

Closing the Gap: How Better Data Can Transform Kidney Injury Detection



In July 2024, I accepted a data manager position within Critical Path Institute’s Data Collaboration Center, unaware of the fact that less than a year later, the knowledge that I gained would have a direct impact on my family. Having never worked in the broader clinical research ecosystem, I underestimated just how complex and resource-intensive data collection is, along with the substantial effort required to curate and share that data in clinical settings. I carried a number of naïve assumptions, chief among them that the pharmaceutical industry operated with effectively unlimited resources and already possessed comprehensive datasets across nearly every disease area. That perspective quickly shifted. I soon came to understand the complexity, fragmentation, and collaborative effort required to generate high-quality clinical data, and the critical role that neutral conveners like C-Path play in enabling its aggregation

and use.

I was assigned to a consortium within C-Path that was working with kidney data — specifically drug-induced kidney injury or DIKI. The goal of our largest project was to promote the use of a panel of non-standard-of-care clinical biomarkers. Our hypothesis was that these biomarkers would be able to inform decision-making regarding whether a patient had undergone acute kidney injury significantly faster than the standard-of-care biomarkers.

From the beginning of my time with the group, the experts had stated that our work was underpowered; thus, acquiring additional relevant data would strengthen our conclusions. I saw how the number of subjects contributing data to our studies fell woefully short in comparison to the originally planned number of subjects due to problems with recruitment, incomplete participation, and not meeting protocol criteria, amongst other things. There were many obstacles to accruing data. Regardless of the reasons and obstacles, it became painfully clear that collecting quality data wasn’t as easy as I had originally thought.

I thought, “Well, we should simply find more relevant studies that have previously been conducted. How hard could that be?” I quickly realized it wasn’t as easy as just going out and getting more data. Those “comprehensive datasets across every disease area” in my imagination didn’t really exist, and if they did, they certainly weren’t accessible to the public. Speaking to partners at pharmaceutical companies and in regulatory agencies like the U.S. Food and Drug Administration (FDA), I realized that there are many legal hurdles to sharing data. And that’s after studies, which can take years and millions of dollars, are completed. In addition, the specific details of the study must align and support the analysis being completed. Just because one company recorded kidney data in a study, doesn’t mean that any analysis involving kidneys will be able to use those data points. As medical science advances, new biomarkers emerge. Assays must be

created to measure these new biomarkers and data must be painstakingly accrued. There is no juggernaut organization, even in 2026, that just magically has all the required data or possesses all the money required to obtain all the vital data needed to inform all decisions. This is why pooling clinical data across multiple sources is so important in advancing medical research and improving patient care.

This realization became very personal in the spring of 2025. My father was suffering from atrial fibrillation. During his treatment, his cardiologist found blockages in his arteries which required coronary bypass surgery.



Andrew, enjoying some quality time with his father.

Thankfully, my father's surgery was successful. But two and a half days later, we received word that his serum creatinine level had increased to more than twice his baseline level. It wasn't until the third day after his surgery that a nephrologist entered the scene. I made it a point to have a conversation with the nephrologist to discuss the novel biomarkers I was working with and their use. We had a great conversation about the biomarkers' use in detecting acute kidney injury. While she had some knowledge of a few biomarkers specifically, she was unaware of the application of the panel and its relevance to early detection of acute kidney injury.

I was frustrated to know that solutions for earlier detection of kidney injury existed, but my father's nephrologist didn't know about them. The biomarker panel that I was working with could identify acute kidney injury in hours versus the clinical standard of serum creatinine, which could take days. Thankfully, the damage to his kidneys was minor and reversible. My father was fortunate. But how many situations play out in which the patient is less fortunate? And in how

many of those situations would learning a patient is experiencing acute kidney injury within hours, rather than days, make a real difference?

With this experience fresh in my mind, I was shown in real time that C-Path's Biomarker Data Repository is a great tool for getting data into the hands of people who need it — first, to qualified researchers, who make the information more accessible to others. Additionally, it can serve as an educational tool, bringing real impact directly to clinicians and patients alike. This kind of access to knowledge matters, because patients should be able to look at their lab data and know when something isn't right.

I'm grateful to be doing this work with C-Path — helping ensure that the data needed to detect kidney injury earlier, and protect patients like my father, is available to the people who can turn it into better care.

Want to learn more about BmDR's life-changing work?

[Get in touch](#)

Andrew Poalilo is a Data Manager II with C-Path's Data Management team. Spending well over a decade in the classroom taught him that the hardest part of working with complex information isn't understanding it — it's communicating it. Over a 15-year career as a high school mathematics educator, Andrew built curriculum, created assessments, mentored students, and developed a discipline for breaking down

complexity that no amount of technical training can replicate. Those insights drive everything he does as a clinical data manager and analyst today.

Andrew manages the full data lifecycle for his research consortium, serving as the primary point of contact for all data-related matters across internal teams and external partners. His work is centered on data transfer, cleaning, and transformation with a focus on SDTM standards compliance, using R as his primary tool alongside Python and SAS. He also oversees the ingestion of curated datasets into relational databases and maintains metadata curation across external-facing data platforms, ensuring data is both discoverable and accurately represented for downstream consumers.

What sets his approach apart is his ability to bridge the technical and the human. He does this by aligning datasets to CDISC standards, developing and delivering SDTM training on the effective presentation and structure of clinical data, and communicating findings to stakeholders with varying levels of data literacy. He is driven by clarity, precision, and the belief that well-managed data leads to better decisions.