Using causal directed acyclic graphs (DAGs) to select patient-important outcomes in transplantation trials—interventions to treat polyomavirus infection as an example

Kidney International (2023) **104,** 628–633; https://doi.org/10.1016/j.kint.2023.07.013

KEYWORDS: acute rejection; causal directed acyclic graph; clinical trial; transplantation

Yue Wu[1], Michael Dymock[2], Ryan Gately[3], Julie A. Marsh[2,4], Carmel Hawley[3], Germaine Wong[1,5] and Thomas L. Snelling[1,2,6,7,8]

[1]*Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales, Australia;* [2]*Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute, University of Western Australia, Nedlands, Western Australia, Australia;* [3]*Department of Nephrology, Princess Alexandra Hospital, Woolloongabba, Queensland, Australia;* [4]*Centre for Child Health Research, Medical School, University of Western Australia, Crawley, Western Australia, Australia;* [5]*Centre for Transplant and Renal Research, Westmead Hospital, Westmead, New South Wales, Australia;* [6]*Department of Infectious Diseases, The Children's Hospital at Westmead, Sydney Children's Hospitals Network, Westmead, New South Wales, Australia;* [7]*School of Public Health, Curtin University, Bentley, Western Australia, Australia; and* [8]*Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia*

Correspondence: *Yue Wu, Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales 2006, Australia. E-mail: yue.wu1@sydney.edu.au*

Transplantation can be life-changing for those with kidney failure because it improves their overall quality of life and survival. However, transplantation is not a cure and comes with complications. With the advent of improved immunosuppressive therapy and surgical techniques, short-term graft and patient survival have improved considerably over the past 3 decades. By contrast, improvements in longer-term graft and patient outcomes have been more modest. The key challenge of transplantation is to optimize graft function whilst avoiding complications associated with over-immunosuppression, such as opportunistic infections and cancers.[1] Although most transplantation centers follow standard immunosuppression protocols, the agents (and doses) used are highly dependent on patient and graft characteristics, clinician preferences, and drug tolerability. As the treatments used in post-transplantation care can have both beneficial and harmful effects, determining the specific outcomes to be assessed when evaluating alternative treatment strategies is crucial.

BK polyomavirus (BKPyV) infection is an opportunistic infection that can cause nephropathy, ureteric strictures, premature graft loss, and return to dialysis following kidney transplantation.[2] Currently, no proven antiviral treatments exist, and the mainstay of treatment involves judicious reduction of immunosuppression to facilitate immune reconstitution and viral suppression before the infection progresses to the point of causing damage to the allograft. However, reducing immunosuppression can trigger acute and subsequently chronic rejection, which is another cause of allograft dysfunction or loss. Proposed interventions for BKPyV infection therefore must be assessed for their effects on the various outcomes that are important to patients and key stakeholders.[3] These outcomes include (or are affected by) allograft function, the risk of opportunistic infection, and allograft loss.[4]

Intravenous immunoglobulin (IVIG) has been proposed as a candidate intervention for the management of BKPyV infection because it may possess BKPyV-specific neutralizing antibodies.[5] Passive immunization with IVIG in the early post-transplant phase has been associated with a lower incidence of viremia and BKPyV-associated nephropathy (BKPyVAN) in high-risk kidney transplant recipients. IVIG also is known to have immunomodulatory effects and may prevent antibody-mediated allograft rejection following a reduction in immunosuppression for the management of the viral infection.[5] However, trial evidence to support the use of IVIG in clinical practice is very limited.

Directed acyclic graphs (DAGs) are being used increasingly to communicate and reason about the mechanisms that underlie complex problems in health and other domains. In a DAG, a node represents a variable (or factor) relevant to the problem of interest, and an arrow (or arc) indicates the presence of a direct influence of predecessor (or parent) variables on their child nodes (nodes descending from other nodes). In a causal DAG, the arcs describe the mechanistic pathways affecting the probability of downstream outcomes, which can, in theory, be intervened upon to change the probability of that outcome. Causal DAGs are being used increasingly to aid in the design and analysis of clinical studies.[6,7] They typically represent the proponent's understanding of the problem domain based on scientific reasoning, knowledge of the literature, and/or data analysis. They serve as an explicit set of causal assumptions that are open to critique by others. With an agreed-on causal framework, researchers can identify and reason collectively when deciphering treatment effects, including potential mediators and confounders, and sources of selection or measurement biases. The construction of a causal DAG may require collaboration of subject-matter experts with those with expertise in casual inference.[8]

In this editorial, we present how we used a causal DAG to help us interpret outcomes of a planned trial involving people with transplant-associated BKPyV infection. The causal DAG was created to capture the clinical challenge of maintaining an acceptable balance between over- and under-immunosuppression in kidney transplantation and to identify potential confounders and competing risks across the different pathways. We hope our example illustrates the purpose of constructing causal DAGs, how best to approach this process, and its potential value in the design and interpretation of medical research.

Methods

We applied a previously described knowledge engineering approach[8] to create the causal DAGs. Preliminary DAGs were drafted by the modeller based on published literature and consultation with domain experts in nephrology and infectious diseases. A knowledge-elicitation survey (see the Supplementary Methods) was designed and conducted, followed by a workshop to consolidate understanding of the survey responses and revise the preliminary DAGs. The revised DAGs were described in written format, which was reviewed by other domain experts (i.e., those who did not participate in the initial elicitation activities) as a form of validation. Although our focus was on understanding causal mechanisms, we allowed some associations between variables that were not strictly causal—for example, for predictive variables that cannot be intervened upon, such as the patient's comorbidities prior to transplantation.

To depict the DAGs during the elicitation sessions and model development process, we used GeNIe software (https://www.bayesfusion.com), which is one of several tools that allow users to construct DAGs, such as DAGitty (http://www.dagitty.net/) and Netica (https://www.norsys.com/download.html).

Results

Introduction of the putative causal DAG for post-transplant BKPyVAN. The thick arrows in the causal DAG (Figure 1) highlight the proposed predominant pathways; perturbations of immune function (p6), either increased or decreased, may increase the risk of allograft damage, leading to graft dysfunction (p16) via either immune-mediated acute rejection (p10) or BKPyVAN (p13). The probability of these complications is influenced by a range of patient factors (dark blue nodes)—for example, patients' adherence to their prescribed immunosuppression (p2) is critical to allograft survival. The recipients' sensitization status (p4), and the degree of human leukocyte antigen (HLA) matching between the donor graft and the recipient (p7) may influence the risk of acute rejection, through their impact on the immune function and humoral response of the HLA donor–specific antibodies (p8). In addition, non-HLA antibodies (p9), such as the angiotensin type 1 receptor (AT1R) antibodies and other unknown immune mechanisms, can also mediate acute rejection. In Supplementary Table S1, definitions for each variable in the causal DAGs are provided.

We use the following 3 nodes to capture separate BKPyV-relevant events: first, latent BK infection (p11), which has a high seroprevalence of 90%; second, reactivation of BKPyV (p12) due to immunosuppression resulting in viruria and viremia; and third, viral infiltration of the allograft (BKPyVAN; prevalence of 3%–5%).[2] Although the mechanisms of allograft damage are complex, we have categorized allograft injury into 2 groups—reversible (p14) and irreversible damage (p15); here, reversible damage subsequently may become irreversible. Acute rejection and BKPyV-related events may cause reversible or irreversible damage, leading to allograft dysfunction (defined by a significant reduction in estimated glomerular filtration rate [eGFR]). Finally, clinicians' management responses (illustrated by red nodes) may affect this causal process directly, including selection of immunosuppressive therapy, its dose, the therapeutic response to acute rejection (p5), and the timing of various investigations (p18). In this context, latency suggests those events, or underlying disease processes (e.g., p10 and p12), that are not observable and therefore must be inferred using observable surrogates—for example, the measured viral load and the eGFR for viral burden and acute rejection, respectively.

In this putative causal DAG (Figure 1), we illustrate how the upstream processes can result in downstream complications, such as increased risk of mortality (p26) and reduced quality of life associated with allograft loss and return to dialysis (indicated by light blue nodes and brown text). We have highlighted how symptoms of kidney failure (e.g., fatigue, pain, lethargy, pruritis; p24) and overall treatment burden (p21) may influence the overall quality of life of the recipients (p27).

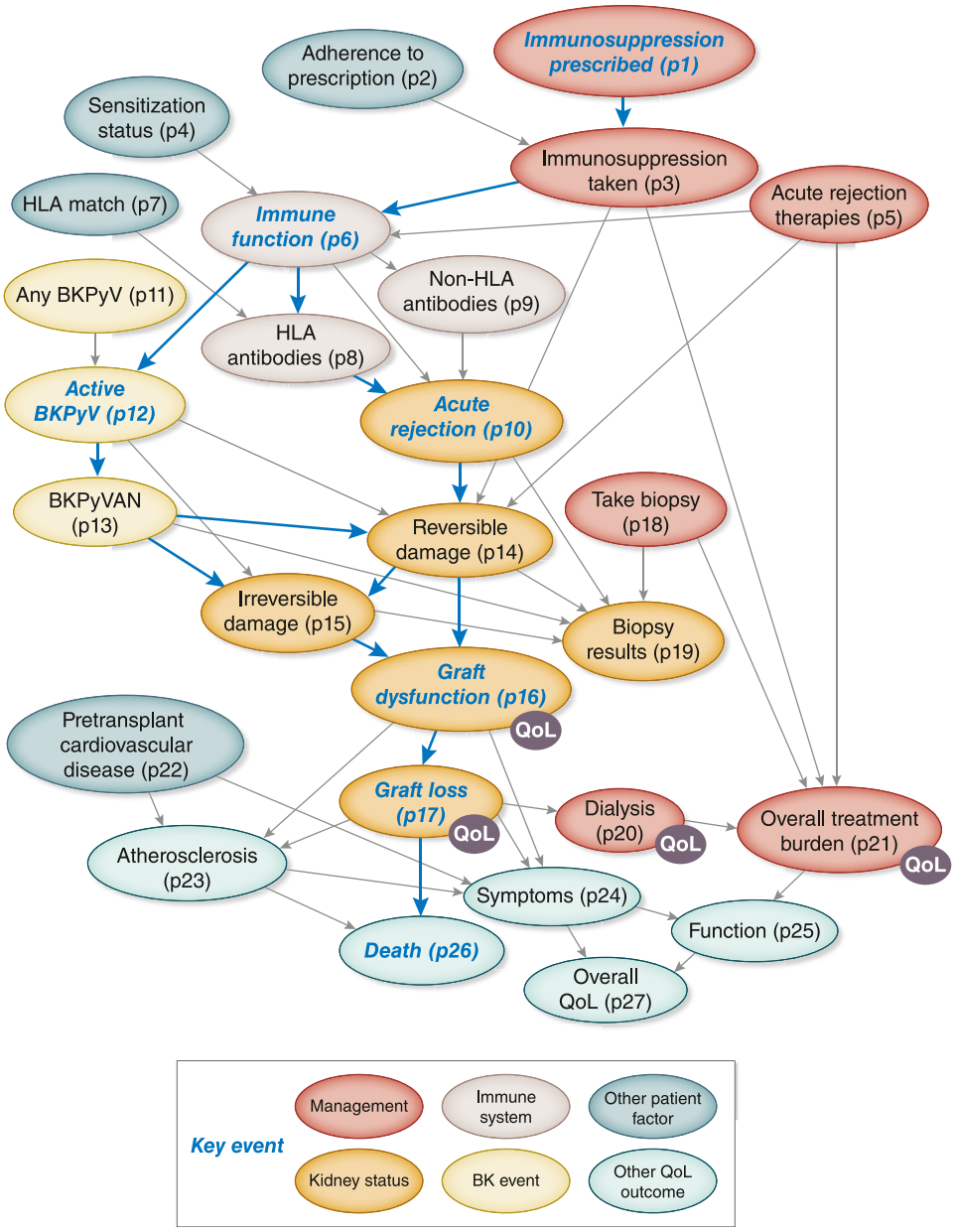

Figure 1 | The causal directed acyclic graph (DAG) for BK polyomavirus (BKPyV) infections and allograft outcomes (the putative [p] DAG). The detailed model structure and the definition for each variable are provided in Supplementary Table S1. BKPyVAN, BKPyV-associated nephropathy; HLA, human leukocyte antigen; QoL, quality of life.

Defining a multicomponent outcome to evaluate a treatment for BKPyVAN. To capture the trajectory of BKPyV infection, we have developed a dynamic DAG that describes the evolution of key events (Figure 1) in the predominant pathways, over both shorter and longer timeframes (Figure 2). We initialize this dynamic DAG by describing the baseline (t0) status of a transplant recipient with respect to immunosuppressive therapy, immune status, BKPyV viral load, risk of acute rejection, and eGFR. Each of these factors informs its own status in the subsequent time step. Although interactions among these key factors are replicated from the DAG structure in Figure 1 within each time step (t1 and t2), 2 new arcs are introduced to capture important feedback loops between time steps that cannot be captured in the single–time step DAG. These arcs represent how BK viral load and the presence of acute rejection can influence the clinician's decision to alter immunosuppression. Specifically, immunosuppression doses will be increased if evidence of acute rejection is

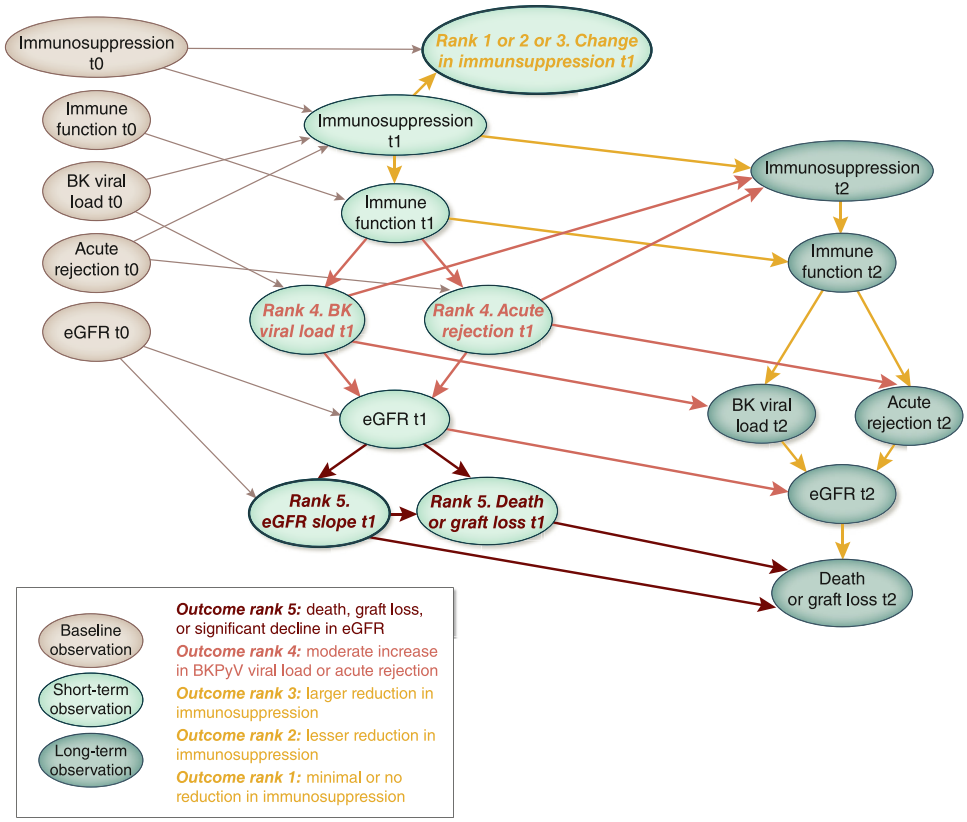

Figure 2 | A dynamic directed acyclic graph (DAG) that guides selection of valid endpoints in a planned clinical trial for the evaluation of intravenous immunoglobulin (IVIG) in the management of BK polyomavirus (BKPyV). Rank 1 indicates the best patient status, and rank 5 indicates the worst. In particular, rank 5 is defined as death, graft loss, or significant decline in estimated glomerular filtration rate (eGFR); rank 4 is defined as a moderate increase in BKPyV viral load or acute rejection; ranks 3, 2, and 1 are defined as larger, moderate, and minimal or no reduction in immunosuppression, respectively. In Supplementary Table S1, we provide details on how each variable in the dynamic DAG is mapped with variables in the putative DAG (Figure 1). t, time.

observed; alternatively, immunosuppression doses will be reduced if high viral loads are detected. We have included 2 summary variables to capture changes in eGFR and immunosuppressive therapy between the baseline and short-term time steps. The short-term risk of death or graft loss is influenced by both the absolute eGFR (as a measure of graft function) and the eGFR rate of decline—that is, the eGFR slope (t1). Short-term and progressive decline in kidney function (eGFR slope t1 and eGFR t2) determine longer-term allograft and patient survival (death or graft loss t2).

Our objective is to use this causal DAG to guide the selection of important and valid patient endpoints for use in the evaluation of IVIG for the treatment of BKPyV in a clinical-trial setting. From a pragmatic viewpoint, endpoints must be measurable (observable) and must occur (observed) within a relatively short timeframe (t1), while also allowing inferences to be made regarding long-term patient-relevant outcomes that are not immediately observable—that is, death or graft loss within a relevant timeframe, such as 10 years (t2). We propose these sets of endpoints to be used to assess the recipient's status within a short timeframe, and to compare patient outcomes between different trial arms.

To address the various possible combinations of the composite measures, outcomes have been consolidated into a single objective ranking. In Figure 2, rank 1 indicates the best patient status, and rank 5 indicates the worst; then, the worst-ranked endpoint for a patient is taken to be the overall ranking of their set of endpoints. We also labelled the causal pathways depicting how each short-term endpoint (t1) is assumed to be related to downstream outcomes of interest (i.e., death or graft loss t2). These ranks are informed by their causal proximity to (and therefore presumed influence on) the

longer-term outcomes of interest. For example, the short-term endpoints of death, graft loss, or significant decline in eGFR are assessed to be strongly influential for the long-term outcomes[9] (via dark-red, light-red, and orange arcs) and are assigned the worst rank (rank 5). If none of the rank-5 events occurs, then the short-term events that are most causally proximal to eventual graft loss and death are assigned rank 4. For example, either a moderate increase in BK viral load and/or concurrent acute rejection is likely to increase the risk of allograft dysfunction (indicated by an eGFR reduction) and thereby increase the risk of graft loss and death in the longer term (via light-red and orange arcs). Moreover, in the absence of rank-5 and rank-4 events, reduction in immunosuppression may increase the risk of subsequent rejection (via orange arcs), leading to allograft dysfunction. In the case in which immunosuppression reduction is required to control the BKPyV viral load, a higher risk of acute rejection is expected. Therefore, a larger reduction in immunosuppression is assigned a higher rank (rank 3) than a lesser reduction (rank 2). The lowest risk of death or graft loss in the longer term is expected to occur in transplant recipients with suppressed BK viral loads, with minimal or no reduction in immunosuppression, without evidence of acute rejection, and without significant short-term eGFR declines (rank 1).

Discussion

Challenges and opportunities for implementation. In this editorial, we present the creation of a causal DAG that offers a framework for assessing the effects of IVIG for post-transplant BKPyV viremia. This DAG informed a 5-point ranked short-term endpoint to derive inferences about important long-term outcomes (e.g., graft loss or death) that are impractical to observe within the trial duration. The causal DAG was necessarily a simplification of the problem domain for our specific context, and it is likely to require modification and extension before it can be used appropriately for other purposes.

The proposed causal DAGs are constrained to variables considered most relevant to our specific problem domain (managing BKPyVAN) and context (clinical trials) based on the opinions of our group of subject-matter experts. We acknowledge that a great many external factors also affect the outcomes of transplant recipients. In the context of a clinical trial, randomization is used to ensure that such factors do not confound the measured treatment effects (i.e., that these unknown or unmeasured factors are balanced across the interventions). Outside of a trial, the potential for those factors to confound measured treatment effects needs to be considered, and causal DAGs provide a coherent strategy for making such considerations. In other words, our DAGs do not aim to represent an exhaustive list of factors that need to be considered for every context. Rather, publishing causal DAGs such as ours offers the following: (i) a knowledge base to facilitate scientific reasoning and debate when experts hold diverging opinions; (ii) a framework for extension and adaptation when other considerations become important—for example, in studies in which treatments are observed rather than assigned, to account for variation in practice across healthcare settings, or when new medical knowledge and technologies become available. The knowledge base can be used to guide data collection, analysis, and interpretation, inform the design of clinical studies including trials, and even guide the development of clinical-decision support tools.[8] Causal DAGs can be extended naturally to inform quantitative models that aim to specify and test the strength of the causal effects depicted in the DAGs; moreover, quantitative approaches can serve as complementary methods to explore alternative causal hypotheses.

In our DAG, almost all key events are "latent" (for example, *graft function*); thus, rendering measurement of the latent status of interest is difficult. In these situations, observable variables, such as surrogate measures, are quantified to deduce the values of the latent variables. Given that the relationship between the surrogate and the latent event of interest is not deterministic, a measured treatment effect on a surrogate endpoint does not equate necessarily to a clinically meaningful effect on a latent outcome of interest. A causal DAG is applied to understand such discrepancies and inform opportunities to improve the selection and timing of observable measures (i.e., data collection). The measurable endpoints (e.g., viral load and eGFR) and outcomes of interest (e.g., BKPyV, acute rejection, and graft function) are likely to vary dynamically over the lifecourse of a transplant recipient, and the strength of any relationship between a long-term outcome and its short-term surrogate is expected to diminish with increasing time gap.

The choice of the desired duration of observation remains uncertain. Selection of the short-term surrogate endpoints, and determining when to measure them, involves an implicit trade-off between trial costs and the timeliness of evidence, on one hand, and the certainty that any measured effect will reflect a clinically meaningful benefit on the other. Quantitative exploration of these scenarios through data simulation may promote a more explicit approach to our assumptions, and thereby help to address these methodological challenges and guide trial design.

Finally, integration of patient-reported endpoints within the causal pathway is crucial,[10] as patient experience and symptom burden are of primary interest. Valid assessment of these symptoms (or symptom clusters) is important because it provides guidance that can be used to evaluate the impact of interventions and inform best practice in clinical management. One of the limitations in assessing treatment and symptom burden using patient-reported outcomes is the lack a well-designed and validated patient-reported outcomes instrument that is specific to transplantation. Such an instrument will provide both important insights into the observed effects in treatment comparisons, and the ability to detect minimally important differences between interventions in transplantation trials. Reliable patient-reported outcome instruments for collecting relevant and important patient data are essential elements of quality clinical care and should be the focus of future research.

Causal inference beyond the context of polyomavirus, kidney transplantation, and clinical trials. Although the use of causal inference framework is attracting attention in the medical field, the literature has gaps regarding how best to create causal DAGs, as well as a lack of protocols to guide their development in a replicable manner. The task of co-creating a robust causal DAG can become increasingly challenging for subject-matter experts and modellers as the problem domain (and its associated data) gets more complex. We believe that the validity of any DAG rests heavily on the building process itself, the documentation of this process, and the description of outcome(s). For example, we highlight the importance of a variable dictionary (Supplementary Table S1), in which variables and their relationships (i.e., nodes and arcs) are explicitly defined and documented, providing explanations that cannot be intuited from the DAG itself. This approach allows the DAG to be challenged by other domain experts, and it allows the DAG to serve as a knowledge base that can be updated with transparency and traceability as the scientific knowledge in a domain advances.

DISCLOSURE

JAM reports participation on the following Data Safety Monitoring Board (DSMB): NEXT Kidney DSMB—for which neither their institute (Telethon Kids Institute) nor they received any payments. The other authors declared no competing interests.

SUPPLEMENTARY MATERIAL

Supplementary File (Word)

Supplementary Methods. Knowledge-elicitation survey questions.

Supplementary Table S1. Variable dictionary.

REFERENCES

1. Chapman JR. What are the key challenges we face in kidney transplantation today? *Transplant Res*. 2013;2(suppl 1):S1.
2. Stolt A, Sasnauskas K, Koskela P, et al. Seroepidemiology of the human polyomaviruses. *J Gen Virol*. 2003;84:1499–1504.
3. McLeod C, Wood J, Mulrennan S, et al. Preferred health outcome states following treatment for pulmonary exacerbations of cystic fibrosis. *J Cyst Fibros*. 2022;21:581–587.
4. Tong A, Gill J, Budde K, et al. Toward establishing core outcome domains for trials in kidney transplantation: report of the standardized outcomes in nephrology–kidney transplantation (SONG-Tx) consensus workshops. *Transplantation*. 2017;101:1887.
5. Myint TM, Chong CHY, Wyld M, et al. Polyoma BK virus in kidney transplant recipients: screening, monitoring, and management. *Transplantation*. 2022;106:e76–e89.
6. Mitra N, Jason R, Dylan S. The future of causal inference. *Am J Epidemiol*. 2022;191:1671–1676.
7. Tennant PW, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50:620–632.
8. Mascaro S, Wu Y, Woodberry O, et al. Modeling COVID-19 disease processes by remote elicitation of causal Bayesian networks from medical experts. *BMC Med Res Methodol*. 2023;23:76.
9. Clayton PA, Lim WH, Wong G, et al. Relationship between eGFR decline and hard outcomes after kidney transplants. *J Am Soc Nephrol*. 2016;27:3440–3446.
10. McLeod C, Norman R, Wood J, et al. Novel method to select meaningful outcomes for evaluation in clinical trials. *BMJ Open Respir Res*. 2021;8:e000877.