A celebration of Harvey Goldstein's lifetime contributions

Harvey Goldstein and his time at the Institute of Child Health

T J Cole
UCL Great Ormond Street Institute of Child Health

Introduction

Harvey Goldstein's career as a statistician was both long and distinguished. It extended from his time as a postgraduate student at UCL in 1962 right up to his death on 9 April 2020, a period of more than 57 years. And it ended just as it had started, with him working at the Institute of Child Health in London (hereinafter the ICH). There is a pleasing symmetry to this, and the symposium dedicated to him on the occasion of his 80th birthday was—appropriately—held at ICH. In retrospect one can see how important the ICH was for his professional development.

After working as a research assistant his first substantive job was at the ICH, then part of London University (now UCL) where in 1964 he started as a lecturer in the department of growth and development, headed by Professor James (Jim) Tanner. The ICH was then, as now, two distinct entities; not only did it play an important role within London University, but it was also the academic arm of the famous NHS tertiary paediatric centre Great Ormond Street Hospital (GOSH)—the two buildings sit back-to-back and are linked both physically and professionally. Many staff in the two institutes have honorary contracts in the other place, and together ICH and GOSH represent a powerhouse of expertise in academic and clinical paediatrics and child health.

When Harvey arrived in 1964 there were several people around who were to prove highly influential to his career. The first was his boss Jim Tanner, who had been in post since 1956 and had developed a formidable international reputation as a paediatrician with a particular interest in the statistics of growth and development. He had set up in 1948 the Harpenden Growth Study, recruiting children from a children's home in Harpenden Hertfordshire, and measuring them repeatedly over time. Later he developed his "Tanner stage" system of grading secondary sexual characteristics in puberty based on these children. His recruiting Harvey was a clear statement of his intent to develop the statistics of growth.

Neville Butler was another influential paediatrician, working at that time in GOSH, who had recently set up the National Child Development Study (NCDS), a cohort of 17,000 infants born during one week in March 1958. In 1965 he revisited them at age 7, and Harvey was involved in analysing their heights. A third influence, Michael Healy, was head of statistics at the Medical Research Council unit at Northwick Park, and he was a close friend of and collaborator with Jim Tanner. Through Jim, Michael became interested in growth statistics and in turned influenced Harvey. Later Michael was awarded the Royal Statistical Society's Guy Medal in Gold for his pioneering work in agriculture and medical statistics.

So from the start, Harvey was immersed in an environment where children were being measured over time to monitor their growth and development, and the statistical challenges that they posed were to occupy him throughout his career.

His first publication in 1963 had nothing to do with children, growth or development, though it did involve age. As a research assistant at the UCL department of statistics he provided a regression analysis to a paper by Jacobs *et al* on age trends in chromosome counts by sex.(1) This was an auspicious start to his publishing career, with the paper appearing in the prestigious journal Nature.

Human growth data provide a wonderful playground for statisticians: the data are easy to collect; they come in several different forms – continuous measurements like height or weight, ordered categorical variables such as the five-category Tanner stages, or binary indicators like menarche status in girls; and they have an intriguing hierarchical structure—repeated measurements in one child are less variable than single measurements from different children. The phrase "growth and development" covers on the one hand continuous measurements that display *growth* over time, and on the other hand binary or ordered variables that reflect *development* as measured by the time to reach some developmental milestone.

To put the scale of Harvey's achievements in perspective, it is useful to sketch out the areas of growth research that he became involved in, from the standpoint of now. To do this we use data from Tanner's Harpenden Growth Study,(2, 3) 4634 height measurements in 418 Harpenden boys aged 1-20 years, seen on 1-32 occasions from 1949-1969.

Figure 1 shows the data plotted against age in three different ways: in Figure 1A they are treated as *cross-sectional* and shown as points; in Figure 1B they are also points but *conditional*, focussing on ages 9 and 13, and in Figure 1C they are *longitudinal*, with the data connected as individual growth curves. The three formats represent different ways not only of viewing the data, but also of analysing them, as shown in Figure 2.

Figure 2A summarises the data in Figure 1A as height centile curves, where each curve identifies a specified percentage of the data lying below it at each age, e.g. 50% below the median. The nine centiles, which are spaced two-thirds of a z-score apart, are estimated using

Generalized Additive Models for Location Scale and Shape (GAMLSS) (4) based on the Box-Cox Cole-Green family. This is now the favoured method for constructing growth charts.

Figure 2B plots height at age 13 conditional on height at age 9 for 70 boys seen at the two ages, with the fitted linear regression line and 95% confidence band showing how well the earlier height predicts the later height. This analysis can be extended naturally to include other covariates.

Finally Figure 2C shows the growth curves of Figure 1C plotted in grey, and in addition adjusted versions of the curves shown colour-coded, where each curve is adjusted for between-child differences in mean height and the timing and intensity of their pubertal growth spurt. The effect of the adjustment is to superimpose the curves—all 418 coloured curves lie within the narrow band—and this highlights the fact that a) adjusting for subject size and timing explains nearly all the variability in Figure 1C, and b) all the individual growth curves are fundamentally the same shape. The curve adjustments are estimated using the SITAR mixed effects model (SuperImposition by Translation And Rotation).(5)

Harvey's list of publications makes clear that he was thinking and writing about these forms of analysis more than fifty years ago. His first solo publication, in 1968, "Longitudinal studies and measurement of change",(6) was in effect a personal manifesto about how to work with longitudinal data—it covered the definition, collection, sampling, editing and data processing, setting up of hypotheses, and statistical analysis of longitudinal data. Running to 25 pages it was mostly text with just two figures and three equations, and for examples he used the NCDS at age 7 and the 1946 National Survey of Health and Development (NSHD).(7) True to the paper's title, the statistical analysis focussed on the measurement of change, where he discussed mixed models for growth curves comparing univariate and multivariate procedures. His references to fixed and random effects were to become a key part of his later multilevel models. In a later paper he expanded on the data processing requirements for longitudinal studies, including software for analysing longitudinal data,(8) which was the precursor of his MLwiN package for analysing multilevel models.

He returned to the topic of growth curves in 1971,(9) where he was uncharacteristically downbeat about the possibility of modelling height effectively during adolescence. He correctly pointed out that such a curve requires a minimum of three parameters per subject: one for final height, one for age at peak velocity and one for peak velocity, and he showed that this was impossible to achieve with a polynomial growth curve. It took over twenty years for the solution to emerge: a mixed effects model with a cubic spline fixed effects mean curve and the three subject parameters as random effects.(5, 10) Fully fifty years after his 1968 manifesto he expanded on this model, showing how to analyse longitudinal growth data including within-individual variability as a function of age.(11)

The first follow-up of the NCDS in 1965 provided Harvey with the opportunity to investigate factors affecting height at age 7 in nearly 15,000 children, 89% of the birth sample. This can be viewed as a conditional analysis, adjusting height at 7 for factors at birth. The factors he focussed on were perinatal and maternal, adjusted for age and sex: birthweight, gestational age, maternal parity, age, height, smoking habits, social class and the number of younger siblings.(12) Despite the large sample size there were no significant interactions by sex, and age could be adjusted for as a linear trend. The results were presented as a series of tables of regression coefficients and variances, in a format that was to become familiar in his later presentation of multilevel models. His Table 6 provided a comprehensive summary of the joint factor effects, with all but one of them significant, and dramatically strong associations with the biological factors of maternal height (0.27 cm per cm) and birthweight (0.21 cm per 100 g), and the more social factors of social class (1.3 cm taller for class 1-2 than 5) and parity (2.8 cm taller for parity 0 compared to parity 3+).(12)

The topic of growth standards (now more commonly known as growth references) also featured prominently in Harvey's early work. In 1972 he described an improved method for estimating growth reference centiles, by fitting to the data a regression line on age and modelling the distribution of the residuals as a function of age.(13) His example involved a linear age trend for the mean and a linear trend for the residual variance, but he emphasised that in principle both could be more complex in form. This foreshadowed by 16 years my own LMS method, where the median, coefficient of variation and skewness of the measurement distribution are estimated as cubic spline curves in age.(14)

Before estimating growth centiles one first needs to collect the data to base them on. The Cuban Growth Study of 1972 was a textbook example of how such studies should be designed—Harvey was the senior author of the study report, and one can hear his voice in much of the writing.(14) It involved the planning and collection of data on over fifty thousand children aged 0 to 20 years, and a key consideration was how many to measure at each age. Harvey recognised that measuring children at random ages between 0 and 20 years was a sub-optimal strategy, and instead he designed the study to over-sample at ages when growth was most rapid, i.e. in infancy and puberty.

By chance, in early 2020 I was researching precisely this question—how to optimally design growth centile studies. Guided by Harvey I recognised that one needed to specify not only the sample size but also the sample composition, i.e. the sample's age distribution. While writing the paper I was continually referring to Harvey's report to sample his wisdom, and I was pleased to submit the paper for publication on 10 April 2020.(15)

It happened that on that same day I heard that Harvey had died. My great sense of loss at the news was exacerbated by the realisation that I would never be able to discuss with him the paper to which he had—unknowingly—contributed so much. To try to capture the feeling I added the dedication: “The paper is dedicated to the memory of Professor Harvey Goldstein, a valued colleague whose many eminent statistical contributions included designing the 1972 Cuban Growth Study.”(15)

References

1. Jacobs PA, Doll R, Goldstein H, Brunton M, Court Brown WM. Chance of human chromosome count distributions with age - evidence for a sex difference. Nature. 1963;197(487):1080-1.
2. Tanner JM, Whitehouse RH, Marubini E, Resele LF. The adolescent growth spurt of boys and girls of the Harpenden Growth Study. Ann Hum Biol. 1976;3(2):109-26.
3. Cole TJ. Tanner's tempo of growth in adolescence: recent SITAR insights with the Harpenden Growth Study and ALSPAC. Ann Hum Biol. 2020;47(2):181-98.
4. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). Appl Stat. 2005;54(3):507-44.
5. Cole TJ, Donaldson MDC, Ben-Shlomo Y. SITAR–a useful instrument for growth curve analysis. Int J Epidemiol. 2010;39(6):1558-66.
6. Goldstein H. Longitudinal studies and the measurement of change. The Statistician. 1968;18(2):93-117.
7. Douglas JWB, Blomfield JM. The home and the school. London: George Allen and Unwin; 1958.
8. Goldstein H. Data processing for longitudinal studies. Appl Stat. 1970;19(2):145-51.
9. Goldstein H. The mathematical background to the analysis of growth curves. Proc Inter Cong Ped Vienna. 1971;XV:39-46.
10. Lindstrom MJ. Self-modelling with random shift and scale parameters and a free-knot spline shape function. Stat Med. 1995;14:2009-21.
11. Goldstein H, Leckie G, Charlton C, Tilling K, Browne WJ. Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. Stat Methods Med Res. 2018;27(11):3478-91.
12. Goldstein H. Factors influencing the height of seven year old children - results from the National Child Development Study. Hum Biol. 1971;43:92-111.
13. Goldstein H. Construction of standards for measurements subject to growth. Hum Biol. 1972;44(2):255-61.
14. Cole TJ. Fitting smoothed centile curves to reference data. J R Statist Soc A. 1988;151:385-418.
15. Cole TJ. Sample size and sample composition for constructing growth reference centiles. Stat Methods Med Res. 2021;30(2):488-507.

Figure 1. The Harpenden Growth Study (2, 3) with 4634 height measurements in 418 Harpenden boys aged 1-20 years, seen on 1-32 occasions from 1949-1969. A: data plotted cross-sectionally; B: data at ages 9 and 13; C: data plotted as longitudinal growth curves.

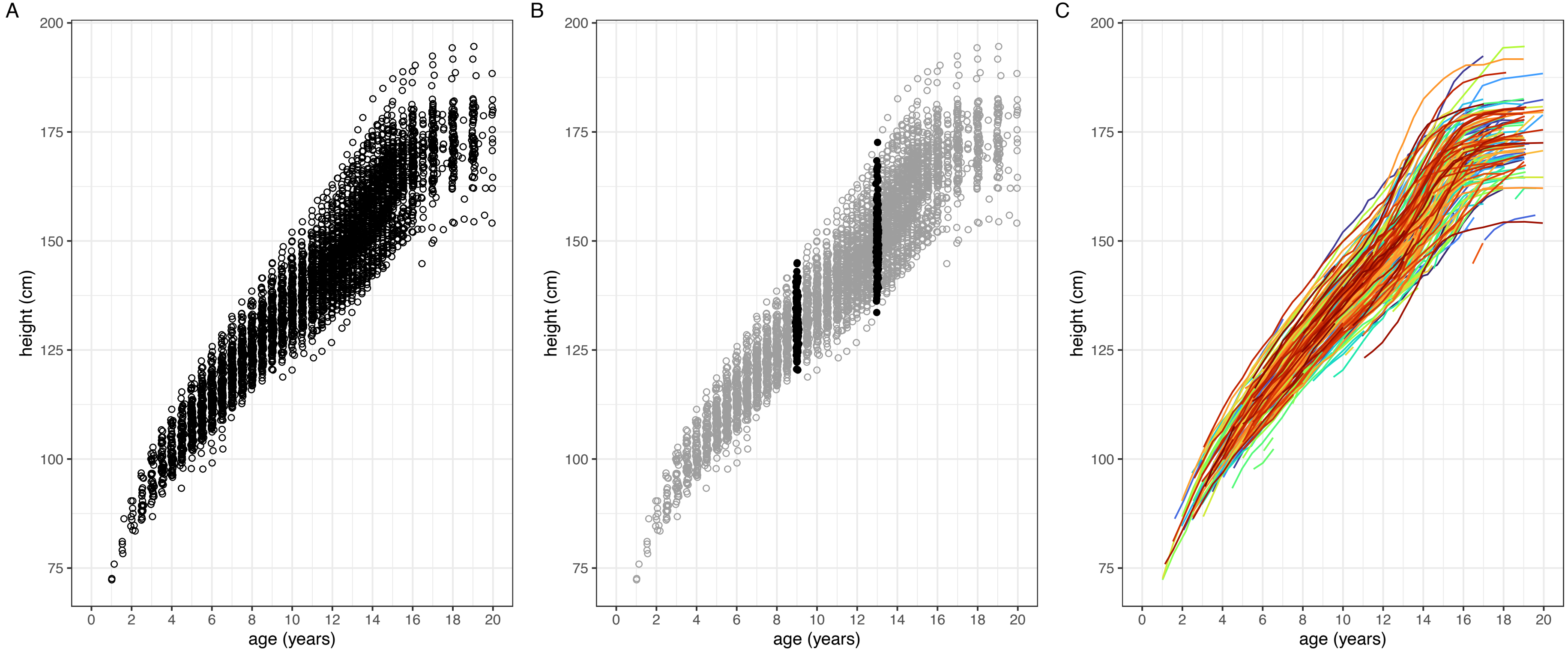

Figure 2. Analyses based on the data as presented in Figure 1. A: nine height centiles constructed using GAMLSS (4); B: conditional regression of height at 13 on height at 9; C: height growth curve analysis using SITAR (5) where the individual curves (grey) when adjusted for differences in mean size, pubertal timing and intensity are superimposed on the mean curve (in colour).