Family ties

Using sophisticated data management techniques, **Dr Joe Rodgers** and his research team have updated two important longitudinal datasets by integrating information on family/kinship relationships

DR JOE RODGERS

What are the primary aims of your project?

Our first aim is to identify the kinship links within the two National Longitudinal Survey of Youth (NLSY) datasets – the original NLSY cohort collected in 1979 (NLSY79) and NLSY-Children (NLSYC) – and to circulate them to academic researchers who use twin, sibling and cousin links in their research. Our second aim is to conduct and publish important and valuable research using these kinship links.

In what ways have your previous experiences prepared you for this project?

In the early 1990s, my research team was among the first social science researchers to recognise the value of family member relationships in NLSY datasets. Our strengths in quantitative methodology, developmental psychology, and strong programming and data management skills provided the necessary knowledge base to develop the first set of NLSY79 and NLSYC kinship links. Originally, we developed the links to facilitate our own research programme and support graduate student research, but it quickly became apparent that other researchers could gain value from access to kinship links, and so the project quickly moved into the public domain.

Could you outline the unique features of NLSY79 and NLSYC?

NLSY79 has several unique features: it is longitudinal, having been collected over a 33-year period from 1979-2012 (and continuing); it is based on a probability sample representative of the US population of households on 31 December 1978; it contains extensive within-family information; and it contains a remarkably broad set of outcome variables related to employment, marriage/reproduction/fertility, intellectual development, social and household environments, delinquency, drug, alcohol and tobacco use, and so on.

NLSYC is a biannual survey of all biological children born to NLSY79 females (over 6,000 females who have produced over 11,000 children), collected from 1986-2012 (and continuing). Extensive information about child, adolescent and young adult development is available within the NLSYC data.

For what purposes can kinship structure be of use to researchers?

Virtually all of the applications from the work by our team and others have public health implications. Previously our work has been used in the development of adolescent pregnancy control, tobacco cessation and teenage drinking programmes. We have also identified causal factors underlying adolescent delinquency and conduct disorder, with some surprising findings. Specifically, family factors are not nearly as strongly causal as previously believed, because differences between families that existed before children were born have a great deal to do with child outcomes. Such findings could not be obtained without the use of sibling and other kinship information.

Have you distinguished different types of sibling relationships?

In NLSY79 we target all kinship links that are specified in the original household rosters. There are over a dozen of those, including twins, full and half siblings, cousins, second cousins, uncles/aunts, adoptive siblings, friends, etc.

In NLSYC data, we primarily distinguish between full- and half-siblings because all of the children are the biological progeny of NLSY79 females. We also identify cousins, via the kinship links between their mothers.

Are you collaborating with other institutions in the course of your work?

Our research team has recently collaborated with several different research groups in addition to the University of Oklahoma research group that conducted the original study on links.

At the University of Chicago, a team led by Ben Lahey and including Brian D'Onofrio, Paul Rathouz, Carol Van Hulle and Irwin Waldman has used NLSYC data to study childhood and adolescent delinquency/conduct disorder, funded through two National Institutes of Health grants. This group has expanded to include activity at Emory University, the University of Wisconsin, University of Indiana and University of Chicago and has published over a dozen papers in the last five years. Several current and former graduate students from these universities are now working actively on projects using the kinship links.

We have ongoing collaborations with Dr Warren Miller at the Transnational Family Research Institute in Aptos, California; Dr Edwin van den Oord, who runs a genetics laboratory at Virginia Commonwealth University; and Hans-Peter Kohler, an economist at the University of Pennsylvania. A separate research effort by the President of the Behaviour Genetics Association, Eric Turkheimer from the University of Virginia, has resulted in a number of publications by himself and his current and past graduate students, who are actively using these data.

Kinship links

After 20 years of analysis by researchers at the **University of Oklahoma**, USA – using increasingly advanced algorithms and database management to validate, cross reference and integrate family connections – kinship links in both NLSY79 and NLSYC datasets are now complete, with a 95 per cent assignment rate

FOR MORE THAN three decades, research on two longitudinal studies has been shaping our perceptions and commonly held views on family relations. Yet, like many psychological and sociological studies, they have limitations in scope and resonance due to methodological or sampling deficiencies.

The 1979 National Longitudinal Survey of Youth (NLSY79) and subsequent National Longitudinal Survey of Youth, Children/Young Adults (NLSYC) collected information in a number of key areas related to family functioning. Both highly regarded as valuable resources for research in many fields, in 1997 they were complemented by a new study – known as NLSY97; it surveyed 9,000 young people between 12-16 years old and their parents. The young participants in NLSY97 are now followed up each year to document their transition from school to work.

Back in 1979, 12,686 young people in the US between the ages of 14 and 22 were randomly selected and surveyed on topics including their family background, job market behaviour, educational experience, health issues, income and assets. They were then surveyed annually until 1994, and have been questioned every other year thereafter. A survey of the children born to NLSY79 women commenced seven years after the original survey. The longitudinal NLSYC survey covers a wide variety of topics beyond demographic and development information, such as the subjects' attitudes to education, their dating and friendship patterns, experience of substance abuse, and the nature of their interactions with their parents. NLSYC continues to collect data every two years.

Until recently, neither survey was explicit on questions about the status of the subjects' siblings and other such relationships. Specifically, whether or not subjects were full or half siblings relied on inferring information from initial parent interviews. To Dr Joe Rodgers, then Professor at the University of Oklahoma, and now Director of the Quantitative Methods programme at Vanderbilt University in Nashville, this was a key limitation to their research utility in a number of areas. "Behaviour geneticists use levels of relatedness to separate out and analyse sources of variance due to genetic and environmental causes," he expands. "Many researchers use family background to control for differences in accounting for causal influences on developing children while many researchers in psychology, sociology and family studies use siblings as their basic unit of analysis."

TOWARDS RELIABLE KINSHIP INFORMATION

Keen to add questions that would improve the quality of the kinship information, Rodgers worked for 10 years with the Center for Human Resource Research and the Bureau of Labor Statistics to include questions about biological relatedness in the surveys. Since 2006 the surveys have included questions that explicitly distinguished between full and half siblings; and over the last three years – to make the full datasets more beneficial to their own research and that of others – Rodgers and his group have created research files, whose purpose is to remedy the gap in the data collected before 2006 and to extend the quality and quantity of the kinship relationship information in NLSY79 and NLSYC.

Rodgers and his core research team – co-Investigator David Bard and Chief Programmer Will Beasley – developed sophisticated algorithms and data management processes to identify the links between subjects. Tracing twins by matching birthdays was relatively simple, but establishing consistent sibling relationship information was more difficult. Kinship pairs were tentatively identified before algorithms were used to specify likely sibling relatedness, and then to assign a level of confidence. Potential inconsistencies, such as siblings disagreeing about the nature of their relationship or a mother's information disagreeing with her children's, required detailed analysis and detective work, including testing against traits within families such as height and resolving conflicting statements given over time: "There are many inconsistencies to resolve within the NLSY datasets, and we need objective and valid ways to decide on kinship relatedness," muses Rodgers. The team therefore applied further algorithms to resolve remaining discrepancies and inferred levels of relatedness from known variables, particularly regarding cousin and aunt-niece/aunt-nephew pairs, before datasets were finally integrated. Ultimately, over 40,000 kinship pairs were defined across the two datasets, including a number of kinship pairs that can only be specified using cross-generational information.

Rodgers and his team developed sophisticated algorithms and data management processes to identify the links between subjects

Research team: From **left** to **right**: Joe Rodgers, David Bard, Kelly Meredith, Will Beasley and Mike Hunter.

WIDE-RANGING RESEARCH POSSIBILITIES

More than 40 research papers were published using the previous versions of the group's kinship link datasets, in addition to seven Master's theses and doctoral dissertations based on their data. Amongst their uses, the data have provided methodological assistance for an age- and gender-based assessment of educational achievement using brothers and sisters, and an investigation of antisocial behaviour among siblings; as well as a substantive means of identifying environmental influences of race-based differences in infant birth weight and determining non-shared environmental influences on delinquent youth behaviour.

Rodgers' own research programme has used the kinship information extensively. Several studies identified the genetic and environmental contributions to individual differences in sexuality and fertility behaviour among adolescents and young adults. Another has studied the causal relationship between smoking during pregnancy and offspring externalising behaviours, concluding that most of the relationship is caused by selection factors that were already present before pregnancy, rather than a direct causal link between maternal smoking and childhood behavioural problems. Another line of research has studied home environmental influences on childhood cognitive development, including identification of the number of books in the home as a significant positive correlate of cognitive development.
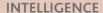
The kinship links for NLSYC were released in May 2012 and the final version of NLSY79 will be released in fall, 2013. "Except for minor tuning after each biannual survey, these links are likely the best that can possibly be produced because we use a combination of explicit and implicit kinship indicators to identify different levels of relatedness," states Rodgers. He and his team have also produced a master file that combines all kinship links within and across both datasets, containing more than 40,000 kinship pairs: "Linking the two provides cross-generation information of a type completely unique to the NLSY data," enthuses Rodgers. While the earlier success rate at linking NLSY79 was around 65 per cent and around 80 per cent for that of the NLSYC links, the success rate in the links in both datasets is now around 95 per cent; a significant leap that will add value for users and open up more opportunities for research.

The group has successfully obtained funding for a support service for researchers interested in using the kinship links of the NLSY datasets. Two trained graduate students are available to work closely with users; the user support office is currently staffed by Kelly Meredith and Mike Hunter, former and current graduate students in quantitative psychology at the University of Oklahoma. There are also examples and templates available online to support researchers in behaviour genetics, developmental psychology and related fields, and if further support is required, then Rodgers is keen to help: "Ultimately, David Bard, Will Beasley and I are also available for research consultations," he offers. The trio can deliver data structured in various ways to reflect the different requirements of, for example, developmental psychologists, economists and behaviour geneticists. They can also provide cross-generational kinship links for aunt-niece, aunt-nephew and cousin and second cousin pairs for both NLSY79 and NLSYC. Files can be emailed and are also available online.

Rodgers and his team now hope to obtain financial support for a further exercise, creating the kinship links for the NLSY97 dataset. Because it is a newer longitudinal survey that approximates the NLSY79 methodology and includes kinship indicators, Rodgers anticipates that this effort will be straightforward.

JOE RODGERS retired in 2012 from the University of Oklahoma, after 31 years on the faculty. He moved to Vanderbilt University in Nashville, Tennessee, where he is now Director of the Quantitative Methods Program in the Department of Psychology and Human Development in the Peabody College of Education. The NIH grant supporting his work on the NLSY kinship links is still administered out of the University of Oklahoma.