**Cohort methods and applications in human biology**

William Johnson[1], Tim J Cole[2]

[1]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough
[2]UCL Great Ormond Street Institute of Child Health, London

WJ is supported by a UK Medical Research Council New Investigator Research Grant (MR/P023347/1) and acknowledges support from the National Institute for Health Research (NIHR) Leicester Biomedical Research Centre, which is a partnership between University Hospitals of Leicester NHS Trust, Loughborough University, and the University of Leicester.

The authors declare that they have no conflict of interest.

Correspondence: William Johnson
School of Sport, Exercise and Health Sciences, Loughborough University, Epinal Way, Loughborough, Leicestershire, UK, LE11 3 TU
Tel +44 (0)1509 228057
Email W.O.Johnson@lboro.ac.uk

Cohort studies can broadly be defined as any non-interventional study in which a group of individuals are recruited (at any point in their lives) and followed up over time. Since the inception of the Society for the Study of Human Biology (SSHB) in 1958, cohort studies have played a central role in the advancement of many of the key pillars of human biology, including growth and development, structural morphology and physiology, ageing and functional decline, and more recently lifelong health and well-being. While previous SSHB conferences have focused on such topics, none has considered the contribution of cohort studies to the field of human biology.

The 60th conference of the SSHB, held in Oxford, United Kingdom (UK) from 9–11 September 2019, considered any abstract that demonstrated either 1) the use of cohort data from one or more studies to answer an important human biology research question (e.g., on life course processes/relationships or secular trends) or 2) methods employed within cohort studies that are relevant to human biology (e.g., statistical analysis techniques or linkage to routine health records). This special issue of the Annals of Human Biology comprises 16 articles, based on the invited talks.

**A brief history**

The first paper in the special issue is a commentary by Professor Noël Cameron on "The growth and development of cohort studies". In the first part of the article, **Cameron** provides a fascinating take on the history of cohort studies designed to investigate the epidemiology of tuberculosis in the late 19th century and early 20th century, featuring prominent scientists such as Sir Austin Bradford Hill (1897-1991) and Sir Richard Doll (1912-2005). The title of the commentary is particularly fitting, however, because longitudinal studies of child growth and development played a central role in the evolution and refinement of cohort studies over the 20th century. In the second part of the article, Cameron describes how Professor James M Tanner (1920-2010, who otherwise needs no introduction) toured the United States of America (USA) in 1947 finding a total of 16 longitudinal studies of human growth and development (Tanner 1981). Perhaps the most powerful of these was the Fels Longitudinal Study, which started in 1929 and continued until recently, providing frequent serial data from birth onwards, on up to four generations within a family, born over more than an 80-year period (Roche 1992). The UK does not have an equivalent "sequential" birth cohort study, but instead has a series of much larger nationally representative birth cohort studies initiated in 1946, 1958, 1970, and 2000 (Elliott and Shepherd 2006; Power and Elliott 2006; Wadsworth et al. 2006; Kuh et al. 2011; Hansen 2012). Against the backdrop of failed initiatives in the UK and USA to establish new mega-cohorts (e.g., the UK Life Study aimed

to recruit 80,000 births), Cameron finishes with a thoughtful discussion of the need for integrated and collaborative cohorts that increasingly operate together.

**The Tanner Memorial Medal**

The Tanner Memorial Medal is awarded for services to human biology. The recipient this year was Professor Tim J Cole, who has contributed to the success of the Society and Journal for over four decades. Cole is the world-leading authority on the analysis of child growth data and most recently has developed the SuperImposition by Translation And Rotation (SITAR) method (Cole et al. 2010). Briefly, SITAR fits a smooth mean curve and estimates a set of three random effects (representing overall size, as well as the timing and intensity of peak velocity) that together capture between-individual differences in growth. In his special issue paper, **Cole** uses longitudinal anthropometric data from the Harpenden Growth Study, which was set up by Tanner in 1949 (Tanner 1981), and the much larger Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd et al. 2013; Fraser et al. 2013). His SITAR model explained 91-99.5% of the variance in 10 linear measurements collected between 7-20 years of age in the Harpenden Growth Study, supporting Tanner's view that pubertal growth can largely be summarised by differences in the timing and intensity of peak velocity. Further, by confirming the good fit of SITAR to height data on different subsamples in ALSPAC, defined according to age and magnitude of peak velocity, Cole explicitly tests how invariant (i.e., unchanging) the height velocity curve is to differences in timing and intensity. The paper finishes with a sincere conclusion: "SITAR works well to summarise pubertal growth. The disappointment is that Tanner did not live to see it in action."

**An evolutionary perspective**

A key tenet of biological anthropology, the discipline most strongly aligned to human biology (as viewed by the Society), is evolutionary life history theory. Assuming that the resources available in any environment are finite, life history theory proposes that each organism has been selected to allocate those resources in a way that maximizes reproductive fitness (Stearns 1992; Wells et al. 2017). In the opening to his article, **Kuzawa et al** describes how life history theory provides a framework for formulating hypotheses, regarding how humans allocate energy between key body functions (maintenance, growth, reproduction, and defence), that can be tested using data from birth cohort studies. The paper goes on to provide a captivating review of two decades of life history theory-motivated work conducted in collaboration with the Cebu Longitudinal Health and Nutrition Survey in the Philippines (Adair et al. 2011). The next paper by **English et al** used ALSPAC data to investigate associations between prenatal anxiety, breastfeeding, and later growth and puberty onset.

The results show how maternal anxiety is related to a lower likelihood of breastfeeding, which in turn is related to slower infant growth and later puberty onset. English et al discuss how these findings fit with non-human studies and predictions from evolutionary life history theory that early adversity results in reduced breastfeeding and, in turn, accelerated development in offspring. The paper illustrates an accelerated life history strategy to ensure reproductive success in the face of a shorter expected lifespan.

**Nutrition and morbidity**

In Africa, The Gambia and Kenya have long-established centres of excellence in cohort research on nutrition and morbidity, among other things. The UK Medical Research Council (MRC) has invested in health research in The Gambia since 1947 and currently funds MRC Unit The Gambia at London School of Hygiene and Tropical Medicine (Hennig et al. 2017). For almost seven decades the field station in Keneba, in the rural West Kiang region, has been the epicentre of numerous, overlapping observational cohort studies, clinical trials, and demographic and health surveys. In a review paper, **Moore** provides a comprehensive description of the observational "Keneba cohorts", which since 2004 include a biobank, an electronic medical records system, and a demographic surveillance system. Moore also highlights some of the key findings and how they have informed nutrition and health policy and planning in The Gambia and other low- and middle-income countries. In Kenya, the African Population and Health Research Center runs the Nairobi Urban Health and Demographic Surveillance System (Green 2017). In his special issue paper, **Matua et al** uses these data to investigate the determinants of vaccination coverage and the subsequent impact on mortality. Children who were not fully vaccinated and those whose vaccinations occurred later than recommended had a higher mortality risk than children who were fully vaccinated on time. The findings of the paper emphasize the importance of complete and timely vaccination for protection against preventable diseases.

**Childhood obesity**

Cohort studies provide the data necessary to investigate how biological measures taken at different stages in the life course are related and might operate via intermediate variables. Rapid infant weight gain, most commonly defined as upward crossing through one major centile band on the UK growth chart (equivalent to 0.67 Z-scores) in the first few years of life, is arguably the strongest correlate of childhood obesity and is widely used in prediction models (Ong et al. 2000; Monasta et al. 2010; Druet et al. 2012). We do not, however, know how infant growth parameters, other than weight, might add to the prediction of childhood obesity. Using data from the Cambridge Baby Growth Study (Prentice et al. 2016), the special issue paper by **Ong et al** provides evidence that the prediction of childhood

percentage body fat by infant weight gain is unlikely to be largely improved upon by incorporation of infant length and skinfold thickness data. The next paper by **Norris et al** also reports a positive association between infant weight gain and childhood adiposity (and blood pressure) outcomes, this time in the Southampton Women's Study (Inskip et al. 2006). The novelty of the study, however, is that it tested whether these well-known associations were modified by fetal growth patterns. The authors found no strong evidence of effect modification by estimated fetal weight at 19 and 34 weeks of gestation, suggesting that the adverse consequences of rapid infant weight gain might occur regardless of whether growth was constrained in-utero. Finally, **Taine et al** used data from the EDEN mother-child cohort study to quantify the extent to which the association of fetal growth (proxied by neonatal anthropometry) with early infant growth was mediated by (i.e., operated via) cord blood leptin levels (Heude et al. 2016). The paper provides new evidence on the biological regulation of early postnatal growth.

**Cognitive ageing and resilience**

With an ageing population, research into healthy cognitive ageing and resilience is more relevant than ever (Kuh et al. 2014). Indeed, this is a key focus of birth cohort studies in which the participants are now reaching older age, and longitudinal studies set up more recently to specifically investigate ageing. The original research article by **Gaysina et al** presents an interesting analysis using data from the UK 1958 birth cohort study. The authors computed measures of cognitive resilience as the residuals (i.e., unexplained variance) from regressions of cognitive function variables (e.g., information processing speed and accuracy) at age 50 years on a cumulative score, representing lifetime affective symptoms according to the Malaise Inventory scale (Rutter et al. 1970). The subsequent results provide evidence that better cognitive resilience in mid-adulthood may be partly determined by higher childhood cognitive ability, education level, and midlife socioeconomic position.

**Epidemiology & data linkage**

The value for money of cohort studies is vastly improved by the ability to link to routine registries, for example, of hospital admissions and causes of death (Harron et al. 2017; Gilbert et al. 2018). The review by **Harron** includes guidance on evaluating linkage quality in cohort studies, focusing on estimating the rate of, and understanding the reasons for, linkage error. The article also provides discussion of the information that should be shared with end-users so that linkage error can properly be accounted for in statistical analyses. Cohort studies in Denmark, Finland, Norway, and Sweden are exemplars because everyone in these Nordic countries has a single personal identity number that enables linkage to nationwide registries (Maret-Ouda et al. 2017). The next two papers are from the same

group and demonstrate the power of linked data for life course epidemiology. They use weight and height measurements between 7-13 years of age from the Copenhagen School Health Records Register of over 350,000 children who ever attended school in the central municipality of Denmark (Baker et al. 2009). The first paper by **Baker el at** involved linkage to the Danish Cancer Registry and provides evidence that a high body mass index (BMI), a short height, and excess BMI gain in childhood are associated with increased hazards of bladder cancer. The second paper by **Aarestrup et al** involved linkage to the Vital Statistics Register and the Danish National Patient Register and provides evidence that the leanest and tallest girls had the greatest risk of endometriosis. Oestrogen exposure is one possible mechanism because this hormone is crucial for the acceleration of linear pubertal growth and high levels have been implicated in the aetiology of endometriosis (Drop et al. 1998; Zondervan et al. 2018).

**Cross-cohort research**

Over the last decade there has been a dramatic increase in national and international initiatives to collate data across multiple longitudinal studies for cross-cohort research. This approach increases statistical power due to larger sample sizes, can help with understanding the extent to which associations might be causal by comparing associations in cohorts with different confounding structures (Brion et al. 2011), and allows investigation of temporal and geographical variation in longitudinal processes and relationships. The UK has a long history of investing in cohort research, and in 2012 the Cohort and Longitudinal Studies Enhancement Resources (CLOSER) initiative was established to maximise the use, value, and impact of the UK studies (O'Neill et al. 2019). Professor Rebecca Hardy was appointed as director of CLOSER in 2019. The special issue paper by **Hardy & O'Neill** is a commentary outlining the work of CLOSER on harmonising existing data across the UK studies, developing a meta-data platform containing information (in one place) on what has been collected in which studies and when, and building capacity in the necessary advanced analytical skills for cross-cohort analyses of longitudinal data. One complexity is that cross-cohort research often involves working with data that vary according to age, period, and cohort (e.g., a measurement might have been taken at 15 years of age, in 1985, on someone born in 1970). The review by **Bell** provides an accessible introduction to the distinction between these three effects and why all three cannot be included as independent variables in a single model due to linear dependency (e.g., 15 = 1985 – 1970). The rest of the paper critically reviews methods that have been proposed to resolve this issue, concluding that it simply cannot be done without making a strong assumption about at least one of the three effects. The final special issue paper by **Johnson et al** uses the CLOSER harmonised weight and height data from the 1946, 1958, 1970, and 2001 birth cohorts to

provide an illustration of longitudinal cross-cohort research. The article investigates how and why the Benn parameter (i.e., the power B in the Benn index weight (kg)/ height $(m)^B$ that minimises the correlation of the index with height) varies across age within each cohort (Benn 1971). Here, the period effect is conflated with the age effect (i.e., period = cohort + age) or can be assumed to be zero. The authors' key finding is that between 1956-2015 the Benn parameter in adolescence initially fell due to a lowering of the weight-height correlation but latterly and more drastically rose due to increasing weight variation. The implications for obesity epidemiology and monitoring of population trends are discussed.

**Concluding remarks**

A purpose of the 2019 SSHB conference was to bring together academics from a wide range of disciplines (e.g., anthropology, nutrition, psychology, statistics, geography, and archaeology) who might not necessarily call themselves human biologists but who clearly conduct cohort research relevant to the advancement of human biology. This special issue reflects the multidisciplinarity and collaborative nature of human biology.

**References**

Adair LS, Popkin BM, Akin JS, Guilkey DK, Gultiano S, Borja J, Perez L et al. 2011. Cohort profile: the Cebu longitudinal health and nutrition survey. International journal of epidemiology. 40(3):619-625.

Baker JL, Olsen LW, Andersen I, Pearson S, Hansen B, Sorensen T. 2009. Cohort profile: the Copenhagen School Health Records Register. International journal of epidemiology. 38(3):656-662.

Benn RT. 1971. Some mathematical properties of weight-for-height indices used as measures of adiposity. British journal of preventive & social medicine. 25(1):42-50.

Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L et al. 2013. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. International journal of epidemiology. 42(1):111-127.

Brion MJ, Lawlor DA, Matijasevich A, Horta B, Anselmi L, Araujo CL, Menezes AM et al. 2011. What are the causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts. International journal of epidemiology. 40(3):670-680.

Cole TJ, Donaldson MD, Ben-Shlomo Y. 2010. SITAR--a useful instrument for growth curve analysis. International journal of epidemiology. 39(6):1558-1566.

Drop SL, De Waal WJ, De Muinck Keizer-Schrama SM. 1998. Sex steroid treatment of constitutionally tall stature. Endocrine reviews. 19(5):540-558.

Druet C, Stettler N, Sharp S, Simmons RK, Cooper C, Smith GD, Ekelund U et al. 2012. Prediction of childhood obesity by infancy weight gain: an individual-level meta-analysis. Paediatric and perinatal epidemiology. 26(1):19-26.

Elliott J, Shepherd P. 2006. Cohort profile: 1970 British Birth Cohort (BCS70). International journal of epidemiology. 35(4):836-843.

Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J et al. 2013. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International journal of epidemiology. 42(1):97-110.

Gilbert R, Lafferty R, Hagger-Johnson G, Harron K, Zhang LC, Smith P, Dibben C et al. 2018. GUILD: GUidance for Information about Linking Data sets. Journal of public health. 40(1):191-198.

Green A. 2017. The African Population and Health Research Center. Lancet. 390(10106):1940.

Hansen K. 2012. Millennium Cohort Study first, second, third and fourth surveys: a guide to the datasets 6th edn. London, UK: Centre for Longitudinal Studies, University of London

Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, Goldstein H. 2017. Challenges in administrative data linkage for research. Big data & society. 4(2):2053951717745678.

Hennig BJ, Unger SA, Dondeh BL, Hassan J, Hawkesworth S, Jarjou L, Jones KS et al. 2017. Cohort Profile: The Kiang West Longitudinal Population Study (KWLPS)-a platform for integrated research and health care provision in rural Gambia. International journal of epidemiology. 46(2):e13.

Heude B, Forhan A, Slama R, Douhaud L, Bedel S, Saurel-Cubizolles MJ, Hankard R et al. 2016. Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development. International journal of epidemiology. 45(2):353-363.

Inskip HM, Godfrey KM, Robinson SM, Law CM, Barker DJ, Cooper C, Group SWSS. 2006. Cohort profile: The Southampton Women's Survey. International journal of epidemiology. 35(1):42-48.

Kuh D, Karunananthan S, Bergman H, Cooper R. 2014. A life-course approach to healthy ageing: maintaining physical capability. The Proceedings of the Nutrition Society. 73(2):237-248.

Kuh D, Pierce M, Adams J, Deanfield J, Ekelund U, Friberg P, Ghosh AK et al. 2011. Cohort profile: updating the cohort profile for the MRC National Survey of Health and Development: a new clinic-based data collection for ageing research. International journal of epidemiology. 40(1):e1-9.

Maret-Ouda J, Tao W, Wahlin K, Lagergren J. 2017. Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data. Scandinavian journal of public health. 45(17_suppl):14-19.

Monasta L, Batty GD, Cattaneo A, Lutje V, Ronfani L, Van Lenthe FJ, Brug J. 2010. Early-life determinants of overweight and obesity: a review of systematic reviews. Obesity reviews : an official journal of the International Association for the Study of Obesity. 11(10):695-708.

O'Neill D, Benzeval M, Boyd A, Calderwood L, Cooper C, Corti L, Dennison E et al. 2019. Data Resource Profile: Cohort and Longitudinal Studies Enhancement Resources (CLOSER). International journal of epidemiology. 48(3):675-676i.

Ong KK, Ahmed ML, Emmett PM, Preece MA, Dunger DB. 2000. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. Bmj. 320(7240):967-971.

Power C, Elliott J. 2006. Cohort profile: 1958 British birth cohort (National Child Development Study). International journal of epidemiology. 35(1):34-41.

Prentice P, Acerini CL, Eleftheriou A, Hughes IA, Ong KK, Dunger DB. 2016. Cohort Profile: the Cambridge Baby Growth Study (CBGS). International journal of epidemiology. 45(1):35 a-g.

Roche AF. 1992. Growth, Maturation and Body Composition: The Fels Longitudinal Study 1929-1991. Cambridge, UK: Cambridge University Press.

Rutter M, Tizard J, Whitmore K. 1970. Education, Health and behaviour. London, UK: Longman Group.

Stearns SC. 1992. The Evolution of Life Histories. New York, USA: Oxford University Press.

Tanner JM. 1981. A history of the study of human growth. Cambridge, UK: Cambridge University Press.

Wadsworth M, Kuh D, Richards M, Hardy R. 2006. Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). International journal of epidemiology. 35(1):49-54.

Wells JCK, Nesse RM, Sear R, Johnstone RA, Stearns SC. 2017. Evolutionary public health: introducing the concept. Lancet. 390(10093):500-509.

Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Vigano P. 2018. Endometriosis. Nature reviews Disease primers. 4(1):9.