How Do Data Sources Inform the Choice of Endpoints and Outcomes in Pragmatic Trials?

Devon Check:

With respect to data sources for endpoints and pragmatic trials, as nicely summarized in a JAMA paper by Weber and colleagues from several years back, the first challenge in using big biomedical data effectively is to identify what the potential sources of healthcare information are and to determine the value of linking these together.

So as Weber suggests, it's often necessary to link across data sources.

Specifically in pragmatic trials, we may need to link across EHR and insurance claims databases. For example, perhaps at baseline we need some more detailed clinical data from the EHR, but then we want to be able to assess utilization outcomes longitudinally. So to do that reliably, we need insurance claims data to capture utilization that might occur across health systems and outside of the health system, for which we have the EHR data.

We're often dealing with multiple health systems, EHRs and multiple payers claims databases.

For example, we can't just rely on Medicare claims if our population of interest includes younger adults as well, who would not be covered on Medicare.

Although linking with claims is often necessary to fully capture all of the care that we're interested in assessing in the pragmatic clinical trial, it's not necessarily straightforward from a technical or legal perspective. In particular, patient consent might be required to leverage data from claims, and that sometimes proves to be a substantial, if not insurmountable barrier.

All to say that it's really a balancing act in terms of measurement and pragmatism.

While patient-reported outcomes are more expensive, less efficient, in other words, less pragmatic, they are often the most salient outcomes from a stakeholder perspective, and therefore, really important to collect as pragmatic clinical trials.