UC Davis
UC Davis Previously Published Works

Title

A novel tool for visualizing chronic kidney disease associated polymorbidity: a 13-year cohort study in Taiwan

Authors

Huang, Chih-Wei
Syed-Abdul, Shabbir
Jian, Wen-Shan
et al.

A novel tool for visualizing chronic kidney disease associated polymorbidity: a 13-year cohort study in Taiwan

Chih-Wei Huang[1,*], Shabbir Syed-Abdul[1], Wen-Shan Jian[2,3], Usman Iqbal[1], Phung-Anh (Alex) Nguyen[1], Peisan Lee[1,4], Shen-Hsien Lin[1], Wen-Ding Hsu[1,5], Mai-Szu Wu[6,7], Chun-Fu Wang[8,*], Kwan-Liu Ma[8,§], Yu-Chuan (Jack) Li[1,9,§]

ABSTRACT

Objective The aim of this study is to analyze and visualize the polymorbidity associated with chronic kidney disease (CKD). The study shows diseases associated with CKD before and after CKD diagnosis in a time-evolutionary type visualization.

Materials and Methods Our sample data came from a population of one million individuals randomly selected from the Taiwan National Health Insurance Database, 1998 to 2011. From this group, those patients diagnosed with CKD were included in the analysis. We selected 11 of the most common diseases associated with CKD before its diagnosis and followed them until their death or up to 2011. We used a Sankey-style diagram, which quantifies and visualizes the transition between pre- and post-CKD states with various lines and widths. The line represents groups and the width of a line represents the number of patients transferred from one state to another.

Results The patients were grouped according to their states: that is, diagnoses, hemodialysis/transplantation procedures, and events such as death. A Sankey diagram with basic zooming and planning functions was developed that temporally and qualitatively depicts they had amid change of comorbidities occurred in pre- and post-CKD states.

Discussion This represents a novel visualization approach for temporal patterns of polymorbidities associated with any complex disease and its outcomes. The Sankey diagram is a promising method for visualizing complex diseases and exploring the effect of comorbidities on outcomes in a time-evolution style.

Conclusions This type of visualization may help clinicians foresee possible outcomes of complex diseases by considering comorbidities that the patients have developed.

Key words: visualize analytic; data visualization; CKD polymorbidity visualization; comorbidity visualization; CKD Sankey diagram

<div style="writing-mode: vertical">RESEARCH AND APPLICATIONS</div>

BACKGROUND AND SIGNIFICANCE

Over the last few years, the rapid proliferation of electronic health records has led to the accumulation of extraordinary amounts of patient-level clinical data.[1–4] Health databases are constantly increasing in volume, variety, and veracity.[5–7] Innovative analysis of these datasets can lead to new insights into the essence of diseases. From an observer's perspective, disease is a spectrum of manifestations that evolve over time. Conjoining pathogenesis could open up new dimensions for the progression and outcome of diseases.[8,9] Therefore, a complete disease model has to include not only the temporal progression but also the branching caused by comorbidities.

In several all-cause mortality studies on chronic kidney disease (CKD) in Taiwan, diabetes mellitus (DM), hypertension (HTN), and hyperlipidemia were found to be independent risk factors associated with the future development of CKD.[10–13] Wen *et al*. reported in his study that smokers and obese individuals in the CKD group were at a significantly higher risk than those without CKD. Previous studies showed that the prevalence and variety of comorbidities may change over different stages of CKD.[10] For instance, the percentage of DM, as a comorbidity of CKD, did not change significantly from CKD stage IV (i.e., 25.6%). In contrast, the percentage of anemia increased from 58.8% to 92.5% in stage V, stage IV,

Correspondence to: Kwan-Liu Ma, PhD, Professor and Chair, Graduate Group in Computer Science, Department of Computer Science, University of California at Davis, 2121 Kemper Hall, One Shields Avenue, Davis, CA 95616, USA; ma@cs.ucdavis.edu
Yu-Chuan (Jack) Li, MD, PhD, FACMI, Professor and Dean, Graduate Institute of Biomedical Informatics, College of Medicine Science and Technology, Chair, Department of Dermatology, Wan Fang Hospital, Taipei Medical University, Wu-Hsing Str. No. 250 Taipei, 11031 Taiwan.
jack@tmu.edu.tw

respectively.[11] From the perspective of prevention, great interest also lies in understanding what happened before CKD (pre-CKD) and how this affects the progression and outcome of the disease. By considering the occurrence of comorbidities in pre-CKD, CKD, and post-CKD states, we will have a more complete view of this complex disease.

CKD has emerged as a global public health burden due to the increasing number of patients, high risk of progression to end-stage renal disease (ESRD), and poor prognosis of morbidity and mortality.[14–16] Compared to other countries, Taiwan has a relatively high incidence and prevalence rates of CKD patients. The prevalence of clinically recognized CKD increased from 1.99% in 1996 to 9.83% in 2003.[11] In particular, CKD was nearly three times more prevalent among low socioeconomic status individuals than among their high socioeconomic status counterparts. Clinicians' and researchers' attention tends to focus on its epidemiology, risk factors, treatment plans, and preventive actions.[17]

Researchers have tried to develop models or scoring systems by weighing common variables associated with CKD. For example: The National Kidney Foundation Kidney Disease Outcomes Quality Initiative guidelines,[18] Creatinine Clearance (Cockcroft–Gault Equation),[19] SCORED,[20] the Kidney Early Evaluation Program,[21] MDRD GFR Equation,[22] and so on. However, using one single model to represent a complex chronic disease such as CKD is difficult and controversial.

OBJECTIVE

"A picture is worth a thousand words" is the biggest advantage of visualization. This web-based visualization tool displays a zoomable time-evolution diagram of pre- and post-CKD comorbidities. Due to the complexity of the relationship between CKD and its comorbidities, it would be difficult for a clinician to associate specific combinations of comorbidities with patient outcomes. Traditional biostatistics and personal experiences may offer little help in terms of determining the time it will take for CKD to evolve into ESRD, much less the time from ESRD to death. Therefore, we aim to marry the big data approach with novel visualization techniques, in order to reveal the true spectrum of CKD. We believe such an approach could lead to a more intuitive understanding of possible patient disease pathways and outcomes.

METHODS AND MATERIALS

In this study, a CKD dataset and a novel multi-dimensional visualization method were used to render the state of the disease. Data visualization processing included data generation, data conversion, and visual presentation. Using CKD as an example, we obtained data from NHIDB, which contains millions of people with CKD medical records. We highlighted eleven underlying CKD-related diseases with nephrologists' opinions including three ESRD procedures and four states illustrating the conditions of CKD patients' entry and exit. We developed algorithms to incorporate the data into visual charts as well as to apply visual analytics and graphics in further research studies.

Study design

The aims of the study were: a) to define CKD progression, including the most related disease causes and outcomes; b) to collect patients' medical records in this CKD study from Taiwan National Health Insurance's one-million cohort dataset; c) to represent and visualize CKD development in an interactive disease spectrum; and d) to interpret the spectrum so that the knowledge discovered may lead to a better understanding of diseases and their associations. The study model is shown in Figure 1 and the data process with visualization approach is shown in Figure S1 of the Appendix.

Study Population

The CKD sample data derived from the one million NHIDB from 1998 to 2011. On the basis of nephrologists' opinions, we selected eleven underlying diseases which were considered to be most closely related to causing CKD and three following procedures within the ESRD stage after patients had been diagnosed with CKD (see Figure 2).

We used the ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) codes from NHIDB to identify CKD patients with outpatient visits (see Table S1 in Appendix). The selection of target subjects is described in Figure S2 of the Appendix. Regarding the input data and attributes in this pilot study, we selected the essential information for constructing the Sankey diagram for CKD individuals, such as anonymous patient ID, date of clinic visit, ICD9-CM codes, and procedure codes. To confirm the date of CKD diagnosis as the index date, the outpatient visit records were checked in 1998 to see whether the patients had CKD before 1999 or not. We excluded the patients who had already been diagnosed with CKD before 1998. We followed a total of 39 928 patients from 1999 to 2011, observing whether they were diagnosed with CKD, their underlying diseases, and the procedures applied in their treatment. Patients who made less than two visits for CKD in each year during the study period were excluded. Finally, we were left with 14 567 eligible patients who fulfilled our study criteria and we monitored them during the study years.

Visualization techniques

We aligned the date of each patient's first CKD diagnosis in "0y," down-sampled and aggregated them in every 2 years, and then clustered them into nonoverlapping groups. The output results are shown in a Sankey diagram–style timeline.

The Sankey diagram was developed by Irish engineer Riall Sankey to analyze the thermal efficiency of steam engines more than a hundred years ago. It is a specific type of flow diagram, in which the width of each state and transition is shown proportionally to the quantity of flow. The use of Sankey diagrams has been standard practice in science and engineering since its development.[23] These diagrams are typically used to visualize energy, materials, or cost transfers between processes. This study is the first to use a Sankey diagram to depict the evolution of CKD.

Our system workflow resembles a visual analytic framework[24,25] and incorporates three types of tasks: a) modeling;

RESEARCH AND APPLICATIONS

RESEARCH AND APPLICATIONS

Figure 1: Study model.
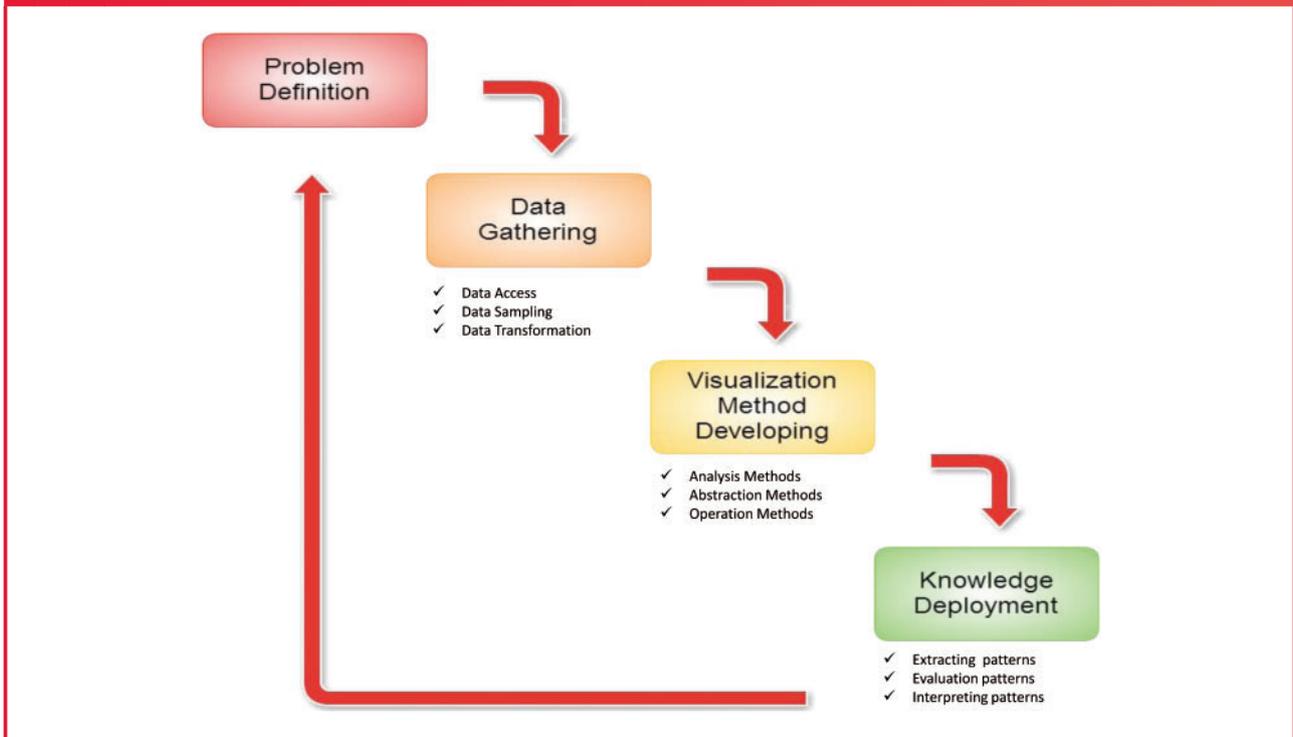
Figure 2: Chronic kidney disease progression from a Nephrologist's perspective.
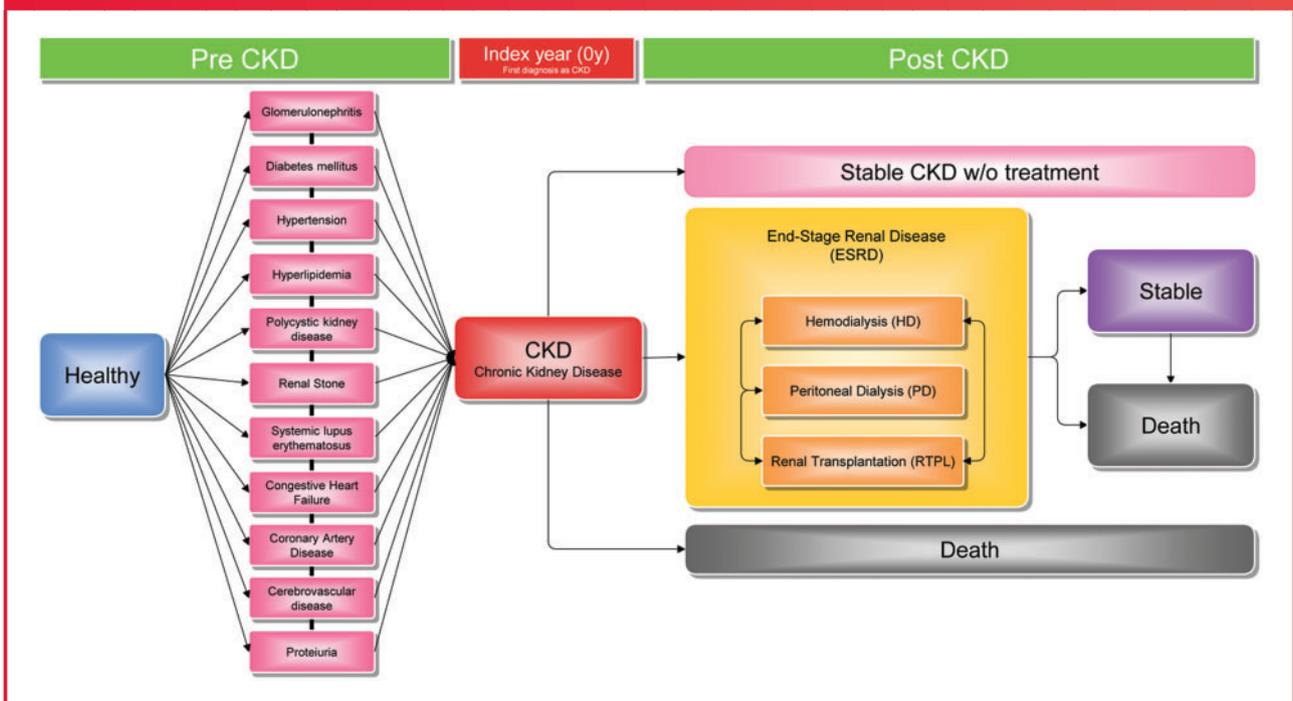
b) similarity analysis; c) visual encoding; and d) interactions to provide the analytical supports. In this study, we conducted six tasks sequentially as follows:

1. Disease/procedure code aggregation: Using domain expert-defined association rules to cluster the 11 024 ICD9-CM codes and 42 351 procedure codes down to 15 representative groups with labels, such as CKD, HTN, hemodialysis (HD), and peritoneal dialysis (PD),etc.

2. Re-constructing patient states: Using outpatient visit records and then group labels to re-construct the patient health states over time.

3. Constructing patient state trajectory network: To compare and study between different patient trajectories, we aligned these trajectories according to the time of their first CKD diagnosis and projected it onto a network where each node is a patient state and each patient trajectory is a path travelling through the network.

4. Clustering with user-specified rules: We simplified the network by: a) aggregating nodes into several 2-year partitions and merging those with the same states in a partition; b) aggregating nodes with small patient quantities; and c) letting users choose label-of-interests for each time partition to remove unwanted information.

5. Visual encoding as Sankey diagram of unique patient states over time: We visualized the network in a representation similar to a Sankey diagram[26] or a parallel set.[27] In addition, the nodes were strictly laid out along the time axis to reveal the temporal patterns. We assigned the color to distinguish between different combinations of diseases. Each node was colored based on patient state, and each edge was colored based on the transition which two nodes connected. However, as there were many more combinations than colors, they were not always uniquely mapped. Thus, you may find two different combinations of diseases sharing the same color. In this case, the user can distinguish each patient's state by hovering the mouse cursor over the cluster to get a popup showing the cluster's name. When the two nodes of an edge were of different states and thus different colors, a gradient was applied for a smoother visual transition.

6. Interactions: The user can select a cohort or an association between two cohorts by clicking on the node or the edge. The system highlights the selected patients with a dark shaded area on the diagram. As a result, it reveals all the paths that the selected patients have been through as well as their proportions to the rest of the cohorts. Within each time step, all patient clusters were nonoverlapped. That means two things: a) each cluster represented a unique combination of diseases; and b) a patient only belonged to the cluster if he or she had the exact combination of diseases which the cluster represented. For example, one patient had diseases A and B, another had diseases B and C; they were put in different cluster groups even though they both had disease B.

In addition, we added four special groups to provide contextual information and support the exploration processes. They were "no records" patients who did not have any disease at all due to a lack of records; "others with/without CKD" for minor groups whose number of patients was below the customizable threshold; "death" patients for whom we could not find any records in 1 year after the last visit until 2011 (Table S1). Therefore we grouped them together and only differentiated between them according to whether they had ever had CKD before or not.

Our system's user interface is web-based, tested with a commodity desktop as the application server and another desktop as the client. Most of the programs are written in Python and raw data were stored as static files in the hard drive. We cache less frequently, updating intermediate results of each step in the analysis process to improve end-user responsiveness. Comparing patient clustering, risk factor aggregation, and patient state reconstruction usually only happen at the beginning of a case study. Caching the output of such steps helps us to reduce unnecessary processing time. For 14 567 patients and 6 031 579 records, it took 6 min to filter and aggregate the risk factors. The average duration to see an update of patient clustering was 5 s.

RESULTS

For each type of disease, procedure, or state extracted from the records, we connect the dates of visits and convert them into a single interval-based event. Thus we have several partially overlapped events, and each of them represents a type of disease, procedure, or state. Figure 3 shows: pre- and post-CKD status, the threshold of patients with minimum to maximum, and the number of patients and their disease transitions. This tool clearly demonstrates the comorbidities that develop CKD and how the comorbidities change the outcomes and influence diseases. Users can explore the data, and uncover and interpret insights from the interactive capabilities of the system.

It is effectively a nonlinear, low-pass filter to reduce trivial transitions caused by a patient's chronic visits or interleaved visits for different purposes. The filtered events are then aligned by the date of each patient's first CKD diagnosis so that the user can easily compare the CKD pathways between patients. Transitions between events are down-sampled with a 2-year period to reduce over-plotting. Meanwhile, all events within a sampling period are aggregated, so that shorter duration events are not lost due to down-sampling.

Our overall visual design follows the principle of a Sankey diagram, which visualizes the transfers between groups with lines and quantities by line width. In our study we focused on the transfers between groups of CKD patients over time. The patients were grouped according to their states, defined by the combinations of the patients' active diseases, procedures, and their states at that time. Note that none of the groups are overlapped. The overall layout has been optimized to reduce the number of line intersections during the transfer.

RESEARCH AND APPLICATIONS

RESEARCH AND APPLICATIONS

Figure 3: Illustration of chronic kidney disease spectrum interface.
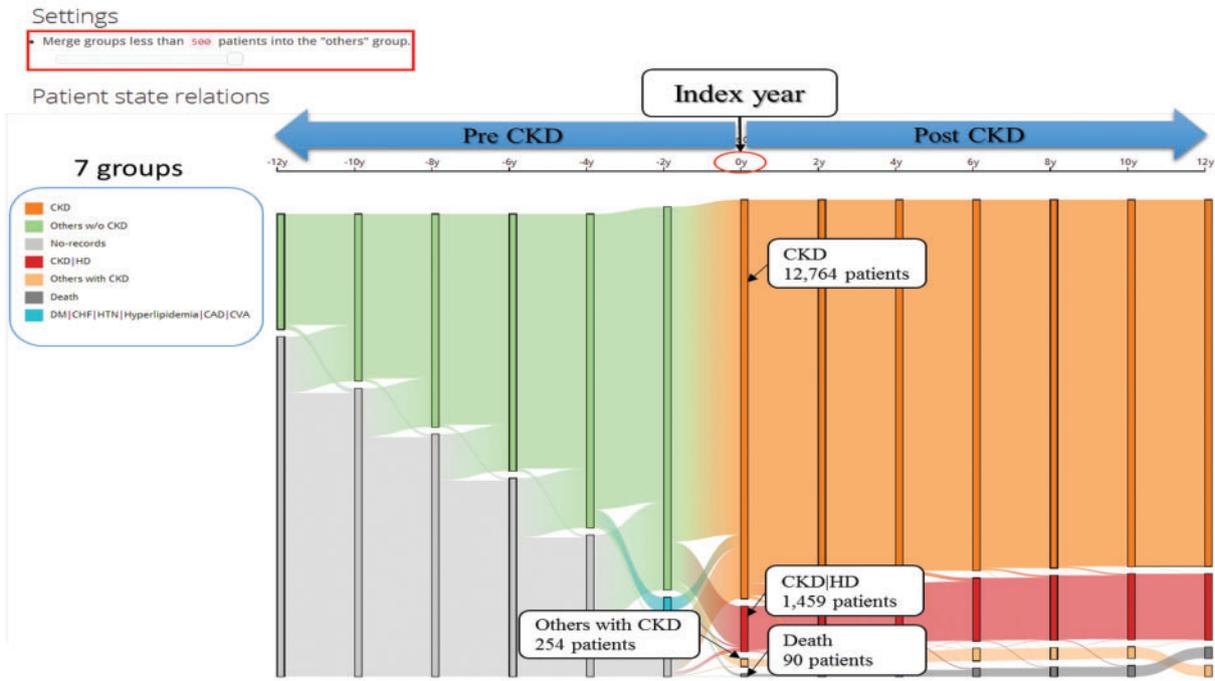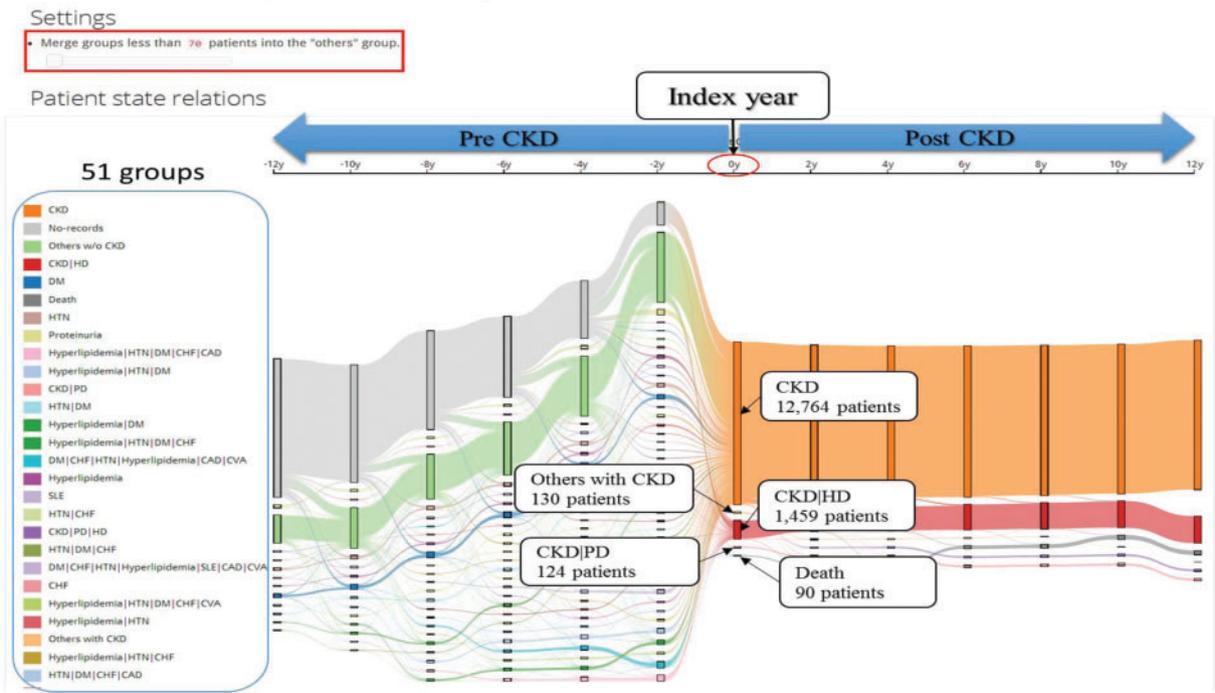
The ICD9-CM disease codes and the procedure codes were used to classify a patient's symptoms and treatments, respectively. However, there are hundreds of different codes and thousands of possible combinations which could be used to drive the Sankey diagram, but we would always end up with over-plotted and unreadable results. Therefore, it is important to classify codes into several focused areas. In our case, the consultation with nephrologists was very helpful for grouping

into focused areas before and after patients were diagnosed with CKD.

Figure 4 shows how this interactive CKD spectrum and their nodes are strictly laid out along the time axis to reveal the temporal patterns. We have chosen one combination of CKD and HD to explain multi-dimension information using the Sankey diagram. The CKD spectrum visually encodes nodes in the flow graph using vertical rectangles. Transition is between two nodes, the patient's progression and patient numbers presented by clicking transitions or nodes. Colors have been used to encode the same disease or disease combinations. In Figure 4A, we set up a threshold of at least 250 patients in one group as an example. We observed 1459 CKD patients who had been diagnosed with CKD first and took the HD within the same year. We found 10 combinations in the pre-CKD stage. Figure 4B shows that three conditions occurred in the post-CKD stage with HD within 2 years. From this group, 1399 patients still took HD, 32 patients died and 28 did not take HD continually. From 28 patients classified in the "Others with CKD" group, we observed 6 patients who started to take PD.

DISCUSSION

In this study, we have demonstrated an innovative approach for the visualization of temporal patterns for diseases and outcomes. We found that using a Sankey diagram in healthcare is a promising method for visualizing a specific disease of interest, especially for observing its development over time. By adjusting the threshold of patient numbers, users can target different aspects to explore how diseases, comorbidities, and polymorbidities affect patients' outcomes.

In this article, we set the threshold at 250 out of 14 567 CKD patients as an example. The interactive CKD spectrum displays the development of the 11 pre-CKD diseases and 3 procedures over the span of 2 years. About 3.4% of patients took HD in the first year of CKD. DM, HTN, and Proteinuria have been found to be single entity diseases which are related to causing CKD. Besides these, other polymorbidities such as a) hyperlipidemia, HTN, and DM group, b) DM, congestive heart failure (CHF), HTN, hyperlipidemia, coronary artery disease, and cerebrovascular disease group, c) hyperlipidemia, HTN, DM, CHF, and coronary artery disease group, d) hyperlipidemia, HTN, DM, CHF group, and e) DM, CHF, HTN, hyperlipidemia, systemic lupus erythematosus, coronary artery disease, and cerebrovascular disease group, include eight main diseases which also lead to CKD. Interestingly, we found 966 patients who did not have underlying diseases prior to being diagnosed with CKD and HD. Therefore, we should pay more attention to discovering patients who did not have any of the underlying diseases suggested by nephrologists. We need to determine what other diseases and comorbidities can potentially lead to a greater risk of developing CKD.

Several related studies have been conducted on visualization methods.[25,28–30] One came from the Outflow project,[24,31] and was conceptually similar to our study, focusing on severe disease crucial issues by using cohort patients' data. The study processed temporal event data and visualized aggregated event progression pathways with diseases. We also aimed to discover correlations within the disease progression by using interactive visualization. However, they built a system by using Framingham symptoms as entities to visualize CHF disease progression.[32] In contrast, our study focused primarily on the temporal analytics by visually emphasizing three main properties of patients: sub-populations' temporal statistics of patients, quality assessment, and uncertainty of patients. Moreover, they used specific indicators for CHF such as Framingham symptoms, whereas there are still a lot of known and unknown risk factors that can lead to CKD.[32] CKD progression and development is more complicated than most diseases. For that reason, we need a useful visualization tool to help improve our understanding of CKD evolution.

In a recent study, "the disease map" visualized how human diseases evolved over time and age.[33] The map visualized the data by classifying into age and gender to represent disease evolution over time. Whilst many of such positive associations between disorders are already well-known and understood, many still remain either entirely unnoticed or not properly recognized. Those unnoticed and unrecognized associations could be novel discoveries. Therefore, the study assumed that certain human disorders can be associated negatively. Although this study provides a novel tool for validating already known diseases and can even be utilized to discover new or hitherto unnoticed empirical association patterns between human diseases, it nonetheless lacks the incorporation of temporal data and information. Moreover, this tool is insufficient for observing one specific disease and its progression and development over time.[30,34] Therefore, our study is novel in providing a complete picture of a focused disease and its progression and development in relation to other underlying diseases. Representing these comorbidities and polymorbies in a spectrum is very effective and simple to understand for both lay people and health professionals.[35–37] Therefore, we think our Sankey diagram visualization technique could provide more authentic and interesting disease spectrums overtime.

As this is a pilot study at this stage, it has its limitations which we must deal with in future work. First, gender and age need to be taken into consideration for the next step of visualization. Some CKD studies have shown that females and elderly people were associated with a high risk of developing CKD.[10,11,15,17] We should also incorporate variables that we ignored in this pilot study, such as drug use and laboratory data which could make the CKD spectrum more useful and functional for clinicians to use.[38,39] After that, we should know more about CKD related features via visualization at a glance. The 11 underlying diseases selected by nephrologists could also limit the generalizability of results. Moreover, we found that CKD progresses by lying in three stages: pre-CKD, post-ESRD, and outcome (stable or death). Thus, we will focus on patients lying in these three stages and their disease variation in the following step of our future research. The NHIDB data used in this study is mainly for billing purposes, which could introduce bias in accurate disease diagnosis.

RESEARCH AND APPLICATIONS

Figure 4: The pre- and post-combination of chronic kidney disease with hemodialysis in the first year.
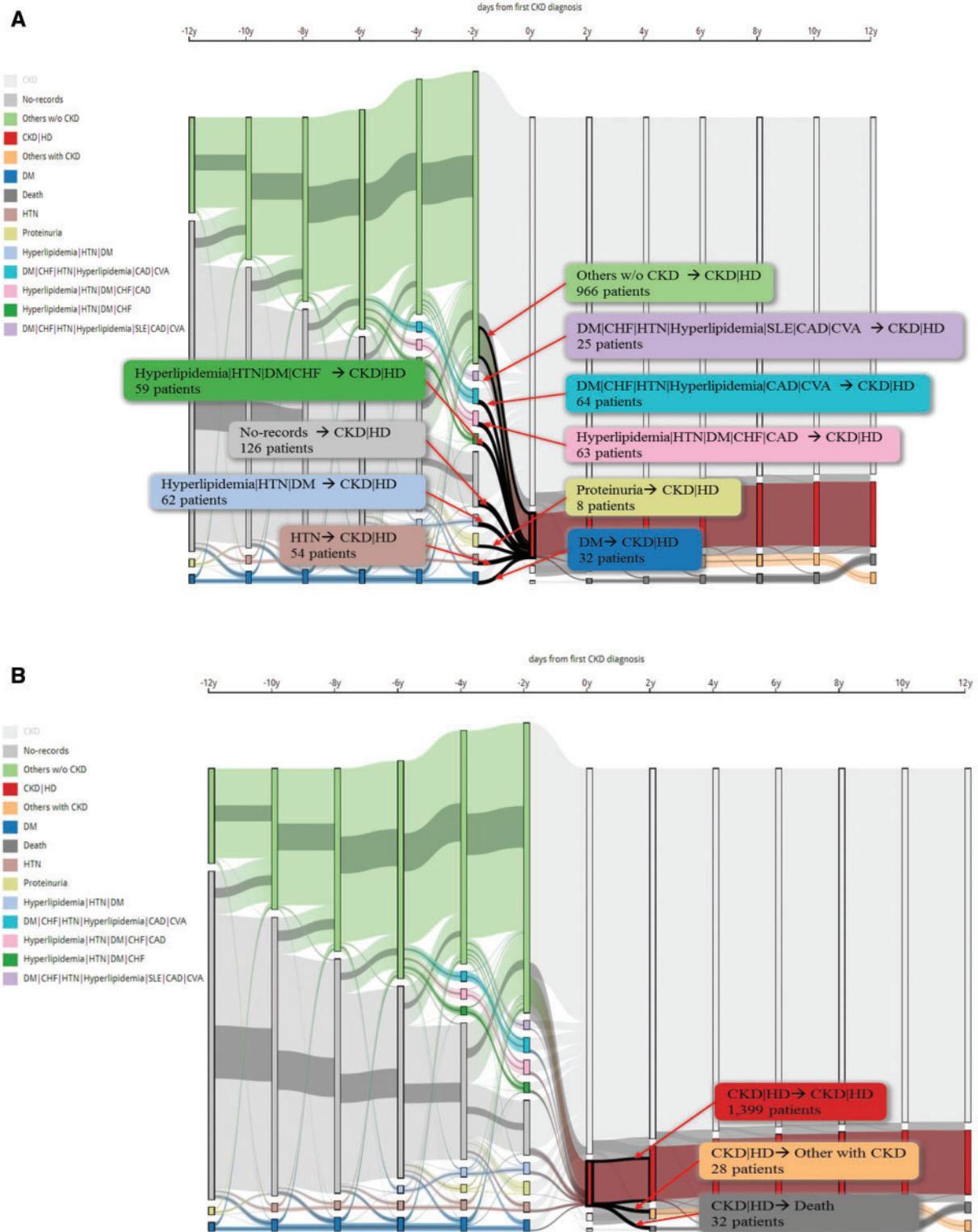
CONCLUSION

We developed a visualization tool based on a Sankey diagram that can represent the comorbidity and progression of CKD over time. This tool has the potential to help clinicians when deciding on the management of pre-CKD and CKD patients. In terms of medical education, it can be used to help students understand the nature of a complex chronic disease by showing the dynamics between the comorbidities and CKD outcomes in a time-evolution fashion. Patients can benefit from this tool in terms of preventive measures to anticipate the plethora of longer term medical outcomes.

We believe that the visualization of comorbidities and outcomes of pre-CKD, CKD, and post-CKD can lead us to a better understanding of underlying pathogenesis. Efforts in this direction will eventually aid in prediction and prevention of the disease, personalization of diagnosis and treatment, as well as the participation of patients in our healthcare system.[40]

CONTRIBUTORS

C-W.H., C-F.W., K-L.M., and Y-C.L. invented and developed the concept. Y-C.L., W-S.J., and S-H.L. obtained the data. W-S.J., K-L.M., S.S-A., and Y-C.L. organized and validated the data. C-W.H, Y-C.L., P-S.L., P-A.N., and S.S-A. reviewed the methods. C-F.W. wrote software, implemented the method, and developed the software. W-S.J. and S-H.L. gave technical support. M-S.W. and W.D-H. performed biological validation. C-W.H., C-F.W., U.I., and Y-C.L. wrote the manuscript.

COMPETING INTERESTS

None.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://jamia.oxfordjournals.org/.

REFERENCES

1. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.* 2012;13(6):395–405.

2. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Informat Assoc.* 2013;20(1):117–121.

3. Iqbal U, Ho CH, Li YC, *et al.* The relationship between usage intention and adoption of electronic health records at primary care clinics. *Comput Meth Prog Biomed.* 2013;112(3):731–737.

4. Lin CW, Abdul SS, Clinciu DL, *et al.* Empowering village doctors and enhancing rural healthcare using cloud computing in a rural area of mainland China. *Comput Meth Prog Biomed.* 2014;113(2):585–592.

5. Buhl HU, Röglinger M, Moser D-KF, Heidemann J. Big data. *Wirtschaftsinformatik.* 2013;55(2):63–68.

6. Wang W, Krishnan E. Big data and clinicians: a review on the state of the science. *JMIR Med Informat.* 2014;2(1):e1.

7. Beyer MA, Laney D. *The Importance of 'Big Data': A Definition.* Stamford, CT: Gartner; 2012.

8. Ou-Yang C, Agustianty S, Wang HC. Developing a data mining approach to investigate association between physician prescription and patient outcome - a study on re-hospitalization in Stevens-Johnson Syndrome. *Comput Meth Prog Biomed.* 2013;112(1):84–91.

9. Huang CL, Nguyen PA, Kuo PL, *et al.* Influenza vaccination and reduction in risk of ischemic heart disease among chronic obstructive pulmonary elderly. *Comput Meth Prog Biomed.* 2013;111(2):507–511.

10. Wen CP, Cheng TYD, Tsai MK, *et al.* All-cause mortality attributable to chronic kidney disease: a prospective cohort study based on 462 293 adults in Taiwan. *Lancet.* 2008;371(9631):2173–2182.

11. Kuo H-W, Tsai S-S, Tiao M-M, Yang C-Y. Epidemiological features of CKD in Taiwan. *Am J Kidney Dis.* 2007;49(1):46–55.

12. Zhang Q-L, Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health.* 2008;8(1):117.

13. Caughey GE, Vitry AI, Gilbert AL, Roughead EE. Prevalence of comorbidity of chronic diseases in Australia. *BMC Public Health.* 2008;8(1):221.

14. Eknoyan G, Lameire N, Barsoum R, *et al.* The burden of kidney disease: improving global outcomes. *Kidney Int.* 2004;66(4):1310–1314.

15. Couser WG, Remuzzi G, Mendis S, Tonelli M. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney Int.* 2011;80(12):1258–1270.

16. Jha V, Wang AY-M, Wang H. The impact of CKD identification in large countries: the burden of illness. *Nephrol Dial Transpl.* 2012;27(Suppl 3):iii32–iii38.

17. Hwang SJ, Tsai JC, Chen HC. Epidemiology, impact and preventive care of chronic kidney disease in Taiwan. *Nephrology.* 2010;15(Suppl 2):3–9.

18. Levy AS, Eckardt KU, Tsukamoto Y. Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int.* 2005;67(6):2089–2100.

RESEARCH AND APPLICATIONS

RESEARCH AND APPLICATIONS

19. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron.* 1976;16(1):31–41.

20. Bang H, Vupputuri S, Shoham DA, *et al.* SCreening for Occult REnal Disease (SCORED): a simple prediction model for chronic kidney disease. *Arch Intern Med.* 2007;167(4): 374–381.

21. McCullough PA, Li S, Jurkovitz CT, *et al.* CKD and cardiovascular disease in screened high-risk volunteer and general populations: the Kidney Early Evaluation Program (KEEP) and National Health and Nutrition Examination Survey (NHANES) 1999-2004. *Am J Kidney Dis.* 2008;51(4):S38–S45.

22. Levey AS, Stevens LA, Schmid CH, *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009; 150(9):604–612.

23. Schmidt M. The Sankey diagram in energy and material flow management. *Jl Ind Ecol.* 2008;12(1):82–94.

24. Wongsuphasawat K, Gotz D. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE T Vis Comput GR.* 2012;18(12):2659–2668.

25. Streit M, Schulz H, Lex A, *et al.* Model-driven design for the visual analysis of heterogeneous data. *IEEE T Vis Comput GR.* 2012;18(6):998–1010.

26. Riehmann P, Hanfler M, Froehlich B. Interactive sankey diagrams. IEEE Symposium on Information Visualization, 2005 (INFOVIS 2005); 2005. pp. 233–240.

27. Kosara R, Bendix F, Hauser H. Parallel sets: interactive exploration and visual analysis of categorical data. *IEEE T Vis Comput GR.* 2006;12(4):558–568.

28. Hsu W, Bui AA. Leveraging domain knowledge to facilitate visual exploration of large population datasets. *AMIA Annual Symposium Proceedings*, 2013; American Medical Informatics Association; 2013. p. 615.

29. Kumar A, Wang YY, Wu CJ, *et al.* Stereoscopic visualization of laparoscope image using depth information from 3D model. *Comput Meth Prog Biomed.* 2014;113(3):862–868.

30. Ramirez-Ramirez LL, Gel YR, Thompson M, *et al.* A new surveillance and spatio-temporal visualization tool SIMID: SIMulation of infectious diseases using random networks and GIS. *Comput Meth Prog Biomed.* 2013;110(3):455–470.

31. Wongsuphasawat K, Gotz D. Outflow: visualizing patient flow by symptoms and outcome. *IEEE VisWeek Workshop on Visual Analytics in Healthcare*; Providence, Rhode Island, USA, 2011.

32. McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. *N Engl J Med.* 1971;285(26):1441–1446.

33. Moldovan M, Enikeev R, Syed-Abdul S, *et al.* Disease universe: visualisation of population-wide disease-wide associations. Advances in Systems Science and Applications 2013;14(2):144.arXiv:1308.2557.

34. Ricketts K, Williams M, Liu ZW, Gibson A. Automated estimation of disease recurrence in head and neck cancer using routine healthcare data. *Comput Meth Prog Biomed.* 2014;117(3):412–424. doi: 10.1016/j.cmpb.2014.08.008.

35. Wicht A, Wetter T, Klein U. A web-based system for clinical decision support and knowledge maintenance for deterioration monitoring of hemato-oncological patients. *Comput Meth Prog Biomed.* 2013;111(1):26–32.

36. Van De Steeg L, Langelaan M, Wagner C. Can preventable adverse events be predicted among hospitalized older patients? The development and validation of a predictive model. *Int J Qual Health Care.* 2014;26(5):547–552.

37. Cramm JM, Nieboer AP. High-quality chronic care delivery improves experiences of chronically ill patients receiving care. *Int J Qual Health Care.* 2013;25(6):689–695.

38. Hassan Y, Al-Ramahi RJ, Abd Aziz N, Ghazali R. Drug use and dosing in chronic kidney disease. *Ann Acad Med Singapore.* 2009;38(12):1095–1103.

39. Inker LA, Schmid CH, Tighiouart H, *et al.* Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med.* 2012;367(1):20–29.

40. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat Rev Clin Oncol.* 2011;8(3):184–187.

AUTHOR AFFILIATIONS

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taiwan

[2]School of Health Care Administration, Taipei Medical University, Taipei, Taiwan

[3]Faculty of Health Sciences, Macau University of Science and Technology, Macau

[4]Institute of Biomedical Informatics, National Yang Ming University, Taipei, Taiwan

[5]Department of Nephrology, New Taipei City Hospital, New Taipei City Government, Taipei, Taiwan

[6]Division of Nephrology, Taipei Medical University Hospital, Taipei, Taiwan

[7]School of Medicine, Taipei Medical University, Taipei, Taiwan

[8]Department of Computer Science, University of California-Davis, California, USA

[9]Department of Dermatology, Wan-Fang Hospital, Taipei, Taiwan.

*Equal Contribution