Published in final edited form as:
Paediatr Perinat Epidemiol. 2004 July ; 18(4): 290–301. doi:10.1111/j.1365-3016.2004.00572.x.

Transition from Birth to Ten to Birth to Twenty: the South African cohort reaches 13 years of age

Linda M. Richter[a], **Shane A. Norris**[b], and **Thea De Wet**[c]

[a]Child, Youth and Family Development, Human Sciences Research Council & School of Psychology, University of Natal, Dalbridge

[b]Mineral Metabolism Research Unit & Birth to Twenty, Department of Paediatrics and Child Health, University of the Witwatersrand, Johannesburg

[c]Department of Developmental Studies and Anthropology, Rand Afrikaans University, Johannesburg, South Africa

Summary

Birth to Ten now Birth to Twenty (BT20), is the largest and longest running longitudinal birth cohort study in Africa. In this paper, the methods, magnitude and significance of recruitment, follow-up and attrition are described. Although more than 5000 births were notified in the area in the 7-week enrolment period in early 1990, only 3275 children were established to have been born to women who were residents in the Greater Johannesburg Metropolitan area for at least the first 6 months of the child's life. Seventy per cent of these children and their families have been followed up for more than 12 years, indicating an average attrition rate of less than 3% per annum, with most attrition occurring in the first 2 years of the study. The most common reason for attrition was movement out of the study area, although detailed follow-up, and the extent of contact re-established at later points, indicate very high levels of circular migration among women and young children between urban and rural areas, as well as very high levels of residential mobility within urban areas. There has been no differential loss of vulnerable families and children. African women living in Soweto are the most consistent participants in the study. A bias, by population group membership and residential area, was introduced in the recruitment phases of the project by the difficulty of enrolling the small proportion of people in the metropolis, largely Whites, who used private delivery services in 1989–90.

Introduction

The majority of studies of child health and development are cross-sectional in nature. While cross-sectional designs are useful for testing hypothesised relationships between variables at one point in time, longitudinal studies are required to determine causality, to understand change, as well as to identify the factors and processes that determine changes in health and development. Prospective longitudinal designs ideally mitigate against the limitations of recall in retrospective studies. However, following cohorts over time is costly and difficult, requiring careful attention to cohort maintenance and attrition, together with ongoing analyses of the characteristics of cases who remain in the cohort and those who leave.1,2

Birth to Ten (BT10) now Birth to Twenty (BT20) is the largest and longest running longitudinal birth cohort study of child health and development in Africa. The children,

Correspondence: Professor Linda Richter, Human Sciences Research Council, Private Bag X07, DALBRIDGE 4014, South Africa. E-mail: lrichter@hsrc.ac.za.

Europe PMC Funders Author Manuscripts

whose life spans the transition to South Africa's democracy, turned 13 years old between April and June 2003. BT10 began in 1989 with planning meetings and pilot studies to guide the implementation of what, to many at the time, appeared to be an overly ambitious programme of research. In 2000, the final data collection wave for the first 10 years of the study took place, and a transitional phase was implemented to continue the study into a second 10 years. BT10 became BT20. BT10 broadly surveyed children's health and development. BT20, while continuing to collect a wide range of data related to young people's health and development, is targeted towards answering specific questions related to risk associated with life-style, including sexual and reproductive disorders, cardiovascular disease and diabetes.

The goals and methods of BT10 have been set out in several previous publications.3-7 BT20 has reorganised the research themes of BT10 to specifically address hypotheses around historical and current factors that place young people at risk during adolescence and in later life for sexual and reproductive disorders and diseases associated with life style.

Given the very high levels of mobility in urban areas in developing countries and the high rates of in-migration from rural areas into cities, one might ask how representative of the population a birth cohort continues to be over time and how generalisable the results are from a study, such as BT20, to the wider population. In a strict sense, the construct *generalisation* refers to the inferences that can be generated about a population from characteristics observed among a *random sample* drawn from the population, and includes the estimates of accuracy of the inferences that can be made. Such a sample is assumed, on the basis of random selection, to be *representative* of the population, in the sense that it should be like any other randomly selected group in the population and therefore like the population as a whole.

Cohort selection is an alternative form of sampling to randomisation, and generalisation does not operate on the same principles, particularly with respect to birth cohorts which are delimited by time. Nonetheless, generalisation is a form of *external validity*, one of the criteria of design adequacy which refers to the relevance of the results from a study for other situations and other samples or populations.8 High external validity is associated not only with sampling, but also with more complex features of design including realism (the likeness of the study to the real world in essential elements), reactivity (e.g. participant reactivity, placebo effects), unique cohort effects, and experimental control. Experimental and field studies differ particularly on the last feature, in that field studies, for example, are frequently more generalisable but less controlled than experiments – or, put another way, field studies frequently have greater external validity but lower internal validity.9 Therefore, the ease and logic with which the results of one particular study can be applied, either theoretically or practically, in other circumstances or among other groups of people is a more complex issue than simply the randomness, or representivity, of the sample. It is as much a conceptual as a statistical issue.9

In addition to the above considerations about the complexity of generalisation *per se*, neither the BT10 nor the BT20 projects comprises a single study. As programmes of research, they are collections of both cross-sectional and longitudinal studies of varying breadth and length, and cover a very wide variety of topics relating to children and families. For this reason, issues of external (and internal) validity need to be addressed by each specific study group with respect to any particular analysis – the issues affecting the validity of studies on lead levels in umbilical cord blood differ, for example, from the issues affecting the validity of studies of behaviour problems among preschool children. However, given the location of all BT20 studies within the larger project, a paper dealing with recruitment, enrolment and attrition is justified at this stage in the extension of the project.

Prospective longitudinal designs are the most powerful strategies for describing and understanding developmental processes and change. However, they are plagued by attrition (or non-response), either initially or over time. Attrition is due to a wide variety of factors, most important of which, in developing countries, is urbanisation, population mobility, circular migration, and the almost unattainable resources needed to track shifting groups of people across increasingly extended distances.10-12 Attrition is a significant threat to internal validity when respondents differ systematically from non-respondents,13,14 and the conditions of the analysis, for example the inclusion of particular kinds of covariates, render non-response particularly problematic.15,16 In longitudinal studies, particularly those focused on children's development, an attrition rate of 10–20% per annum is anticipated.17,18 In addition to attrition, in common with cross-sectional designs, bias can be introduced through differential enrolment of groups of people who differ systematically from one another.

In studies conducted in the West, vulnerable groups of people (those who are classified as mobile, from lower social classes, or with a high propensity for dysfunction or disorder) have been found to disproportionately fall out of longitudinal studies presumably because of higher mobility and their more tenuous institutional connectedness.19-21 This form of attrition has been shown, in some studies, to lead to significant underestimations of mortality and morbidity differentials. Although there have been very few studies of attrition in longitudinal studies conducted in developing countries, a review from the Cameroon found that drop-out was greater among higher social class groups, more intact families and healthy children. The authors concluded that the finding suggests an unwillingness of people from higher socio-economic areas to tolerate long-term studies, and to be interviewed repeatedly.22 As a result of the paucity of data and experience, there are few guidelines with respect to what can be expected about attrition rates in developing countries and how they an be minimised, and the BT20 programme can provide some insight into this issue.

In addition to the possibility of specific personal and social characteristics being systematically associated with retention and attrition in longitudinal studies, the quality of the data collected can further attenuate the effects of attrition – large amounts of missing data and mismatched cross-year identifiers, for example, can lead to the exclusion of subjects from analyses.17 Lastly, methods to reduce sample attrition are not neutral – each method retains people with particular characteristics, so that initial recruitment bias can accumulate through systematic attrition over time.19 From many respects, attrition is a complex issue, more so in a multidimensional programme of research, such as BT20.

The aims of this paper are to: (1) describe the processes of enrolment into BT10; (2) to outline the methods used to minimise attrition during follow-up; and (3) to demonstrate the difficulty of assessing both recruitment and attrition in the context of high population mobility.

Methods

Methods for three aspects of the study will be described: the pilot studies, the recruitment of cases into BT20, the strategies used to maintain the cohort and avert attrition, and available information on lost cases.

Pilot studies

Pilot studies conducted during the planning stages for BT10 in 1989 and 1990 examined several issues in an attempt to seek clarity on the anticipated numbers of children who could be enrolled into a cohort and the potential for their follow-up. These included: the monthly birth rate and seasonality of births in Soweto-Johannesburg, the scope and accuracy of

routinely collected health services and local ordinance data, the feasibility and logistics of follow-up, and estimates of sample losses, at least over the short-term.3 These studies indicated, amongst others, that an average of 2680 children were born each month within the defined study area; random patterns of births pertained amongst all groups of people, except for October, during which a significant increase in births occurred;23 approximately 16% of hospital deliveries were not recorded in routine delivery ward records because the births were directed through surgery and other specialised treatment services; routinely collected data was not suitable, without modification, for translation into research information because of differences in formats across service centres, lack of standardisation of measurements, unreliability, and incompleteness.24,25 The results of the pilot studies were used to operationalise procedures for the enrolment of the birth cohort.

The pilot studies also indicated that 20% of deliveries were immediately lost to follow-up because mothers could not be found after delivery at the addresses they had given or because the addresses they had recorded were patently false. In discussion with health service personnel and as confirmed on more intensive follow-up during the first year, it was concluded that this group of people consisted of non-resident women who planned to deliver their babies in Soweto-Johannesburg for personal or social reasons, and that they were either highly mobile within the area, or that they gave false addresses when admitted to a delivery centre. Some of the reasons why women from rural and peri-urban areas came to Johannesburg-Soweto to deliver their babies were, at the time, that they continued to believe that the place of birth of their child determined access to work as was the case under Apartheid laws to control the movement and employment of black people; they assumed that the health services in the city were superior to those in rural areas, and they came to the urban areas where their migrant husbands were working to receive financial and social support during their confinement. These women returned to their customary places of residence, usually in rural areas, soon after delivery and, for this reason, could not be considered part of a residential birth cohort in Soweto-Johannesburg.

Of the remaining deliveries, 80% were traceable 9 months after their birth. The conclusion from the pilot studies was that 64% of all births, and 80% of the resident birth cohort, were likely to be traceable for at least a year. At the time of the pilot studies, the BT10 scientific team concluded that BT10, as planned, could satisfy the basic conditions of a prospective longitudinal study – that ‘the success of a prospective birth cohort depended on an initial sample size large enough, and representative of the population, to (i) reflect a sufficient range of exposures, (ii) allow for generalisability of findings, and (iii) provide sufficient subjects who have stayed in the study throughout the period of investigation, a group known as the core sample’.4, p. 449 In other longitudinal studies, eventual cohort sizes of around 1500 have been regarded as sufficient.15 In the light of experience gained in the course of the pilot studies, a plan for the recruitment and retention of cases was developed, and this plan has been continually refined as the study has progressed.

The recruitment of cases into BT10

The recruitment of cases into BT10 is described in detail because bias in enrolment and loss to follow-up are both, to some extent, influenced by recruitment procedures.

Cohort studies examine subpopulations (cohorts) as they change over time. Typically a cohort is an age group, although it is not restricted to this.25 In the case of BT10, the cohort was defined by the timing of a singleton birth within a defined period, as well as residence for at least 6 months after the birth of the child within the designated metropolitan area. Based on information gathered in the pilot studies, a 7-week period from April 1990 was identified for the cohort enrolment time frame. Residency was restricted to the areas covered by the extant town councils governing Johannesburg, Soweto and Diepmeadow. This area

covered approximately 200 square kilometres, and included close to 3.5 million people, with about 400 000 informal housing units.

A pre-enrolment phase was undertaken to try and identify women who would deliver within the cohort frame, by interviewing women attending antenatal clinics during their 26th to 32nd week of pregnancy. The net for this pre-enrolment phase was cast deliberately wide, and included women in excess of those who eventually delivered within the cohort time frame. Of the 3000-odd women interviewed at this time, some delivered before, and some after, the designated cohort enrolment time frame. Once the cohort enrolment period began, extensive efforts were made to identify and trace records of every birth that occurred in the Soweto-Johannesburg metropolis. This included placing an interviewer in all public delivery centres for the duration of the cohort enrolment period; accessing the records of public and private delivery services, birth notification records as required by local ordinance, mortuary records; and retrospective tracing of women and their babies who later attended postnatal health services in the area. A database was constructed to contain information from these diverse sources, and to facilitate electronic selection and refinement. Cases were systematically deleted from the database in terms of the following criteria – duplicate entries, residential addresses given outside of the area, and multiple births.6

At the beginning of the study, two decisions were made which influenced enrolment and follow-up. These decisions were variously methodological, logistical and political. The first decision concerned residency, and the second concerned the use of private health services. With respect to private health services, it was clear at the planning and pilot stages of BT10 that it would be very difficult to obtain cooperation from private practitioners, clinics and hospitals for the enrolment phases of BT10. There were also political concerns about expending excessive study resources to trace a minority of people. Under Apartheid, the private sector services comprised about 15% of the population, which consisted largely of White and Indian families.24 This meant that there was an initial under-enrolment of largely middle-class White and Indian children and their families. This decision was discussed with several international experts at the time and the conclusion was reached that the value of the study lay, not in descriptions of the features of proportional samples of families from different population groups, but in the analysis of individual and family characteristics and their longitudinal determinants among the majority of children in the metropolis. This majority consisted largely of Black and Coloured families. All terminology relating to so-called ‘race’ categorisation in South Africa refers to the apartheid system of population group classification into Black, White, Coloured and Indian groups. It is retained for no other reason than its familiarity as population descriptors. Planned analyses based on health sector use and/or population group membership would have required a stratification strategy to over-sample White and Indian families. Subsequent substudies in BT10 have adopted this approach.

With respect to residence, the cohort was defined by birth and also continued residence in the defined area during the first six months of life. As children and families subsequently started to leave Soweto-Johannesburg to live in other areas of the country and even fast-growing formal and informal neighbouring areas in the metropolis, they were labelled as lost to follow-up even though some of these departing cases were, in principle, traceable. The reasons for not following up these cases were largely based on financial and logistical limitations. With additional funding, this strategy has been revisited and, in the last 3 years, families are now being traced within a radius of 100 kilometres beyond the study area.

Cohort maintenance strategy

The literature indicates that longitudinal studies that implement comprehensive procedures for contact and follow-up have higher retention rates than those that do not have a

systematic follow-up strategy.10,12,13 For this reason, a follow-up protocol using a wide range of methods was devised and implemented. As in other urban longitudinal studies, follow-up was hindered by the high mobility of people who lodge and/or rent accommodation in urban areas and who circulate between extended families in rural and urban areas; by the lack of employment in the formal sector and potential contacts through places of work; by inadequate street naming and numbering, particularly in Soweto; and by the lack of telephones in the area.

BT20 used a two-tier strategy to maintain the cohort:

Tier 1: 'Community and cohort retention' to retain participants over time

1. The creation of trust and confidence in the study and stability in its relationship with the community. The study was planned in cooperation with the health services and a Community Advisory Board that was created. In addition, field workers were recruited from the area, most of whom have been with the study from its inception and thus have personal relationships with the families in the neighbourhoods in which they collect data.
2. Detailed contact information is acquired, and stored in a computerised system. Up-to-date records of all recruits and at least two additional contactable people are recorded.
3. Intensive tracking procedures can be intrusive, giving rise to a number of ethical issues. Maintaining confidentiality is of primary importance – each BT20 family is given a unique identification number; files are locked; interviewers are trained in research ethics and participant protection; and child and caregiver interviews are conducted in private with assurances of confidentiality.
4. Compensation for transport costs is provided for participants who come into one of the two BT20 data collection sites. The value of the compensation is R50 or 4 GB pounds at the current exchange rate of R12 = 1 GB pound.
5. The length of time between phases of data collection is kept to a minimum. Contact with families is maintained between data collection phases through a number of mechanisms: tangible reminders which vary from time to time – stickers, keyrings, calendars, fridge magnets; birthday cards; reports on the study in newspapers and on the radio; regular newsletters, and subsidised social events to which random groups of children are invited – a sports day, a visit to the planetarium, etc. These procedures have been found in other longitudinal studies to be helping in reducing attrition.26
6. A limited social and health service is incorporated into the study – for example, a toll-free telephone number and a free-post address is installed at BT20 to enable families to contact the office for advice and information; referral notes to local services are given to families when health or social problems are detected; and the BT20 office responds as best it can to desperate appeals for help, including providing money towards the burial costs for deceased BT20 mothers and children. 27

Tier 2: 'Participant contact and retention' occurs through the following mechanims:

1. Children and caregivers are telephoned at home, at work, or through nominated contactable relatives, neighbours or friends.
2. Letters to children and caregivers are delivered to home.
3. Children and caregivers are visited at home.

4. Children are contacted through local schools.

Results

In this section, several characteristics of BT20 participants are described, including the total births that occurred in the area, the residential BT20 birth cohort, the core longitudinal sample, and cross-sectional study samples in BT20. In addition, some data with respect to attrition is provided, although a report on a more detailed analysis of loss to follow-up is in preparation.

Total births

As a result of the systematic collection and filtering of all available data, it was concluded that 5449 singleton births took place, during the defined time period, to women who gave a residential address in Soweto-Johannesburg. This population has been described in detail elsewhere,5 and characteristics of the total births are given in Table 1. However, not all of these births were to women who were resident in the defined area, as described below.

The residential BT20 birth cohort

The residential cohort refers to those women and children who actually resided in the Greater Johannesburg Metropolitan area at the time of the baby's delivery and also for at least the first six months of the BT20 child's life. In this way the residential cohort is distinguished from transient visitors to the metropolis who might have attended antenatal clinics in the area late in their pregnancy, and who delivered their babies in the city. The residential cohort also excludes stillbirths ($n = 112$) and child deaths in the first week of life ($n = 61$).

Recruitment of the residential cohort occurred in four waves. The first was through the interviews at antenatal clinics referred to earlier, the second was through contact made with women in clinics and hospitals at the time of the delivery of their babies, the third when children were 6 months old and the fourth around the time of the children's first birthday – the last two contacts both occurred through the well-baby services at which immunisations are given on a regular schedule during the first 18 months of the child's life. In addition, intensive drives to trace people not found before and/or missed in previous follow-ups were launched at child age of six months, three years, and when children were 7–8 years of age. At cohort age 12 years, a school survey has been undertaken of children born in the cohort range attending primary schools in the Greater Johannesburg Metropolitan Areas and beyond. The survey was designed to find cases lost to follow-up and to draw a sample of immigrants to the city during the preceding 10 years for comparison with the BT20 group.

Contacts were made with women through a number of mechanisms described earlier, mainly routine health service use, telephone calls and home visits. A total of 2176 mothers and their children (40% of the total recorded births) were not found during the year following the child's birth, nor during the intensive search processes launched at six months, three years, and 7–8 years. All the addresses registered for these women were visited at least once, and up to three times. During the school survey, which covered close to 700 primary schools in the areas, 81 cases were found who fulfilled the date and area criteria for enrolment, but who had never been recruited into the survey. Thus recruitment failure is known to have occurred in 4% of these cases. A paper describing these cases, and the reasons for failing to recruit them, has been submitted for publication (G. Del Fabbro, L. Richter and S. Norris, unpublished). The analysis of these cases indicates that they all experienced high rates of residential mobility, which is also the primary cause of intermittent loss to follow-up in the study.

Europe PMC Funders Author Manuscripts

A 40% reduction from total births to the actual residential birth cohort is very close to the 36% that was predicted on the basis of the 1989/1990 pilot studies. On the basis of these findings, the residential birth cohort was calculated to consist of the 3273 children (60% of all recorded births) who were ascertained had resided in the area for at least the first six months of the BT20 infant's life. This group is referred to as the *BT20 cohort*.

Although the longitudinal perspective afforded by the study leads to the firm conclusion that about 36% to 40% of the women who delivered in the metropolis were not residents, the question arises as to whether the subgroup of the total births classified as not belonging to the BT20 cohort differs in systematic ways from the defined BT20 residential birth cohort, particularly given the initial exclusion of people using private sector health services from the recruitment procedure. This comparison can be made using birth data from the notification forms mandated by law for all deliveries in the city and which are available for the total births. Characteristics of the total births, the BT20 cohort and the non-cohort births, are shown in Table 1.

Statistically significant differences between the two groups (BT20 cohort and non-cohort children) were found in terms of two proxies of social class (population group membership and residential area), as well as one possible indicator of infant vulnerability (gestational age). More African women living in Soweto and Diepmeadow who used public services were included in the cohort, while more White women residing in the suburban areas that used private services were non-cohort cases. The cohort was thus not unduly influenced by the loss of large numbers of non-resident women who used the city's clinic delivery systems and then left the area, as these occurred in proportion to residents. But the cohort was differentially affected by the loss of White middle-class women through the exclusion of the private delivery services from the recruitment strategy. Differences between groups do not automatically lead to bias,24 but studies within the BT20 programme using social class and or race as explanatory variables have to evaluate carefully the implications of the cohort composition for any particular analysis undertaken. The statistical difference in gestational age between cohort and non-cohort cases falls away when cases with unknown values are removed ($\chi^2 = 1.34$; d.f. = 1; NS).

The BT20 programme is currently in contact with about 2300 children and their families; that is, cases that have been seen up to 12 years of age. This comprises 70% of the BT20 residential cohort, which means that, over more than a decade, there has been an attrition rate of less than 3% per annum, a figure substantially lower than normally occurs in child longitudinal studies.

The core longitudinal sample

The core longitudinal sample consists of those children and their families who have been interviewed and investigated at each major point during the first 10 years of the study; that is, antenatally, at delivery, in the 1st year of life, in the 2nd–3rd year of life, in the 4th–5th year, in the 7th-8th year, and between 9 and 10 years of age (7 time points). This core longitudinal sample consists of 1495 cases, or 46% of the BT20 cohort. This is an exceptional group of children in Africa, and even in the world, who have been followed up regularly for 10 years and examined intensively on a wide range of social, environmental, psychological, educational, biological and health parameters. The core longitudinal sample is being maintained into the second decade, that is, from 11 to 20 years of age. Characteristics of the core longitudinal sample from the first 10 years in comparison with the non-core sample of the BT20 cohort are shown in Table 2.

The core longitudinal sample closely resembles the BT20 cohort on all parameters except residential area. African mothers living in Soweto were significantly more likely than other

groups to be part of the core longitudinal sample. Conversely, White suburban women were much less likely to have participated in all BT20 data waves. The latter trend is due to both an under-enrolment at the start of the study and later differential catch-up enrolment as the study became established.

Overlapping cross-sectional study samples

At each BT20 data collection point, and at some stages in between, related and independent but overlapping cross-sectional study samples have been assembled, some of which have been larger, and some of which have been smaller than the BT20 cohort. For example, the total births were used to do an analysis of correlates of caesarean section births in Soweto-Johannesburg.28 In addition, more than 2800 women were interviewed antenatally in anticipation of their possible inclusion in the BT20 cohort. In fact, only 1976 met the time and residential criteria for inclusion in the BT20 cohort. A study of very-low-birthweight children coincided with the BT20 cohort, but the study included also very-low-birthweight children who did not meet the exact time and residential criteria for inclusion in the BT20 cohort.

In addition to cross-sectional study samples that are parallel and/or overlapping with the BT20 cohort, the BT20 programme also includes cohorts with differently defined inclusion criteria. For example, a longitudinal study of lifestyle disease precursors is defined by intake at age 5 years of BT20 cohort children, when the first blood samples were taken for cholesterol analyses.29

Attrition

Attrition from the BT20 cohort has been tracked from enrolment to date. Attrition occurs for a number of reasons: child and/or maternal death; child abandonment or adoption; migration of the mother, family, and/or child out of the study area; transfer of children to the care of relatives in a rural area; study fatigue and so on. The reasons for attrition were tracked by enquiring after participants through neighbours and through contact addresses provided, and by requesting participants to phone or send in change-of-address forms.

The tempo of attrition over the first 8 years of the study and the reasons for loss are shown in Table 3. As indicated before, during the first 10 years of the study, intensive tracing took place at 6 months, three years, and 7–8 years. The attrition data presented here do not include new information collected during the school survey. In the very early stages of the study, maternal and child deaths accounted for a sizable proportion of cases not found. However, the most common reason for not finding a family at any data point was that either or both the mother and child had moved out of the area or the address given was problematic in some way. Successfully making contact also became a significant problem as time passed and families moved or broke up, mothers returned to work, and participants were less inclined to participate in the study as a result of research fatigue. However, loss to follow-up is also a function of resources. When the children were aged 9–10 years, the study was very short of resources, and failed contact was largely due to the inability of the small number of staff retained to seek and find families.

What the attrition data most vividly demonstrates is a longitudinal perspective on attrition. Between a third and a half of all cases logged at any particular time as ‘not found’, that is, as cases of attrition, were found in the metropolis and interviewed again at a later point. This demonstrates the high levels of family and child mobility in the area, and the extent of circular migration between urban and rural areas. It also demonstrates the difficulty of ascertaining attrition at any one time in a longitudinal study of several years' duration.

Discussion

It was the intention of this paper to illustrate the complexity of recruitment, participation and attrition in a study of 10 or more years' duration. The complexity derives from the social and economic context of the cohort which, together with the longitudinal dimension, renders the parameters of enrolment, participation and attrition difficult to pin down. The pilot studies undertaken prior to the launch of the BT20 project predicted with a fair degree of accuracy the mobile nature of the metropolitan child-bearing population, with women moving in and out of the metropolis shortly before and after the birth of a child, for a variety of personal, social and economic reasons. Only with hindsight, have we been able to ascertain with any surety the group of resident families who make up the BT20 cohort. Even so, the recently completed survey of several hundred primary schools in the area found 81 cases recorded in the total births, but who had never been found during the preceding decade.

After 10 years, the project has retained 70% of the cohort, with a rate of attrition very much lower than could have been anticipated from published data. In common with the Cameroon and other studies, bias in enrolment and attrition has, firstly, been due to low levels of participation by higher socio-economic groups, a problem that besets census and other national survey efforts in South Africa. Secondly, loss to follow-up of enrolled cases is primarily due to high mobility in informal urban areas.11,12,22

A unique core group of children and their families, consisting of close to half of the cohort has been retained with complete longitudinal data across the first 10 years of the project. This is a considerable achievement in view of the very rapid urbanisation, as well as other forms of political and social change, that have taken place in South Africa over the time period of the study reported here.

Contrary to expectations generated from Western studies, there has been no discernible loss of identifiably vulnerable families and children. On the contrary, and consistent with the one other comparable African study, better-off families have shown a greater reluctance to participate in repeat interviews and assessments over time. A bias was introduced into the project in the recruitment phases by the exclusion of the small group of people in the metropole, mainly White, who used private health facilities in 1989–90. The costs, in time and money, to enrol this group would have far exceeded the budget of the study and would have detracted resources from the commitment of the project to enrol the majority of the population of the city into the project. It has also been more difficult to trace White families. Many White families in middle-class areas live in high-security enclosures and tend to deny access to strangers. In addition, large numbers of White families are isolated in their neighbourhoods. We have frequently found that no-one in the vicinity of an address given by a White family knew anything about them or their whereabouts since they left the area. A kind of social vacuum was created by their departure. Black families are, on the whole, more socially connected to their neighbours, who often know where the departed BT20 family has gone and how to contact them.

The differential enrolment, retention and attrition of White and Black families, in particular, mean that, a decade down the line, simple population group comparisons cannot be validly undertaken because of the unknown selection factors operating in the already small group of White participants in the project. However, with the perspective across South Africa's transition to democracy, it is uncertain whether, at this stage, simple population group comparisons would yield as interesting information as comparisons based on other stratification criteria. For example, using BT20 data, Barbarin and Richter30 have examined the relationship between socio-economic status, neighbourhood levels of danger and children's psychological functioning.

Europe PMC Funders Author Manuscripts

As indicated at the start of the paper, the issues of external validity, including generalisation, are complicated. In a longitudinal study such as BT20, changing population trends also affect the representivity of a cohort. Movement in and out of the metropole during the first 10 years of the study has resulted in marked population changes. Close to 50% of children currently attending schools in the Greater Johannesburg Metropolitan area were not born in the city. Given this changing scenario, which population does the BT20 cohort 'represent'? Clearly, BT20 represents the children born in the metropolis in 1990, who have continued to live in the city and surrounding environs, and not necessarily the population currently resident in the city. For these reasons, response rates in longitudinal studies can be presented in a number of ways – as a proportion of those cases initially recruited into the project, as a proportion of those currently eligible for recruitment, as a proportion of those who responded at the last phase, and so on.17,31,32 Each approach is justified by the specific purposes of the analysis.

BT20 is a study of children born in 1990 in Soweto-Johannesburg, and it has tracked the physical and psychological development, as well as the material and social circumstances, of a very large number of children for more than 10 years. The study is continuing to 20 years, by which time a considerable proportion of the child cohort will be parents themselves. The BT20 cohort embodies the changes and transitions of a decade, the first decade of South Africa's democracy. For that reason alone, the cohort is of enormous social significance.

In order to systematically compare the BT20 cohort with the current population of the Greater Johannesburg Metropolitan area, the BT20 School Survey began in 2002 to survey all children attending schools in the metropolis. Close to 700 schools were visited, and a short demographic questionnaire sent home with more than 8000 children, for completion. By this means we hope to define the population characteristics of 12-year-olds currently residing in Johannesburg-Soweto in comparison with the BT20 cohort.

At age 11 years and beyond, data are being collected for the next phase of the study (10–20 years) which is focusing on growth and pubertal development in relation to risk factors that dispose young people to sexually transmitted infections and HIV/AIDS, and diseases associated with life style, related to diet, activity levels and substance use.

Acknowledgments

BT20 is funded by the Wellcome Trust (UK), the Human Science Research Council of South Africa, the Medical Research Council of South Africa, the University of the Witwatersrand, and the Anglo-American Chairman's Fund.

References

1. Bijleveld, C.; van der Kamp, L. Longitudinal Data Analysis: Designs, Models and Methods. Sage Publications; London: 1998.
2. Little, T.; Kai, S.; Baumert, J., editors. Modelling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches and Specific Examples. Lawrence Erlbaum; Mahwah, NJ: 2000.
3. Yach D, Padayachee N, Cameron N, Wagstaff L, Richter L. 'Birth to Ten' – a study of children of the 1990s living in the Johannesburg-Soweto area. South African Medical Journal. 1990; 77:325–326. [PubMed: 2321102]
4. Fonn S, M De Beer, Kgamphe S, McIntyre J, Cameron N, Padayachee N, et al. Birth to Ten' – pilot studies to test the feasibility of a birth cohort study investigating the effects of urbanization in South Africa. South African Medical Journal. 1991; 79:449–454. [PubMed: 2020885]
5. Yach D, Cameron N, Padayachee N, Wagstaff L, Richter L, Fonn S. Birth to Ten: child health in South Africa in the 1990s. Rationale and methods of a birth cohort study. Paediatric and Perinatal Epidemiology. 1991; 5:211–233. [PubMed: 2052483]

6. Anderson A, Richter L. 'Birth to Ten': error detection and correction of a longitudinal address database. South African Journal of Epidemiology and Infection. 1994; 9:16–19.
7. Richter L, Yach D, Cameron N, Griesel R, De Wet T. Enrolment into Birth to Ten (BTT): population and sample characteristics. Paediatric and Perinatal Epidemiology. 1995; 9:109–120. [PubMed: 7724410]
8. Ray, W. Methods Toward a Science of Behavior and Experience. Brooks/Cole Publishing Co; Pacific Grove, CA: 1997.
9. Neumann, W. Social Research Methods: Qualitative and Quantitative Approaches. Allyn and Bacon; Needham Heights, MA: 1997.
10. Sullivan C, Rumptz M, Campbell R, Eby K, Davidson W. Retaining participants in longitudinal community research: a comprehensive protocol. Journal of Applied Behavioral Science. 1996; 32:262–276.
11. Barros F, Victora C, Vaughan J. The Pelotas (Brazil) birth cohort study 1982–87: strategies for following up 6,000 children in a developing country. Paediatric and Perinatal Epidemiology. 1990; 4:205–220. [PubMed: 2362877]
12. Haaga J, DaVanzo J, Peterson C, Nai Peng T. Twelve year follow up of respondents and their adult children in a panel survey in peninsular Malaysia. Asia-Pacific Population Journal. 1994; 9:61–72. [PubMed: 12288227]
13. Brown-Peterside P, Rivera E, Lucy D, Slaughter I, Ren L, Chiasson M, et al. Retaining hard-to-reach women in HIV prevention and vaccine trials: Project ACHIEVE. American Journal of Public Health. 2001; 91:1377–1379. [PubMed: 11527761]
14. Robinson S, Marsland L. Approaches to the problem of respondent attrition in a longitudinal panel study of nurses' careers. Journal of Advanced Nursing. 1994; 20:729–741. [PubMed: 7822610]
15. Foster E, Bickman L. An evaluator's guide to detecting attrition problems. Evaluation Review. 1996; 20:695–723.
16. Plewis, I. Analysing Change: Measurement and Explanation Using Longitudinal Data. John Wiley and Sons; Chichester: 1995.
17. St. Pierre R. Planning longitudinal field studies: considerations in determining sample size. Evaluation Review. 1980; 4:405–415.
18. Aylward G, Hatcher R, Stripp B, Gustafson N, Leavitt L. Who goes and who stays: subject loss in a multicenter, longitudinal follow-up study. Developmental and Behavioral Pediatrics. 1985; 6:3–8.
19. Dodds S, Furlong A, Croxford L. Quality and quantity: tackling non-contact attrition in a longitudinal survey. Sociology. 1989; 23:275–284.
20. Janus M, Goldberg S. Factors influencing family participation in a longitudinal study: comparison of pediatric and healthy samples. Journal of Pediatric Psychology. 1997; 22:245–262. [PubMed: 9114646]
21. Simmons C, Ritchie S, Mullet M, Mingarelle C. Subject loss and its implications for a high risk population. Infant Behavior and Development. 1989; 12:139–146.
22. Kuate Defo B. Mortality and attrition processes in longitudinal studies in Africa: an appraisal of the Iford surveys. Population Studies. 1992; 46:327–348.
23. Sparaco, A. An Investigation of the Birth Patterns of the Four Population Groups Resident in the Johannesburg Area: July 1979–June 1989. University of the Witwatersrand; Johannesburg: 1990.
24. Ellison G, Richter T, de Wet T, Harris H, Griesel R, McIntyre J. Improving the accuracy of birth notification data: lessons from the Birth to Ten study. South African Journal of Epidemiology and Infection. 1997; 12:91–96.
25. Ellison G, Richter L, de Wet T, Harris H, Griesel R, McIntyre J. The reliability of hand-written and computerised records of birth data collected at Baragwanath Hospital in Soweto. Curationis. 1997; 20:36–40. [PubMed: 9287552]
26. Boys A, Marsden J, Stillwell G, Hatchings K, Griffiths P, Farrell M. Minimising respondent attrition in longitudinal research: practical implications from a cohort of adolescent drinking. Journal of Adolescence. 2003; 26:363–373. [PubMed: 12770532]

Europe PMC Funders Author Manuscripts

27. Hayman R, Taylor B, Pearl N, Galland B, Sayers R. Participation in research: informed consent, motivation and influence. Journal of Paediatrics and Child Health. 2001; 37:51–55. [PubMed: 11168870]

28. Matshidze K, Richter L, Ellison G, Levin J, McIntyre J. Caesarean section rates in South Africa: evidence of bias among different 'population groups'. Ethnicity and Health. 1998; 3:71–79. [PubMed: 9673465]

29. Steyn K, De Wet T, Richter L, Cameron N, Levitt N, Morell C. The cardiovascular disease risk factors in five-year-old urban children in South Africa: the Birth to Ten study. South African Medical Journal. 2000; 90:719–726. [PubMed: 10985136]

30. Barbarin O, Richter L. Adversity and psychological functioning of children growing up in black townships in South Africa: effects of community danger and socioeconomic status. American Journal of Orthopsychiatry. 1999; 11:16–25.

31. Babbie, E. The Practice of Social Research. 4th Edition. Wadsworth Publishing; Belmont, CA: 1986.

32. McBroom W. Sample attrition in panel studies: a research note. International Review of Modern Sociology. 1988; 18:231–246.

Table 1

Comparison of the BT20 residential cohort with cases from the total births who were excluded from the cohort (including perinatal deaths)

Characteristics	Total births (*N* = 5449) *n* (%)	BT20 cohort (*N* = 3273) *n* (%)	Non-cohort (*N* = 2176) *n* (%)
Population group			
African	4029 (74)	2568 (78)	1461 (67)
White	677 (12)	207 (6)	470 (22)
Coloured	546 (10)	383 (12)	163 (8)
Indian	191 (4)	115 (4)	76 (3)
Not known	6 (0.1)	0	6 (0.3)
χ^2 = 10.9.3; d.f. = 3; $P < 0.01$			
Residential area			
Soweto/Diepm	3384 (62)	2429 (74)	955 (44)
Suburban Jhb	1132 (21)	343 (11)	789 (36)
Former Indian/Coloured areas	620 (11)	432 (13)	188 (9)
Inner city	313 (6)	69 (2)	244 (11)
Not known	0	0	0
χ^2 = 27.8; d.f. = 3, $P < 0.01$			
Maternal age (years)			
<17	155 (3)	92 (3)	63 (3)
17–19	598 (11)	392 (12)	206 (9)
20–38	4526 (83)	2692 (82)	1834 (84)
39+	168 (3)	95 (3)	73 (3)
Not known	2	2	0
χ^2 = 0.44; d.f. = 3; NS			
Gravidity			
1	1742 (32)	1094 (33)	648 (30)
2–4	3108 (57)	1875 (57)	1233 (57)
≥5	479 (9)	304 (9)	175 (8)
Not known	121 (2)	0	121 (5)
χ^2 = 0.11; d.f. = 3; NS			
Gestational age (weeks)			
<37	717 (13)	388 (12)	329 (15)
37–41	4305 (79)	2773 (85)	1532 (70)
42+	24 (0.4)	11 (0.3)	13 (0.5)
Not known	403 (7)	101 (3)	302 (14)
χ^2 = 8.9; d.f. = 3, $P < 0.05$			
Birth weight (g)			
<1500	104 (2)	30 (1)	74 (3)
1500–2499	563 (10)	322 (10)	241 (11)
2500–3999	4614 (85)	2827 (86)	1787 (82)

Europe PMC Funders Author Manuscripts

Characteristics	Total births (*N* = 5449) *n* (%)	BT20 cohort (*N* = 3273) *n* (%)	Non-cohort (*N* = 2176) *n* (%)
4000+	149 (3)	89 (3)	60 (3)
Not known	19 (0.3)	5 (0.2)	14 (0.6)
$\chi^2 = 1.13$; d.f. = 3; NS			

Table 2

Characteristics of the core longitudinal sample in comparison with the BT20 cohort

Characteristics	BT20 Non-core cohort (*N* = 1778) *n* (%)	Core longitudinal sample (*N* = 1495) *n* (%)
Population group		
African	1300 (73)	1268 (85)
White	169 (9)	38 (3)
Coloured	226 (13)	157 (10)
Indian	83 (5)	32 (2)
$\chi^2 = 5.58$; d.f. = 3; NS		
Residential area		
Soweto/Diepmeadow	1158 (65)	1271 (85)
Suburban Johannesburg	298 (17)	45 (3)
Former Indian/Coloured areas	254 (14)	178 (12)
Inner city	68 (4)	1 (0.1)
$\chi^2 = 16.33$; d.f. = 3; $P < 0.01$		
Maternal age (years)		
<17	32 (2)	60 (4)
17–19	183 (10)	209 (14)
20–38	1511 (85)	1183 (79)
39+	52 (3)	43 (3)
$\chi^2 = 1.55$; d.f. = 3; NS		
Gravidity		
1	511 (29)	584 (39)
2–4	1088 (61)	786 (53)
≥5	179 (10)	125 (8)
$\chi^2 = 2.25$; d.f. = 2; NS		
Gestational age (weeks)		
<37	199 (11)	189 (13)
37–41	1508 (85)	1255 (84)
42+	6 (0.3)	5 (0.3%)
Not known	65 (3.6)	46 (3.0)
$\chi^2 = 0.22$; d.f. = 3; NS		
Birthweight (g)		
<1500	14 (0.8)	16 (1)
1500–2499	193 (11)	129 (9)
2500–3999	1516 (85)	1311 (88)
4000+	52 (3)	37 (2)
Not known	3 (0.2)	2 (0.1)
$\chi^2 = 0.47$; d.f. = 3; NS		

Europe PMC Funders Author Manuscripts

Table 3

Attrition across the first 10 years of the BT20 programme

	Follow-up time				
Attrition reason	**0–1 years**	**2–3 years**	**4–5 years**	**7–8 years**	**9–10 years**
Mother or child died	148	9	10	6	0
Mother and/or child moved out of the study area	104	359	326	246	2
Mother and/or child moved within the study area	42	123	264	193	52
Problems with given address	53	132	475	66	55
Reported migration between urban and rural area	0	3	19	6	0
Unavailable; no response to contact attempts	22	19	208	148	0
Not contacted – for a variety of reasons, including resources	0	505	35	568	1810
% Attrition due to mobility	54%	54%	81%	41%	6%
Participants seen during the data collection cycle	2904	2124	1937	2041	1355
Total	3273	3273	3273	3273	3273