International Journal of Population Data Science

Journal Website: www.ijpds.org

Yin and yang: why big data needs little data

Wright, John^{1*}

¹Bradford Institute for Health Research

Objective

Big data and data linkage offer great potential for improving health, but the shallowness of much routine data is a major limiting factor. We explore how connecting wide routine data (big data) and deep research data (little data) can harness the real potential of data linkage.

Approach

We have linked routine clinical data from education and health (primary and secondary care) for a well-characterised birth cohort (Born in Bradford) with phenotype and genotype data on almost 14,000 families. We explore the potential for this combination of big and small data to address key research priorities in health and education research.

Results

We present examples of the complementarity of routine and research data linkage in four varied domains:

1. Health care: how does postnatal mental health need (small data) match with mental health demand (big data)?
2. Education: how do early life exposures (small data) influence school readiness and standardised assessment tests (big data)?
3. Genetics: what is the impact of rare mutations (small data) on health service uptake (big data)?
4. Public health: how can big data and small data be used to evaluate the effectiveness of early life interventions?

Pros and cons of both big data and small data are identified. Some lifestyle and demographic factors are more likely to accurate from bespoke research data collection, but clinical and educational measures may be better gleaned from routine

records. The reliability of the different sources of data is discussed.

Conclusions

Our results illustrate the symbiosis of combining research and routine datasets. Opportunities for harnessing this power through combining routine data with cohort studies, clinical trials and national surveys are explored.

*Corresponding Author:

Email Address: john.wright@bthft.nhs.uk (J. Wright)