

Executive summary

Understanding youth transitions requires information on young people's individual background characteristics and the circumstances under which they grow up. In fact, the ability to assemble information about family and community background, physical health and psycho-social development, as well as academic achievement and the broader school environment, into a coherent data stream, from infancy right through to adulthood, is invaluable for developing effective policy settings. However, no single data source in Australia currently provides coverage of young people's developmental trajectories from birth and early childhood, to tertiary education and entry into the labour market.

One option for addressing the lack of life-course data is to link an existing flagship youth survey such as the Longitudinal Surveys of Australian Youth (LSAY) to existing administrative collections, such as the National Assessment Program – Literacy and Numeracy (NAPLAN), Medicare Australia and others. Data linkage refers to the process of matching records on the same person held in different data sources, such that the different sources are combined to present more comprehensive information on individuals. With an increasing number of Australian and international research projects capitalising on the advantages of data linkage, the idea of supplementing LSAY with data from administrative collections is well worth exploring.

Another option for creating life-course data from existing sources is to enhance LSAY with information from the Longitudinal Study of Australian Children (LSAC). While both surveys collect data on background characteristics and key life events, they do so for different sets of individuals and across different age groups. LSAY could be complemented with information from the Longitudinal Study of Australian Children via 'statistical matching', whereby individuals from both surveys who are statistically equivalent on a number of key background characteristics are merged into a fictitious individual to observe the impact of socio-demographic attributes and key interventions on transition outcomes over time.

The purpose of this discussion paper is to evaluate the feasibility of both approaches. In the first part of the report, the potential for enhancing LSAY with information from administrative collections through data linkage is investigated. The second part explores the viability of combining relevant data from LSAY and the Longitudinal Study of Australian Children via statistical matching.

When exploring the possibilities of data linkage with LSAY it is necessary to consider the challenges inherent in the process. These challenges revolve around technical issues, cost, and legal/ethics considerations. The latter generally represent the largest obstacle, given that legal consent needs to be sought from LSAY respondents (and possibly their parents or legal guardians) in order to proceed with any form of data linkage. Privacy regulations further require that specific protocols be followed to ensure the protection of privacy and confidentiality. These regulations include the de-identification of linked data, use of an independent agency as data custodian and integrating authority, and secure storage of linked data. The costs associated with observing privacy regulations would likely be offset by the advantages data linkage can bring to LSAY; namely, broadening the scope of the questionnaire without increasing respondent burden. Existing questions could also be supplemented with administrative data, allowing scope for new questions.

The specific administrative data sources considered for linkage with LSAY in this discussion paper are the National Assessment Program – Literacy and Numeracy, Medicare Australia, the Department of Human Services (Centrelink), the Higher Education Statistics Collection, the National VET Provider Collection and the Australian Census. While all six collections could add considerable value to LSAY, the largest initial benefit would be derived from linking the National Assessment Program – Literacy and Numeracy and Medicare data to LSAY. Centrelink data could then be linked in a second step to further enhance the breadth and depth of LSAY.

The Longitudinal Study of Australian Children provides valuable data on early childhood development (among a host of other relevant information), whereas the Longitudinal Surveys of Australian Youth focus on the transition experience from 15 years of age onwards. If it were feasible to combine relevant information from the Longitudinal Study of Australian Children and LSAY, researchers would be able to analyse a powerful dataset that captures aspects of a person's developmental trajectory from birth up to about 25 years of age.

Given that LSAY and the Longitudinal Study of Australian Children do not contain the same individuals, a method known as 'statistical matching' would have to be employed to combine both surveys. Statistical matching combines records of individuals who are statistically similar on key characteristics and which are available in both datasets. From a conceptual perspective, it is important to understand that combining LSAC data with LSAY via statistical matching would result in a synthetic dataset, in which each matched record represents a fictitious individual who proxies the combined trajectory of two real individuals who are statistically equivalent on a number of key characteristics.

Although statistical matching can be useful in certain situations, this report concludes that it is not advisable to combine information from LSAY and the Longitudinal Study of Australian Children into a synthetic dataset for further analysis. A statistical match between LSAY and the Longitudinal Study of Australian Children is an interesting empirical exercise, yet the methodological obstacles are such that any results from an analysis of a matched Longitudinal Study of Australian Children–LSAY dataset would lack the necessary robustness to inform policy and practice in meaningful ways.

The overall conclusion from this discussion paper is that strong consideration should be given to concrete plans for linking administrative collections to LSAY, beginning with the National Assessment Program – Literacy and Numeracy and Medicare data. Once a process for data linkage has been developed for LSAY, linking with either Centrelink data or the Australian Census could be investigated, based on a detailed cost–benefit analysis.