Journal of Modern Applied Statistical Methods

Volume 18 | Issue 1 Article 17

4-7-2020

A Glossary on Building Longitudinal, Population-Based Data Linkages to Explore Children's Developmental Trajectories

Jennifer E. V. Lloyd
University of British Columbia, jennifer.lloyd@ubc.ca

Jacqui Boonstra
University of British Columbia

Lisa Chen
University of British Columbia

Barry Forer
University of British Columbia

Ruth Hershler
University of British Columbia

See next page for additional authors

Recommended Citation

Lloyd, J. E. V., Boonstra, J., Chen, L., Forer, B., Hershler, R., Milbrath, C., Poon, B. T., Razaz, N., Rowcliffe, P., & Schonert-Reichl, K. (2019). A glossary on building longitudinal, population-based data linkages to explore children's developmental trajectories. Journal of Modern Applied Statistical Methods, 18(1), eP3180. doi: 10.22237/jmasm/1556670060

A Glossary on Building Longitudinal, Population-Based Data Linkages to Explore Children's Developmental Trajectories

Cover Page Footnote

Thank you to the Canadian Institutes of Health Research (CIHR), Michael Smith Foundation for Health Research (MSFHR), the Djavad Mowafaghian Foundation, the Social Sciences and Humanities Research Council (SSHRC), Population Data BC, Edudata Canada, the BC Ministries of Health, Education, and Child & Family Development, BC's Vital Statistics Agency, and Statistics Canada for supporting HELP's developmental trajectories (DT) program of research. Thanks also to Joanne Schroeder, Dr. Jean Shoveller, and Emilia Walton for their insights on an earlier draft. This manuscript is dedicated to the memory of Dr. Clyde Hertzman.

Authors

Jennifer E. V. Lloyd, Jacqui Boonstra, Lisa Chen, Barry Forer, Ruth Hershler, Constance Milbrath, Brenda T. Poon, Neda Razaz, Pippa Rowcliffe, and Kimberly Schonert-Reichl

Journal of Modern Applied Statistical Methods
May 2019, Vol. 18, No. 1, eP3180.
doi: 10.22237/jmasm/1556670060

בס"ד

ISSN 1538 − 9472

A Glossary on Building Longitudinal, Population-Based Data Linkages to Explore Children's Developmental Trajectories

Jennifer E. V. Lloyd
University of British Columbia
Vancouver, British Columbia

Jacqui Boonstra
University of British Columbia
Vancouver, British Columbia

Barry Forer
University of British Columbia
Vancouver, British Columbia

Rush Hershler
University of British Columbia
Vancouver, British Columbia

Constance Milbrath
University of British Columbia
Vancouver, British Columbia

Brenda T. Poon
University of British Columbia
Vancouver, British Columbia

Neda Razaz
University of British Columbia
Vancouver, British Columbia

Pippa Rowcliffe
University of British Columbia
Vancouver, British Columbia

Kimberly Schonert-Reichl
University of British Columbia
Vancouver, British Columbia

Population-based, person-specific, longitudinal child and youth health and developmental data linkages involve connecting combinations of specially-collected data and administrative data for longitudinal population research purposes. This glossary provides definitions of key terms and concepts related to their theoretical basis, research infrastructure, research methodology, statistical analysis, and knowledge translation.

Keywords: Developmental trajectories, data linkage, longitudinal analysis, population health, early child development

Introduction

Population-based, person-specific, longitudinal child and youth health and developmental data linkages (PPL data linkages) involve connecting combinations of specially-collected data and administrative data for longitudinal population research purposes. Such linkages provide state-of-the-art opportunities to explore children's developmental trajectories – that is, changes in health, development, and well-being over time – at the level of the population.

doi: 10.22237/jmasm/1556670060 | Accepted: August 11, 2019; Published: April 7, 2020.
Correspondence: Jennifer E. V. Lloyd, jennifer.lloyd@ubc.ca

As a leading example from Canada, the University of British Columbia's Human Early Learning Partnership (HELP) has linked together individual-level data from four sources: (1) HELP's Early Development Instrument (EDI) school readiness records collected from 172,221 Kindergarten children at age 5/6 years (comprising 12 annual cohorts, representing at least one population-based cohort of >90% of Kindergartners from British Columbia's 60 school districts); (2) British Columbia Ministry of Education grade school and standardised assessment records, to allow us to investigate school attainment of children in the EDI cohorts; (3) British Columbia Ministry of Health; and (4) BC Vital Statistics Agency health, birth, and death records, to allow us to explore birth-through-school age health trajectories of children in the EDI cohorts. These linkages make British Columbia one of the first jurisdictions in the world to have such extensive data describing the early life course of children. Some examples of output from these linked data include: Gagné et al. (2018), Guhn et al. (2019), Guhn et al. (2016), Oberle et al. (2014), Thomson et al. (2017), and Thomson et al. (2019).

Although PPL data linkages are becoming increasingly available, they are novel to many research settings not only in population health, but also in education and the social sciences. For this reason, readers interested in creating their own PPL data linkages are provided with this glossary. First, key terms relating to the theory underlying the construction of PPL data linkages are defined. Then, terms relating to the research infrastructure required to construct PPL data linkages are defined. Finally, terms relating to post-construction aspects of PPL data linkages are defined – specifically, research methodology, statistical analysis, and knowledge translation.

Theoretical Basis

Causal Inference

An association between an exposure and an outcome without confounding or bias. In epidemiology, establishing causality is key in planning public health interventions aimed at modifying the probability of the targeted health outcome (Szklo & Nieto, 2006). Hill (1965) provides a set of guidelines about the minimum conditions required to establish a causal relationship between an exposure and an outcome.

Developmental Trajectory

Distinct courses or pathways of a particular behavior or characteristic over time, whether it be biological or social in nature (Hertzman & Boyce, 2010). Nagin (2005) refers to a developmental trajectory as the "evolution of an outcome over age or time" (p. 1).

Human Development

The dynamic interplay between biological and social factors, and how this interplay shapes human health, education, well-being, and development, from infancy through old age, and across diverse social and geographical contexts. Human development is innately longitudinal and encompasses factors from multiple levels of social aggregation (Hertzman & Boyce, 2010).

Interdisciplinary Research

Because PPL linkages involve working with intersectoral partners from varied backgrounds and disciplines, their creation is an inherently interdisciplinary endeavor. The Committee on Facilitating Interdisciplinary Research and Committee on Science, Engineering, and Public Policy (2004) define interdisciplinarity as "a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice" (p. 2).

Life-Course Epidemiology

An approach for studying the intra- and inter-generational long-term effects of physical and social exposures on health and disease risk during gestation, childhood, adolescence, young adulthood, and later adulthood (Ben-Shlomo & Kuh, 2002). Particular attention is paid to studying the gradual accumulation of risk over the life course, critical periods of health and development where exposures during critical windows of time have irreversible impacts, the temporal ordering and interrelationships of exposures, and biological, behavioral, and psychosocial pathways operating across the life course to influence health and disease outcomes (Ben-Shlomo & Kuh, 2002).

Pathways

Sequences in which early life experiences influence the probability of other later-life experiences, as well as related health and developmental outcomes (Hertzman & Boyce, 2010).

Population/Subpopulation

A population is a complete set of individuals or objects of interest that share a certain property or properties (e.g., all children born in British Columbia between 2000 and 2010). A subpopulation is a particular subset of a population. A population is often more comprehensively understood if it is first separated into distinct subpopulations (e.g., male vs. female, low vs. high birth weight). PPL data linkages avoid problems related to small sample sizes, sampling variability, and an exclusive focus on high-risk sub-populations (Lloyd & Hertzman, 2009; Lloyd et al., 2009).

Research Infrastructure

Agreements

A research agreement (RA) is a legally-enforceable information/data sharing agreement between a researcher(s) and a data steward(s), which outlines the terms, conditions, and obligations associated with access and use of defined fields of data approved for a specific research project. The RA defines the research project (e.g., research questions, study design, analytic methods), the ethics review, the study population (e.g., specific demographics, date ranges, geographic areas, database sources), data security and access, use of data terms, and adherence to privacy legislation. For some data providers, the RA is submitted as one document. For others, a data acquisition request (DAR) that defines the project, study population, data security, and access is submitted first. The DAR is then subsumed into the RA upon approval by the data steward.

Coding, Recoding, Geocoding

Researchers must often transform data upon receipt from a data partner, for the relevance of the project at hand. Coding is the process of transforming data, either quantitative or qualitative, into numeric or string categories for data analytic purposes. Recoding involves transforming data into different or the same variables when the categories for a given variable in the data being analyzed need to be

collapsed into fewer meaningful categories. Recoding may also be useful if the order of the categorical data does not adequately reflect the implied meaning of the variable. Geocoding is the process of assigning specific latitude and longitude coordinates to a specific geographical point (e.g., based on street address, city, province/state, and postal/zip code) such that the point can be placed accurately on a map.

Confidentiality

Anonymized data refer to those in which any identifiable information on an individual (person, typically) has been irreversibly removed. Anonymizing individual level data preserves the confidentiality of personal information being attributed to any particular individual. De-identified data are typically individual-level data from which identifiable information has been removed. During this process, each record for this individual is typically assigned a unique study identification number. The unique identifier can then be linked back to the individual's identifying information via the study ID. This number would be attached to a separate, securely stored file that contains the identifiable information and thus allows for re-identification.

Data Security

The implementation of digital privacy measures to protect data from corruption, and from unsolicited actions of unapproved users. Given the large volumes of data involved in PPL data linkages, the safeguarding of personal information is of paramount importance. Typically, any identifiable personal information is removed from the linked data, encrypted, and stored separately from the rest of the content data that are used for the research activities. The data are stored in highly secure data storage facilities.

Data Sources

In order to study population-based developmental trajectories, one requires population-based data sources. National censuses can provide a regular snapshot of socioeconomic and demographic variables at different aggregate geographies; these include official boundaries such as census subdivisions or school districts, or custom boundaries such as those which HELP uses to define neighborhoods. Other national databases, such as income tax filer data, are available for the same aggregate geographies, often yearly. These databases are complemented by a number of large, individual-level, linkable administrative databases, such as those

routinely collected by government ministries (e.g., health, education). Non-administrative population-level databases are relatively rare, due to the effort and expense required to produce and update them [that said, the Early Development Instrument (EDI) and registries of health disorders do make possible individual linkages (properly de-identified) to trajectory-relevant information in administrative databases (Kershaw & Forer, 2010)].

Ethics

The core values, principles, and practices for conducting research that reflect overall respect for human dignity and principles of respect for persons (e.g., autonomy), welfare (e.g., risks and potential benefits), and justice (e.g., fairness and equity) (Canadian Institutes of Health Research et al., 2010). In research involving linkage of de-identified data, for example, linkage may be both a potential threat to privacy and also an opportunity to provide information for the public good (Kelman et al., 2002). Protection of privacy reflects a core principle of concern for welfare; consequently, national ethics guidelines in Canada call for researchers to describe security measures taken to safeguard the information and to assess and minimize risks for re-identification of individuals (Canadian Institutes of Health Research et al., 2010).

Meta-Data

Meta-data assist in the retrieval of information and are understood to be an especially complex undertaking as human information grows exponentially. Originally used by Bagley (1968), meta-data describe information "about" data, such as identifiers, domain prescriptors specifying the domain(s) of data values, and access codes that limit conditions of data access. Structural meta-data refer to the structure, design, and organization of the data. Descriptive meta-data describe the data content. Administrative meta-data provide data related to managing a data resource, such as information about its creation, or the property rights associated with it, as well as information needed to archive or preserve a data resource (National Information Standards Organization, 2004).

Partners

A data broker collects and sells information on individuals or companies, typically on behalf of a data steward. A data steward is responsible for the management and oversight of an organization's data holdings. This person is responsible for data usage and security policies with regard to these data holdings. In addition, the data

steward identifies data quality issues, data definitions, and is responsible for all meta-data pertaining to a particular data holding. A research 'collaboratory' is an organization whose role it is to facilitate the researchers' process of submitting research agreements and data acquisition requests to varied data stewards and/or data brokers, to assist with data security and storage, and with data education and training. Intersectoral partnerships refer to collaborations between three sectors: government, business, and civil society (e.g., non-governmental organizations, non-profits, community groups).

Passive Consent

The creation of PPL data linkages relies on passive consent. Passive consent is customary in population health studies because it is not often possible to get the active consent of a population or subpopulations of children. With passive consent, parents are fully informed about the nature of the research project and the proposed use of the linked data, but they do not have to actively complete a consent form.

Record Linkage

The process of connecting data about one population group to other relevant data on the same population in order to identify trends or patterns. Linkage is only permitted under strictly approved and controlled circumstances. Direct linkage refers to the merging of multiple data sets where each data set contains the same unique identifier (e.g., Personal Health Number, Social Insurance Number). Deterministic linkage is a technique whereby links between records are determined based on the perfect match of a set of common identifiers or the match of a subset of the identifiers. With probabilistic linkage, identifiers are given weights according to how strong a likelihood it is that a set of identifiers pertains to the same individual.

Research Methodology, Statistical Analysis, and Knowledge Translation

Commensurability

A central assumption in the study of developmental processes (Schaie et al., 1998), commensurability requires that the same or comparable behavior or characteristic has been measured across all waves (periods of data collection) of a longitudinal study. From a statistical point of view, researchers must assemble multiple sources of psychometric evidence about the validity and reliability of the scores of

longitudinal measures (forms, tests, scales, assessments) to provide empirical confirmation that commensurability has been met. Only then may longitudinal measures' scores be included in such statistical analyses as individual growth modelling and group-based trajectory modelling (Lloyd, 2010).

Contextual vs. Compositional Effects

The neighborhood effects literature commonly makes a distinction between two distinct sources of variation in outcomes of interest. Compositional effects refer to place-based differences in the characteristics of the individuals in each place (e.g., level of education). Contextual differences refer to differences in the characteristics of the places themselves (e.g., early childhood education services). That said, the composition/context distinction is often considered artificial, as many outcomes are now seen as co-created in reciprocal and non-recursive ways through complex person-place relationships (Cummins et al., 2007).

Knowledge Translation

Activities that promote utilization of knowledge (i.e., all forms of knowledge but, in this context, scientific knowledge) to develop strategies for improving society and responding to human problems (Backer, 1991). Central to knowledge translation is involvement of relevant stakeholders (e.g., researchers, decision-makers) who exchange knowledge and engage in processes of knowledge creation and application for the aim of turning knowledge into action (Graham et al., 2006). For example, the terminology and assumptions of knowledge translation are diverse and embedded in many disciplines; researcher/knowledge-user collaboration facilitates establishment of a common understanding for a shared knowledge-to-action approach (Estabrooks et al., 2006).

Longitudinal vs. Repeated Cross-Sectional Research

Longitudinal research involves following outcomes of the same individuals (or groups of individuals) over time. Because the same individuals provide repeated measurements on the outcomes of interest, both within individual (dynamic) changes as well as differences (heterogeneity) among individuals may be explored. Cross-sectional research typically is motivated by an interest in comparing independent groups with different attributes or under contrasting conditions. It provides a snapshot of different individuals or groups of individuals at a single point in time and affords a static comparison across multiple variables or conditions at a specific time or occasion. Whereas longitudinal analysis must contend with missing

data and selection biases due to (e.g.) participant attrition, cross-sectional analysis rests on the assumption that independent participants are equated/matched on variables not of interest but that may influence outcomes (Frees, 2004).

Missing Data

Missing values in PPL data linkages may be due to attrition, data collection problems, or changes in data recording. Missingness in complex data can impact time-varying and time-invariant variables (Lloyd et al., 2013). It is crucial to appropriately handle missing data because ignoring or incorrectly handling it can lead to biased parameter estimates and incorrect inferences (Enders, 2010; Schafer & Graham, 2002). Modern methods, such as maximum likelihood (ML) estimation or multiple imputation (MI), are recommended over traditional methods (e.g., list-wise deletion, mean substitution). Enders recommends ML over MI, when possible, because ML is easier, more flexible, and less time consuming to implement using structural equation models. Conversely, when linked data include missing item-level data that are summed/averaged to produce scale scores, then MI is a more flexible method. MI is also better for missingness in explanatory variables because ML results in loss of cases. MI poses unique challenges with complex data, however, because it requires specialized algorithms to preserve between-cluster differences (Enders, 2010) and cross-level interactions (Lloyd et al., 2013).

Multilevel Analysis

Inter-individual variability in outcomes of interest is likely to be a function of influences at a variety of levels. Beyond the influence of particular individual-level variables, individuals' outcomes may also be influenced by membership in groups (e.g., schools, neighborhoods), such that some proportion of the inter-individual variability is accounted for by group membership. Multilevel analyses assume two or more nested levels (e.g., individuals within schools), each with variables that may influence the overall variation in outcomes across individuals (Snijders & Bosker, 2012). Multilevel modeling, therefore, allows researchers to make inferences about the relative effects of explanatory variables across levels.

Practical vs. Statistical Differences

Practical significance refers to whether or not an obtained result is meaningful and useful (Kraemer et al., 2003). Statistical significance addresses whether or not an obtained result is due to chance. Population studies with very large participant sizes can yield statistically significant results for relationships that are quite weak and of

little practical value. Practical significance must consider the strength of the relationship as well as the real benefit, cost, potential harm, or side effects. There is no formal statistical test of practical significance, but researchers typically use one of several measures of effect size to assess the strength or size of the association as a guide (see Kraemer et al., 2003).

Spatial Analysis

An analysis that applies traditional statistical methods within a geographical context, meaning that data are assumed to be spatially dependent. Each data point has a precisely defined location (at a variety of scales), leading to analyses and models that address where phenomena or objects exist, or how these attributes form gradients over space or time (de Smith et al., 2013). Given the dependence on location, geospatial research methods typically rely on observation and natural experiments, rather than true experimental methods. Geospatial patterns are often portrayed using maps, both static and dynamic.

Typologies

Similar patterns in trajectories followed by discrete subgroups of children within a population that are not detectable at the outset of a study by such sociodemographic variables as gender or socioeconomic status (Nagin, 2005). Group-based trajectory modelling – a "finite mixture modeling application that uses trajectory groups as a statistical device for approximating unknown trajectories across population members" (Nagin & Odgers, 2010, p. 111) – is one way in which to explore such typologies.

Unit of Analysis

The entity of interest in statistical analyses. In trajectories research, the unit of analysis is typically the individual person. Important to distinguish, however, is that interpretation of statistical analyses of such trajectories occurs at varied higher-level organizations and geographies (e.g., school district, neighborhood, province), and not at the level of the individual person.

Conclusion

Longitudinal, population-based data linkages are advancing knowledge about children's developmental trajectories over time, at the level of population. The aim

in preparing this glossary was to provide a resource that defines key terms and concepts relating to the theory, construction, methodology, statistical analysis, and knowledge translation of PPL data linkages. More generally, however, it is hoped that this glossary offers readers insights into the breadth of considerations that population health researchers, working alongside their intersectoral partners, must make when undertaking projects involving PPL data linkages.

Acknowledgements

Thank you to the Canadian Institutes of Health Research (CIHR), Michael Smith Foundation for Health Research (MSFHR), the Djavad Mowafaghian Foundation, the Social Sciences and Humanities Research Council (SSHRC), Population Data BC, Edudata Canada, the BC Ministries of Health, Education, and Child & Family Development, BC's Vital Statistics Agency, and Statistics Canada for supporting HELP's developmental trajectories (DT) program of research. Thanks also to Joanne Schroeder, Dr. Jean Shoveller, and Emilia Walton for their insights on an earlier draft. This manuscript is dedicated to the memory of Dr. Clyde Hertzman.

References

Backer, T. E. (1991). Knowledge utilization: The third wave. *Knowledge, 12*(3), 225-240. doi: 10.1177/107554709101200303

Bagley, P. (1968). *Extension of programming language concepts*. Philadelphia, PA: University City Science Center.

Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology, 31*(2), 285-293. doi: 10.1093/ije/31.2.285

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council Canada, Social Sciences and Humanities Research Council of Canada, & Interagency Secretariat on Research Ethics (Canada). (2010). *Tri-council policy statement: Ethical conduct for research involving humans*. Ottawa, Ontario: Interagency Secretariat on Research Ethics.

Committee on Facilitating Interdisciplinary Research, & Committee on Science, Engineering, and Public Policy. (2004). *Facilitating interdisciplinary research*. Washington, D.C.: National Academies Press.

Cummins, S., Curtis, S., Diez-Roux, A. V., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. *Social Science & Medicine, 65*(9), 1825-1838. doi: 10.1016/j.socscimed.2007.05.036

de Smith, M., Longley, P., & Goodchild, M. (2013). *Geospatial analysis: A comprehensive guide to principles, techniques, and software tools* (4th edition). Leicester, UK: Matador.

Enders, C. K. (2010). *Applied missing data analysis.* New York: Guilford Press.

Estabrooks, C. A., Thompson, D. S., Lovely, J. J. E., & Hofmeyer, A. (2006). A guide to knowledge translation theory. *Journal of Continuing Education in the Health Professions, 26*(1), 25-36. doi: 10.1002/chp.48

Frees, E. (2004). *Longitudinal and panel data: Analysis and applications in the social sciences.* Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511790928

Gagné, M., Janus, M., Milbrath, C., Gadermann, A., & Guhn, M. (2018). Early emotional and communication functioning predicting the academic trajectories of refugee children in Canada. *Educational Psychology, 38*(8), 1050-1067. doi: 10.1080/01443410.2018.1475627

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W., & Robinson, N. (2006). Lost in knowledge translation: Time for a map? *Journal of Continuing Education in the Health Professions, 26*(1), 13-24. 10.1002/chp.47

Guhn, M., Emerson, S. D., Mahdaviani, D., & Gadermann, A. M. (2019). Associations of birth factors and socio-economic status with indicators of early emotional development and mental health in childhood: A population-based linkage study. *Child Psychiatry & Human Development, 51*, 80-93. doi: 10.1007/s10578-019-00912-6

Guhn, M., Gadermann, A. M., Almas, A., Schonert-Reichl, K. A., & Hertzman, C. (2016). Associations of teacher-rated social, emotional, and cognitive development in kindergarten to self-reported wellbeing, peer relations, and academic test scores in middle childhood. *Early Childhood Research Quarterly, 35*, 76-84. 10.1016/j.ecresq.2015.12.027

Hertzman, C., & Boyce, T. (2010). How experience gets under the skin to create gradients in developmental health. *Annual Review of Public Health, 31,* 329-347. doi: 10.1146/annurev.publhealth.012809.103538

Hill, A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine, 58*(5), 295. doi: 10.1177/003591576505800503

Kelman, C. W., Bass, A. J., & Holman, C. D. (2002). Research use of linked health data – A best practice protocol. *Australian and New Zealand Journal of Public Health, 26*(3), 251-255. doi: 10.1111/j.1467-842x.2002.tb00682.x

Kershaw, P., & Forer, B. (2010). Selection of area-level variables from administrative data. An intersectional approach to the study of place and child development. *Health and Place, 16*(3), 500-511. doi: 10.1016/j.healthplace.2009.12.008

Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(12), 1524-1529. doi: 10.1097/00004583-200312000-00022

Lloyd, J. E. V. (2010). Construct commensurability and the analysis of change. *Educational and Psychological Measurement, 70*(2), 252-266. doi: 10.1177/0013164409355689

Lloyd, J. E. V., & Hertzman, C. (2009). From Kindergarten readiness to fourth-grade assessment: Longitudinal analysis with linked population data. *Social Science & Medicine, 68*(1), 111-123. doi: 10.1016/j.socscimed.2008.09.063

Lloyd, J. E. V., Irwin, L. G., & Hertzman, C. (2009). Kindergarten school readiness and fourth-grade literacy and numeracy outcomes of children with special needs: A population-based study. *Educational Psychology*, *29*(5), 583-602. doi: 10.1080/01443410903165391

Lloyd, J. E. V., Obradović, J, Carpiano, R. M., & Motti-Stefanidi, F. (2013) JMASM 32: Multiple imputation of missing multilevel, longitudinal data: A case when practical considerations trump best practices? *Journal of Modern Applied Statistical Methods 12*(1), 261-275. doi: 10.22237/jmasm/1367382480

Nagin, D. (2005). *Group-based modeling of development.* Cambridge, MA: Harvard University Press. doi: 10.4159/9780674041318

Nagin, D. & Odgers, C. L. (2010). Group-based Trajectory Modeling in Clinical Research. *Annual Review of Clinical Psychology, 6,* 109-38. doi: 10.1146/annurev.clinpsy.121208.131413

National Information Standards Organization. (2004). *Understanding metadata.* Bethesda, MD: NISO Press.

Oberle, E., Schonert-Reichl, K. A., Hertzman, C., & Zumbo, B. (2014). Social-emotional competencies make the grade: Predicting academic success in early adolescence. *Journal of Applied Developmental Psychology, 35*(3), 138-147. doi: 10.1016/j.appdev.2014.02.004

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177. doi: 10.1037/1082-989x.7.2.147

Schaie, K. W., Maitland, S. B., Willis, S. L., & Intrieri, R. C. (1998). Longitudinal invariance of adult psychometric ability factor structures across seven years. *Psychology and Aging*, 13(1), 8-20. doi: 10.1037/0882-7974.13.1.8

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd edition). London: Sage.

Szklo, M., & Nieto, J. (2006). *Epidemiology: Beyond the basics* (2nd edition). Sudbury, MA: Jones and Barlett Publishers.

Thomson, K. C., Guhn, M., Richardson, C. G., Ark, T. K., & Shoveller, J. (2017). Profiles of children's social-emotional health at school entry and associated income, gender and language inequalities: A cross-sectional population-based study in British Columbia, Canada. *BMJ Open, 7*(7), e015353. doi: 10.1136/bmjopen-2016-015353

Thomson, K. C., Richardson, C. G., Gadermann, A. M., Emerson, S. D., Shoveller, J. A., & Guhn, M. (2019). Association of childhood social-emotional functioning profiles at school entry with early-onset mental health conditions. *JAMA Network Open, 2*(1), e186694. doi: 10.1001/jamanetworkopen.2018.6694