The article "Population-Level Administrative Data: A Resource to Advance Psychological Science" by Leah S. Richmond-Rakerd and colleagues (2024) highlights the potential of administrative data for advancing psychological research. Administrative data refers to the information collected as individuals interact with various public systems, such as healthcare, social welfare, criminal justice, and education. The authors argue that these datasets are underutilized in psychology compared to other fields like economics and public health, despite their immense potential to deepen understanding of human behavior, development, and well-being.
The article identifies three key advantages of population-level administrative data: its comprehensive and low-burden collection, large scale, and the possibility of unique linkages across domains. The authors emphasize that these datasets are collected without burdening participants or incurring high research costs, and they avoid biases common in self-report or survey data. Administrative data also often span many years, allowing researchers to examine long-term outcomes and track the impact of early life experiences across the lifespan. For example, researchers have leveraged decades of Danish register data to explore the long-term effects of depression on dementia risk.
Another strength is the large scale of administrative datasets, which makes it possible to study rare conditions and marginalized populations. This is crucial for psychological research, as smaller cohort studies often lack sufficient statistical power to investigate low-frequency disorders like schizophrenia. Administrative datasets also provide opportunities to examine the experiences of individuals facing adversities such as parental incarceration or foster care, which may be underrepresented in traditional research. Additionally, the size of these datasets facilitates natural experiments, allowing researchers to infer causal relationships in ways that would be unethical or impractical in a lab setting.
The authors also discuss the benefits of linking administrative data across domains. For example, data from health, education, and criminal justice systems can be integrated to explore how childhood experiences influence adult outcomes. Administrative data can be linked across generations, allowing for the study of cross-generational effects and sibling comparisons, which can help control for familial confounds. This kind of linkage has been used to better understand the intergenerational transmission of risk factors, such as the role of maternal smoking during pregnancy.
However, the article also highlights ethical considerations in using administrative data. While much of this data is deidentified, researchers must be cautious about potential re-identification risks when combining different datasets. The article advocates for transparency in data use and ensuring that research using administrative data maintains a "social license," meaning it is acceptable to the public and stakeholders. The authors also warn against the potential for findings to generate stigma, particularly when studying vulnerable populations or sensitive behaviors like mental health disorders or criminality.
Methodological limitations of administrative data include the fact that these records often only capture individuals who interact with specific services, which may lead to biases in estimating the prevalence of conditions like alcohol use disorder. Additionally, administrative data might reflect the norms and practices of dominant cultures, which can overlook the experiences of marginalized groups. To address these challenges, the authors suggest linking administrative data with survey and cohort study data to provide a richer understanding of psychological processes.
In terms of future directions, the authors call for greater cross-national comparative studies to assess the consistency of psychological outcomes across cultures, as well as the development of administrative data resources in low- and middle-income countries. They also encourage psychologists to engage with ongoing data-linkage initiatives and to incorporate administrative data into randomized experiments to strengthen causal inference. Lastly, the article highlights the need for more training in large-scale data analysis and ethical issues related to administrative data research, noting that interdisciplinary collaborations can enhance the use of these datasets in psychology.
In conclusion, Richmond-Rakerd and colleagues argue that administrative data is a valuable, underutilized resource for psychological science, offering unique opportunities to study behavior, development, and well-being on a population level. The authors advocate for increased familiarity and use of these datasets in psychology to further advance the field and improve understanding of complex psychological processes.