused randomized clinical trials. As a result, provider data represent a robust platform to define the scope of a given issue, allow the identification of best practices, and evaluate the comparability, feasibility, and/or scalability of one or more potential solutions. Having this clinical insight into real-world practices on a national level informs policy decisions (13).

Limitations of Provider Databases

Nevertheless, despite the strengths of large-provider databases, they do have limitations (Table 1). One advantage that the USRDS data have over them is the capacity to collate and then evaluate comorbid illnesses derived from not only the dialysis provider claims but the larger universe of Medicare claims, including hospitals. Provider databases tend to have less comprehensive comorbidity data than the information obtainable from Medicare files. In fact, large dialysis providers had limited success when they tried to improve capture of comorbidity data because of the need to document these data as case mix adjusters for reimbursement. Dialysis providers have mounted significant programs to capture the case mix burden predicted by the Medicare data at a facility level but have not been able to find even 50% of case mix adjusters predicted by the CMS (14).

Another limitation is the lack of transparency in the clinical events that occur external to the outpatient dialysis facility, such as treatment during hospitalization. For example, surveillance for events, such as central line-associated bloodstream infections or transfusions in ESRD patients, may be fragmented if there is a lack of coordinated documentation between dialysis providers and hospitals. Inasmuch as the integration of various aspects of laboratory, vascular access, and pharmacy-type data within the LDOs are all in varying stages of completion, more work needs to be done with integration and data sharing with hospital-based systems, physician practices, skilled nursing facilities, and other silos of patient care. This work is particularly important, because what happens in dialysis units is a crucial component of the coordination of care required to impact hospitalization and readmissions to the hospitals in ESRD patients. Such integration could be done by direct data transfer from hospitals or access by providers to the Medicare common working file, which contains all claims data on a given patient with an acceptable delay. Thus, the provider DIS and databases provide an opportunity to fulfill the promise of health information exchange (15–17).

Although the data from LDOs provide extensive granularity for individual, facility-based, and population-based analytics, there are nuances in patient care that are currently not being captured in the databases and that require critical and conservative interpretation of the findings. For example, although not unique to these large databases, many historical analyses of intradialytic changes in blood pressure were based solely on the difference between baseline readings taken before initiating hemodialysis treatments and final readings taken after the treatment had been completed—missing the up and down changes that were measured in between; however, this limitation and the implications on interventions may not be apparent to the reader (18). The data may require additional internal validation and review, particularly if the analytical results are sensitive to the potential range of expected values of the specific variable. For example, dialysate sodium will need to be clarified as to whether it was based on a prescription, a laboratory-measured dialysate sodium, the number provided by the dialysis machine from conductivity measurements, or some other method of determination to allow for accurate interpretation. Furthermore, attention to the completeness of ascertainment for particular data fields is necessary. In addition, when key variables that may be major prognostic or deterministic factors are not available for analysis, it is appropriate to acknowledge limitations related to residual confounding.

Finally, an often unrecognized but clearly significant limitation of databases from DIS is the fact that they are living records— in essence, changing day to day as clinical information becomes available. In some instances, sufficient time for completing data entry must be allowed when extracting analytical files, ranging from 2 weeks to 2 months depending on the variable of interest. Furthermore, depending on starting assumptions or when the analytical period is ended, certain types of data may be misreported or missing (for example, blood culture results may not be available for data extracts, discharge summaries may be pending for patients returning from the hospital and/or patient transfers among dialysis facilities, and correct identification of patients transferred among dialysis facilities for whom audit trails have yet to correctly identify and connect may be lacking [they may initially be recorded as two separate patients]). Such aspects of

the DIS may have implications for data sharing when systems that receive data extracts (*e.g.*, for CROWNWeb) have built-in business rules that may require automatic cutoff periods and/or restrict provision of updated data from the previous data period. Therefore, an understanding of the data flow is crucial when enabling communication and health information exchange between systems (Figure 1).

Future Directions

The limitations listed above are not insurmountable and actually, are dwarfed by the potential use of these large-provider databases, both as currently constructed and with forthcoming planned enhancements. The data platforms used by different dialysis provider DIS and their databases were built with the vision of increased flexibility to participate in electronic health information exchange. Therefore, future integration or communication with selected physician practices, participating hospital systems, and other patient care sites is not a question of if but when it will occur. The dialysis patient at home will eventually become linked to the dialysis facility as improvements in telehealth and electronic communications allow for interaction between facility care staff, the home apparatus, the patients, and even the caregivers when appropriate.

Therefore, as a chronic care condition, patients on maintenance dialysis provide an ideal setting for testing the patient-centered care concept—with the informatics hub of care being the dialysis unit and the overall director of care being the nephrologist. These large-provider databases can become repositories of diagnostic tests and procedures performed elsewhere (*e.g.*, radiographic imaging or angiography), prescriptions (*e.g.*, not limited to medicines but even for physical therapy or supplemental oxygen), and even consultation/medical/surgical notes. Now, imagine if later-stage patients with CKD could be initiated into a programmatic education and care before they reach ESRD stage through the partnerships built with managing nephrologists. These patients can be shepherded through the CKD process with coordinated care—some opting for pre-emptive kidney transplantation, others opting for home-based therapies, and others opting for conservative management or palliative care—and then, there are patients who will opt for in-center hemodialysis. This collaboration, when thrust on the pre-existing data-dense informatics of LDOs, creates an extremely well-suited environment to deliver integrated healthcare, consistent with the national healthcare goal to develop accountable care organizations (19).

The ability to longitudinally track interventions and outcomes as well as make corrections and decisions close to real time (and often at the point of care) creates the opportunity to unify an otherwise system of fragmented care for the CKD patient as they approach the need for maintenance dialysis or kidney transplantation and embark on the chosen form of renal replacement therapy. It provides for opportunities to compare performance based on physician and physician practice-based analytics (perhaps the next stage in evolution after facility-based analytics), particularly when the processes of care associate with multiple physician practices that are reflected in the mixed outcomes observed from many dialysis units. Furthermore, potential benefits of linking clinical databases with each other as well as linking clinical databases with claims databases to provide additional insight into patient treatment may be complicated by the lack of standardization of data capture (*e.g.*, the use of Medical Dictionary for Regulatory Activities terminology to capture safety events). This concern is another hurdle to overcome.

Finally, with the prohibitive costs of randomized clinical trials, the value of large-provider databases is increased with developing acceptance of the concepts of and support for pragmatic clinical trials. The initial epidemiologic analyses can be used to frame the hypotheses and determine statistical power, whereas the proposed interventions can be implemented through randomly selected or matched patient care clusters; they can even be stratified by facility, area, region, or provider. Outcomes may be determined from routinely collected variables and supplemented by external elements as necessary. Therefore, as we embark not only on improved quality and safety of patient care, we develop an enhanced ability to determine comparative effectiveness, efficiency, feasibility, and scalability of interventions.

Conclusion

In summary, large-provider databases create a data-dense environment that captures a large proportion of healthcare provided to CKD patients on maintenance hemodialysis. These databases currently bridge DIS across dialysis facilities within provider systems and provide links to or are being developed to provide lins to laboratory, vascular access, and CKD-related medication information systems. They provide longitudinal follow-up of patient conditions and patient care that allows for analytics that can be performed close to real time. These analytics address or inform queries regarding patient distributions, clinical indicators, problematic conditions, best practices, and patient outcomes at a level of granularity that is not always achievable through claims-based data. These analyses inform issues that may involve decisions on public policy or clinical practice. The value that these data collection systems bring will continue to increase as we look to a data-driven future, where the integration of medical information becomes the foundation of patient-centered care models, such as accountable care organizations.

Acknowledgments

The authors would like to thank Dr. Frank Maddux and Dr. Allen Nissenson for their helpful comments.

The views expressed in this editorial review belong to the authors and do not necessarily reflect the position of Fresenius Medical Care, North America or DaVita, Inc.

Disclosures

M.K. and H.M.W. are employees of DaVita Clinical Research, a wholly owned subsidiary of DaVita, Inc. E.L. is an employee of Fresenius Medical Care.

References

1. Briggs JP: Evidence-based medicine in the dialysis unit: A few lessons from the USRDS and the NCDS and HEMO trials. *Semin Dial* 17: 136–141, 2004

2. United States Renal Data System (USRDS): United States Renal Data System (USRDS). Available at: http://www.usrds.org/. Accessed May 18, 2012
3. University of Michigan Kidney Epidemiology and Cost Center: University of Michigan Kidney Epidemiology and Cost Center. Available at: http://www.sph.umich.edu/kecc. Accessed May 20, 2012
4. Dialysis Facility Reports: Dialysis Facility Reports. Available at: http://www.dialysisreports.org/. Accessed May 20, 2012
5. United Network for Organ Sharing: United Network for Organ Sharing. Available at: http://www.unos.org/. Accessed May 20, 2012
6. International Quotidien Dialysis Registry: International Quotidien Dialysis Registry. Available at: http://www.quotidiandialysis.org/. Accessed May 20, 2012
7. Dialysis Outcomes and Practice Patterns Study (DOPPS): Dialysis Outcomes and Practice Patterns Study (DOPPS). Available at: http://www.dopps.org/. Accessed May 20, 2012
8. Wingard RL, Chan KE, Lazarus JM, Hakim RM: The "right" of passage: Surviving the first year of dialysis. *Clin J Am Soc Nephrol* 4[Suppl 1]: S114–S120, 2009
9. Lacson E, Teng M, Ong J, Vienneau L, Ofsthun N, Lazarus JM: Antibody response to Engerix-B and Recombivax-HB hepatitis B vaccination in end-stage renal disease. *Hemodial Int* 9: 367–375, 2005
10. Lacson E Jr, Xu J, Lin SF, Dean SG, Lazarus JM, Hakim RM: A comparison of SF-36 and SF-12 composite scores and subsequent hospitalization and mortality risks in long-term dialysis patients. *Clin J Am Soc Nephrol* 5: 252–260, 2010
11. Lacson E Jr, Ikizler TA, Lazarus JM, Teng M, Hakim RM: Potential impact of nutritional intervention on end-stage renal disease hospitalization, death, and treatment costs. *J Ren Nutr* 17: 363–371, 2007
12. Lacson E Jr, Wang W, Lazarus JM, Hakim RM: Hemodialysis facility-based quality-of-care indicators and facility-specific patient outcomes. *Am J Kidney Dis* 54: 490–497, 2009
13. Lacson E Jr, Teng M, Lazarus JM, Lew N, Lowrie E, Owen W: Limitations of the facility-specific standardized mortality ratio for profiling health care quality in dialysis. *Am J Kidney Dis* 37: 267–275, 2001
14. Mayne TJ, Burgess M, Weldon J: Finding the case mix adjusters in the bundle. *Nephrol News Issues* 24: 34–40, 2010
15. Blumenthal D: Wiring the health system—origins and provisions of a new federal program. *N Engl J Med* 365: 2323–2329, 2011
16. Blumenthal D: Implementation of the federal health information technology initiative. *N Engl J Med* 365: 2426–2431, 2011
17. Fineberg HV: Shattuck Lecture. A successful and sustainable health system—how to get there from here. *N Engl J Med* 366: 1020–1027, 2012
18. Agarwal R, Metiku T, Tegegne GG, Light RP, Bunaye Z, Bekele DM, Kelley K: Diagnosing hypertension by intradialytic blood pressure recordings. *Clin J Am Soc Nephrol* 3: 1364–1372, 2008
19. Nissenson AR, Maddux FW, Velez RL, Mayne TJ, Parks J: Accountable care organizations and ESRD: the time has come. *Am J Kidney Dis* 59: 724–733, 2012

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