Harnessing administrative data to study health inequality

Homelessness, substance abuse, and psychosis are enduring challenges that tend to occur in tandem at the bottom of the social structure. Matters are often made worse by public policies that manage society’s most vulnerable via correctional facilities rather than care facilities. Simultaneous exposure to such a complex configuration of pathogenic factors is plausibly associated with poor health and shortened lifespans. However, due to a lack of access to relevant data, the extant scientific literature has struggled to empirically substantiate the impacts of co-occurring homelessness, opioid dependence, psychosis, and criminal justice involvement on health and health inequality. These impacts have proved difficult to map at the population level because commonly used data sources are subject to various biases, including high non-response and attrition rates.

In The Lancet Public Health, Emily Tweed and colleagues step into this breach by offering a novel population-based analysis using linked administrative datasets from multiple sectors in Glasgow, UK, between 2010 and 2019. The Article makes at least two important contributions. First, it demonstrates how administrative big data can be fruitfully harnessed for purposes of population health research. In contrast to small-scale survey data, administrative data—ie, data derived from the day-to-day operation of public agencies to facilitate the administration of welfare programmes—involve the assemblage of comprehensive data files comprising virtually entire populations over time. By use of records from homelessness, criminal justice, and health services as well as mortality records disaggregated by cause of death, Tweed and colleagues provide their analysis with an unusually rich and robust empirical anchoring.

Second, on this basis, they generate credible estimates of mortality gaps between those with zero, one, or multiple exposures of interest in a population cohort comprising more than half a million individuals. Specifically, they find consistently higher age-adjusted mortality risks among those with at least one exposure than for the unexposed reference group (for example, homelessness plus other exposures vs no exposures: hazard ratio 8.4 [95% CI 7.3–9.5]; homelessness alone vs no exposures: hazard ratio 2.2 [1.9–2.5]). These findings are unsurprising but nonetheless important to document in detail, especially because the authors also unveil important health gradients in previously understudied domains, such as non-communicable disease burdens. The study provides a much-needed evidence base that improves our knowledge and understanding of the needs of these uniquely vulnerable populations and offers a basis for future research and intervention.

The Article represents an important example of descriptive epidemiology. To improve population health and reduce inequalities, the next step—and the remaining challenge for future research—is to identify the mechanisms by which the observed empirical patterns are generated. For this purpose, administrative big data can be used to derive compelling inferences geared towards uncovering causal associations rather than simple correlational patterns. For example, when assessing the causal impact of a welfare benefit programme on population health outcomes, large sample sizes facilitate the construction of appropriate and statistically well-powered comparison groups that obviate the need for methodologically convoluted adjustments for confounding and selection bias. Additionally, greater statistical power allows researchers to study important but relatively rare events, such as the co-occurrence of multiple adverse exposures (such as the ones described in this study) or wide-ranging policy reforms that impact some potentially vulnerable individuals but not others. Moreover, the inherently longitudinal nature of administrative data allows researchers to gather many linked observations over time across a wide range of key variables. This has the virtue of capturing temporal sequences that are central to social epidemiological enquiries, including individual life course trajectories or intergenerational transmissions of social advantages or disadvantages.

Central to such causal knowledge is also specificity: knowing that variable X has an aggregate causal impact on variable Y is useful but insufficient for devising suitable health policies. As acknowledged by Tweed and colleagues, the causal forces at work at the interface of disease epidemics and social inequality are complex: they are operant at the level of individuals, neighbourhoods, and local communities; certain interventions can help some but hinder others; and the
underlying determinants of health inequality might differ from one location to another. Future research should therefore seek to unpack this heterogeneity and specify the conditions under which it can be appropriately addressed.

Overall, this novel study breathes fresh understanding into crucial questions that have only partially been answered by previous scholarship. The Article shows that, if suitably repurposed for scientific research, administrative data have the potential to fuel high-quality empirical investigations. Hopefully, the work by Tweed and colleagues paves the way both for rich descriptive accounts, as provided here, and equally rich causal analyses that will inform future public policy.

I declare no competing interests.

Elias Nosrati
elias.nosrati@fafo.no

Fafo Institute for Labour and Social Research, Oslo 0650, Norway

1 Nosrati E, Marmot M. Punitive social policy: an upstream determinant of health. Lancet 2019; 394: 376–77.
2 Tweed EJ, Leyland AH, Morrison D, Katikireddi SV. Premature mortality in people affected by co-occurring homelessness, justice involvement, opioid dependence, and psychosis: a retrospective cohort study using linked administrative data. Lancet Public Health 2022; published online July 27. https://doi.org/10.1016/S2468-2667(22)00159-1.
3 Card D, Chetty R, Feldstein M, Saez E. Expanding access to administrative data for research in the United States. National Science Foundation. White Paper. September 2010. https://rajchetty.com/wp-content/uploads/2021/04/NSFdataaccess.pdf (accessed July 12, 2022).
4 Connelly R, Playford CJ, Goyle V, Dibben C. The role of administrative data in the big data revolution in social science research. Soc Sci Res 2016; 59: 1–12.