



Published in final edited form as:

Alcohol Clin Exp Res. 2004 February ; 28(2): 330–340.

Collecting Longitudinal Data Through Childhood, Adolescence, and Young Adulthood: Methodological Challenges

Vivian B. Faden, Nancy L. Day, Michael Windle, Rebecca Windle, Joel W. Grube, Brooke S.G. Molina, William E. Pelham Jr., Elizabeth M. Gnagy, Tracey K. Wilson, Kristina M. Jackson, and Kenneth J. Sher

National Institute on Alcohol Abuse and Alcoholism (VF), Bethesda, Maryland; Western Psychiatric Institute and Clinic, Pittsburgh, Pennsylvania (NLD); University of Alabama at Birmingham (MW, RW), Birmingham, Alabama; Prevention Research Center (JWG), Berkeley, California; Departments of Psychiatry and Psychology (BSGM), University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; Departments of Psychology and Pediatrics (WEP, EMG), State University of New York at Buffalo, Buffalo, New York; ADD Program (TKW), University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; and Missouri Alcoholism Research Center (KMJ, KJS), University of Missouri-Columbia and Department of Psychology, University of Missouri-Columbia, Columbia, Missouri

Abstract

This article presents the proceedings of a workshop at the 2003 Research Society on Alcoholism meeting in Fort Lauderdale, Florida. The organizers and chairs were Vivian Faden and Nancy Day. The presentations were (1) Lessons Learned From the Lives Across Time Longitudinal Study, by Michael Windle and Rebecca Windle; (2) Methodological Issues in Longitudinal Surveys With Children and Adolescents, by Joel Grube; (3) The Pittsburgh ADHD Longitudinal Study: Methodological and Conceptual Challenges, by Brooke Molina, William Pelham, Elizabeth Gnagy, and Tracey Wilson; and (4) Lessons learned in Conducting Longitudinal Research on Alcohol Involvement: If Only I Had Known Before Hand! by Kristina Jackson and Kenneth Sher.

Longitudinal studies have been extremely important in helping us to understand the risk and protective factors for and the development of alcohol-related problems. Such studies will continue to play a critical role as we seek to further enhance our understanding of alcohol-related problems and explore the roles of genes and environmental factors and their interaction in a developmental framework that encompasses an entire lifetime. All longitudinal work presents significant methodological challenges, but working across childhood, adolescence, and young adulthood, a period of life characterized by multiple dramatic developmental transitions, presents particular challenges.

The researchers who discuss their work in the articles that follow generously offered to share what they have learned from many years of longitudinal work so that ongoing and future studies can benefit from their knowledge. They discuss their successes and failures as well as highlight some of the difficult conceptual issues with which they have struggled in designing, executing, and analyzing their studies. The authors address the unique challenges of studying people from childhood to young adulthood, a period of rapid physical, emotional, intellectual, and social development. They also reflect on the future of longitudinal work in a changing research environment and, on the basis of their experience and expertise, consider how future studies

may need to adapt what is done and how it is done to ensure the continued success of this vital kind of research. The specific populations in the studies discussed span the age spectrum from childhood to young adulthood and represent general, clinical, and high-risk populations.

Michael Windle and Rebecca Windle discuss their experiences collecting prospective longitudinal data in a two-generational family study of alcohol use and problem behaviors. The study, which included >1000 adolescents ascertained when they were either sophomores or juniors in high school, is now in its sixth wave of data collection. The authors highlight challenges that they faced in conducting this research, such as tracking and retaining participants across time in an increasingly electronically sophisticated environment. These and other obstacles are discussed with regard to historical changes that have both impeded and enhanced efforts to retain subjects.

Joel Grube draws on his experience with four National Institute on Alcohol Abuse and Alcoholism/National Institute of Child Health and Human Development–funded longitudinal studies that used differing methods for sample recruitment, data collection, and linking data over time. He highlights problems and advantages associated with (1) sampling using random-digit dialing procedures, list-assisted sampling, and school-based sampling; (2) the use of computer-assisted self-administered interviews (CASI), computer-assisted telephone interviews (CATI), and self-administered questionnaires; and (3) the use of self-generated codes for anonymously linking data over successive survey waves.

Brooke Molina, William Pelham, Elizabeth Gnagy, and Tracy Wilson highlight experiences from a prospective study of the association between attention-deficit/hyperactivity disorder (ADHD) and alcohol abuse. This study, which is following >350 children ascertained in their elementary school–aged years, is currently in its third wave of data collection. The authors discuss a number of complex issues faced, including the wide age range of the participants and associated measurement challenges as participants mature, and describe specific techniques used to locate, recruit, and retain a clinical population characterized by disorganization and psychiatric comorbidity.

Kristina Jackson and Kenneth Sher discuss their experiences studying college students from their freshman through their senior years and beyond. The study involved six waves of data collection and recently has begun additional longitudinal data collection in a new cohort of college students recruited as freshmen. The study is addressing a wide range of alcohol-related issues in this high-risk population. Drawing on their extensive experience in the college and young adult population, the authors emphasize issues related to study design and logistics.

LESSONS LEARNED FROM THE LIVES ACROSS TIME LONGITUDINAL STUDY

Michael Windle and Rebecca Windle

This research was supported by Grant R37-AA07861 awarded to Michael Windle from the National Institute on Alcohol Abuse and Alcoholism.

The importance of long-term longitudinal studies has become increasingly recognized as researchers seek to understand causative factors and time-course issues related to the onset, escalation, maintenance, termination, and relapse of alcohol and other substance abuse disorders. In this contribution, we report on some of the challenging issues that we have confronted with our ongoing longitudinal study of adolescents and their parents.

Overview of Study

Lives Across Time: A Longitudinal Study of Adolescent and Adult Development (LAT), currently in its 16th year of funding, is an ongoing prospective study of adolescents and their parents. The purpose of LAT is to investigate developmental processes and the roles of risk and protective factors that contribute to alcohol use, alcohol disorders, and other health outcomes from adolescence through adulthood. The study was initiated in Western New York in 1988, when adolescents were aged 15.5 years, and we are currently in the sixth wave of data collection, when these young adults are aged 30.5 years. The initial sample size for the study was 1218 adolescents, and at wave 5, data were collected from one or more family members from 941 families. The study may be described in three phases.

Phase I focused on middle adolescence (teens aged 15–18 years) and was titled the Middle Adolescent Vulnerability Study (MAVS). Its purpose was to identify vulnerability factors that were predictive of health outcomes among middle adolescents. Vulnerability factors included family history of alcoholism, childhood behavior problems (e.g., conduct-disordered behaviors), temperamental attributes (e.g., mood quality, activity level), quality of peer relationships (e.g., support from and conflict with friends), quality of adolescent family relationships (e.g., family support), stressful events, and functioning of the adolescent's primary caregiver (e.g., level of caregivers' depressive symptoms, alcohol and substance use). Adolescent health outcomes included alcohol and illicit drug use, alcohol-related problems, delinquent behaviors, depressive symptoms, and suicidal ideation and attempts.

The MAVS used a research design that involved four waves of data collection spaced 6 months apart. An important conceptual consideration that guided the MAVS design was that middle adolescence is a phase in the lifespan of substantial change (e.g., pubertal development, physical maturation, personal identity issues), as well as when alcohol, tobacco, and illicit drugs may escalate to high levels of use and substance-related problems and/or disorders. The intensive data collection of multiple variables across a short time period facilitated efforts to capture changes in emotional and behavioral functioning occurring in middle adolescence. Data were collected from the adolescents within their high school settings using paper-and-pencil methods, and primary caregivers completed mail surveys. A selective listing of publications with the MAVS data are provided at the end of this article.

In phase II, the title of the study was changed from MAVS to Lives Across Time: A Longitudinal Study of Adolescent and Adult Development (LAT). During this phase, data were collected from young adults when they were 23 to 25 years of age, and the focus was on the transition from adolescence to young adulthood. This period in the lifespan is a time when young adults are beginning to establish and define adult life roles (e.g., as a spouse, parent, employee). During phase II, structured interviews were conducted to assess psychiatric and substance abuse disorders, and questionnaires were used to assess significant developmental tasks of young adulthood related to the selection of a partner, the quality of intimate relationships, and employment experiences. Also, the study was expanded from including only the target child and one parent (usually the mother) to including fathers. Phase III is a 5-year follow-up of phase II and is ongoing. It includes the assessment of age-relevant issues for young adults (e.g., adjustments to parenting, career choices) and their parents (e.g., retirement, partner loss, barriers to exercise), the assessment of psychiatric and substance abuse disorders, and a more comprehensive assessment of family history of disorders.

Major Barriers to Subject Participation and Tracking

Perhaps one of the most difficult issues in the conduct of long-term longitudinal research is the retention of subjects. The following provides a list of the more prominent barriers that we encountered in maintaining the sample for our study.

- Technological advances related to telephones, such as caller ID and voice mail, have made it increasingly difficult to have direct personal contact with subjects to facilitate recruitment and retention. In addition, cell phones and the rapid changes of subjects switching cell phone companies and telephone numbers increased the difficulty of contacting subjects.
- The age period (18–24) for the wave 5 follow-up, from high school to young adulthood, is associated with a number of age-normative transitional events, such as going to college or entering the military, which can require intensive tracking.
- This young adult age group is also highly mobile with regard to changing residence or occupying more than one residence (e.g., dorm room and family-of-origin home).
- This young adult age group also changes jobs frequently; hence, work settings may not provide a stable referent for tracking subjects.
- Many women marry one or more times, often with accompanying name changes that make them difficult to track.
- We found that members of the family of origin were sometimes unwilling or unable to provide us with contact information for the young adults.
- Some of the young adults already occupied multiple social roles (e.g., spouse, parent, employee), and these demands interfered with willingness to participate.
- For some young adults with high levels of discretionary income, the monetary incentives that we offered were not powerful.

Subject Retention

Given the barriers previously enumerated, the retention of the sample across this 16-year period has required the use of multiple, systematic strategies, as well as adapting new strategies in response to both age-normative secular changes (e.g., greater mobility by young adults, later age of marriage) and technological changes (e.g., caller ID). We had very few difficulties in retaining subjects during the high school years, when the majority of students could be located in their high school settings. Our retention rates exceeded 90% during the first four waves of the study. However, retention from wave 4 (in high school) to wave 5 (5–6 years later) presented some challenges. The following provides a list of the major procedures that we used to track our study participants who had moved since wave 4.

- We contacted family members.
- We used local directory assistance and telephone books.
- We contacted the post office and used their “request to forward” service.
- We conducted nationwide searches with on-line electronic databases.
- We searched public records, such as the Department of Motor Vehicles, Credit Bureau, and death certificates.
- We went back to the schools and obtained high school reunion lists.
- We monitored local papers for wedding announcements. This was especially valuable for tracking women who had already been married, because the local paper provided the names of the bride’s parents, which often corresponded with the names of the subject during the high school years.
- We used a “snowball” approach in which we asked subjects who were participating in the study whether they knew the whereabouts of classmates whom we were unable to track.

- On a trial basis, we used a detective agency to identify the addresses or telephone numbers of subjects whom we were trying to track. For the amount that we were willing to pay the detective agency, using our methods we were equally successful in tracking subjects as the detective agency.

Using these tracking methods, we were able to track 983 (80.7%) of the participants from adolescence to young adulthood to provide them with the opportunity to continue to participate in the study. We were unable to locate 235 participants for the wave 5 interview; however, we have pursued these subjects for the wave 6 interview and thus far have contacted >60 of them.

Conclusions

Data collection for prospective research is a dynamic process, requiring ongoing adaptations to secular changes (e.g., cell phones, caller ID). It is important to obtain personal identifying information (e.g., social security number, driver's license number) to track subjects using many of the major databases (e.g., Department of Motor Vehicles). Sensitivity in collecting such information from subjects is required, as well as clear communication concerning the confidentiality of these data. It is important to try to maintain periodic contact with subjects via birthday and holiday cards. This will also help to inform you about address changes in a timely manner. It is also important to budget sufficient funds for tracking and ongoing maintenance of the subject database. If possible, it is advantageous to establish positive relationships with significant others (e.g., other family members), who may facilitate tracking, although as we found in our study, it is important to maintain these contacts across time. The same parents who had provided informed consent for their child to participate in our study during the adolescent years were distrustful of us several years later even when we clearly identified who we were and informed them of the child's previous participation in the study. Finally, and perhaps most important, be prepared for the unanticipated!

METHODOLOGICAL ISSUES IN LONGITUDINAL SURVEYS WITH ADOLESCENTS AND CHILDREN

Joel W. Grube

This paper describes some lessons that we have learned in the course of conducting longitudinal surveys with children and adolescents over the past 15 years. These surveys have investigated a range of potentially sensitive behaviors, including drinking, drug use, drinking and driving, and sexual experience. We have used a number of survey methods, including telephone, mail, face-to-face interviews, self-administered questionnaires, and most recently computer-assisted self-administered interviews (CASI). This presentation is not largely data driven. Rather, it draws more informally on our experiences over the course of these studies. In particular, I discuss three broad areas: (1) sampling—how do you get representative population samples of young people; (2) survey mode—which survey methods seem to yield the best data on sensitive behaviors with young people; and (3) how to link surveys over time with the same respondents while maintaining confidentiality and anonymity.

Sampling

In terms of sampling, we have used three primary methods: school-based samples, random-digit dial (RDD) samples, and list-assisted samples. I discuss the advantages and disadvantages of each of these approaches in turn.

School-Based Samples—In school-based sampling, respondents are obtained through lists provided by the school. There are certain advantages to using school-based samples. Most notably, it is relatively easy and inexpensive to use school-based sampling. Moreover, the lists are fairly complete if you are targeting youths who are in school. The disadvantages are, first,

that the sample excludes dropouts, chronic truants, absentees, those in alternative schooling, and older youths who have completed school. Some of these groups are known to be at higher risk for drinking and other problem behaviors. As a result, we may be missing some of the youths in whom we are most interested. A second disadvantage is that school-based sampling is becoming increasingly more difficult. In California, where we have conducted most of our surveys, state privacy laws are making it more difficult to survey students in school on sensitive behaviors. Some school districts have even more stringent requirements than the state and preclude surveys on certain topics or altogether. In some cases, school districts have interpreted the state laws as prohibiting providing researchers with students' names and addresses for surveys outside the school setting. These factors all make school-based survey sampling more difficult and introduce unknown biases into the sample if district decisions about participating are not a random event.

RDD—RDD methods begin with a randomly generated list of telephone numbers in known working telephone exchanges. An attempt is made to contact each number, and a household enumeration is taken to determine eligibility—that is, to ascertain whether there is a child in the target age group present. RDD sampling has some advantages. First, it avoids the difficulties of working with schools, including limitations that may be placed on survey subject areas. Second, it may be more inclusive in the sense that you have the opportunity to include dropouts, truants, chronic absentees, those who are in alternative schooling, and older youths. This is particularly important if the behaviors that you are studying are more prevalent in these populations or if these groups are the population of interest. There are also major disadvantages to RDD sampling. In particular, RDD sampling techniques are inefficient and very expensive. In one of our studies, we estimated on the basis of census data that 11% of households had children in the target age group. That meant that we would have to call a minimum of 10 to 12 numbers to locate one eligible household. In actual fact, the number of necessary calls was much higher because contact was never made with a large number of households despite repeated attempts. The growing use of caller ID and caller screening has exacerbated this problem. In addition, response rates are falling on telephone surveys generally. The samples obtained using RDD techniques also may be biased. Although there is 90 to 97% coverage, telephone subscription varies by socioeconomic status (SES) and ethnicity. Thus, RDD techniques may underrepresent lower income or minority respondents. Given the large number of noncontacts, substantial and unknown biases can enter into the sample. Finally, because of the uncertainty about which households are eligible among noncontacts, estimation of response rates gets somewhat complicated. A number of studies that use RDD techniques simply report completion rates: the number of completed interviews divided by the number of known eligible households that were contacted. This method seriously overestimates response rates because it does not take into account that an unknown number of households that were never contacted were actually eligible for the study. It thus is necessary to estimate which proportion of the noncontacted households would have been eligible if you had been able to reach them. Usually, this is done by assuming that the proportion of eligible households among the noncontacted numbers is the same as that among the contacted households. This approach, however, probably underestimates response rates because it is likely that fewer numbers in the noncontacted group are actually eligible.

LIST-ASSISTED SAMPLING

In list-assisted sampling, you purchase a list of telephone numbers that have a higher probability of having eligible households. These lists may be prescreened to eliminate nonresidential numbers. List-assisted samples are often based on consumer information regarding purchases, magazine subscriptions, warranty registration, and related information databases. The advantage of list-assisted sampling is that it is more efficient and thus cheaper than RDD. List-assisted sampling, like RDD, may also include groups excluded from school samples. The

major disadvantages to list-assisted sampling are that it may be nonrepresentative and response rates are difficult to estimate. List-assisted samples are nonrepresentative for all of the reasons that RDD is nonrepresentative (e.g., differential telephone coverage) but also because certain groups (e.g., lower SES) may be underrepresented in the lists because they are less likely to appear in the databases from which the samples are generated. As with all telephone survey methods, increasingly low response rates are becoming problematic. List-assisted sampling, however, can allow you to send advance letters to households describing the study and inviting participation. Such letters may improve response rates.

Survey Administration

In the course of our studies, we have used personal interviews, self-administered questionnaires (SAQ), computer-assisted telephone interviews (CATI), and CASI. Each of these survey modes has advantages and disadvantages.

PERSONAL INTERVIEWS

Face-to-face or personal interviews have some clear advantages over other methods. First, because a trained interviewer administers the survey, they produce very high data quality in the sense that there are few missing items or errors. Personal interviews also allow moderately complex skip patterns and allow open-ended or semistructured questions that are more difficult to implement with some other techniques. Major disadvantages are that personal interviews are very expensive, are not anonymous, and may result in underreporting of sensitive behaviors compared with other techniques. Data entry and editing are also required, adding some expense to the study.

SAQ

SAQ have certain advantages. In particular, they are relatively inexpensive, particularly if given in school, other group settings, or by mail. Importantly, SAQ can provide levels of confidentiality and anonymity that are not possible with telephone or personal interviews. As a result, they may give better estimates of sensitive behaviors. The disadvantages include lower data quality because of missing or skipped items, the inability to build in complex skip patterns, and the difficulty of including and coding open-ended questions. As with personal interviews, SAQ require data entry (or scanning) and data cleaning.

Telephone Interviews

CATI are relatively inexpensive (putting aside sampling costs for RDD surveys) and can allow complex skip patterns. They also produce very high quality data with few missing items and allow open-ended questions. CATI techniques also require minimal data entry and editing. Telephone interviews, however, are not anonymous and may suffer from underreporting of sensitive behaviors. As discussed under RDD sampling, there are also issues of coverage and declining response rates.

Computer-Assisted Self-Administered Interviews (CASI)

CASI are a relatively new development through which survey data are obtained by presenting questions on a computer screen and having participants directly enter their responses. CASI allow complex skip patterns. This is particularly useful in studies of especially sensitive behaviors (e.g., sexuality) in which you may want to skip some respondents over questions that would be inappropriate because of age or level of experience as ascertained from responses to previous items. CASI generally produce very high-quality data. They also allow investigators to embed audiovisual files and other stimulus materials directly into a survey. There is some evidence that CASI improve reporting of sensitive behaviors. The computer

technology can also be used to directly record responses to open-ended questions. Another advantage is that CASI can incorporate audio-assist capabilities whereby the questions are presented aloud through headphones as well as on the screen. It is interesting that in our experience, few youths use the audio assist because it slows the interview process. CASI also require minimal data entry and cleaning. Possible disadvantages are that CASI require respondents to have a minimal level of computer literacy or at least to be comfortable responding on a computer. CASI are also initially expensive to implement because of the costs of obtaining sufficient numbers of computers to run a study efficiently.

Anonymity—In general, it is desirable to maintain anonymity, especially when conducting surveys of sensitive behaviors. Comparisons of national surveys of youths, for example, show that higher reports of drinking and other sensitive behaviors occur for those studies being conducted anonymously and in group settings (Monitoring the Future, Youth Risk Behavior Survey) relative to studies being conducted in the home or using face-to-face interviews (National Household Survey on Drug Abuse). Maintaining anonymity in longitudinal studies in which surveys from individuals must be linked over time is particularly difficult.

SELF-GENERATED CODES

Although rarely used, one viable approach is to use self-generated codes. Self-generated codes allow data to be matched over successive survey waves by using stable attributes (e.g., sex, year of birth, month of birth, middle initial, number of older siblings, initial of mother's first name) to create a unique code for each respondent (e.g., Grube et al., 1989). Although self-generated codes can be used effectively, some elements do not work (e.g., race, ethnicity) because they are not reliably reported. Research and experience suggest that a minimum of seven elements be used. In addition, self-generated codes work best when they are implemented within smaller sampling units (e.g., schools) rather than applied to larger sampling units.

Because young people make mistakes in codes, it is necessary to compensate for respondent errors. To this end "off-one" procedures can be useful. Off-one procedures first pair all surveys that match on all code elements. Then they pair all surveys that match on all elements but one, after eliminating duplicate codes. Off-one procedures can increase matching success by 25% or more, and mismatches occur in <2% of cases (Grube et al., 1989; Kearney et al., 1984). With the use of off-one procedures, matching success rates as high as 90% can be achieved. Moreover, using off-one procedures improves prevalence estimates and does not seem to reduce reliability or bias structural parameters in prediction models (Grube et al., 1989).

CONCLUSION

A number of possible sampling strategies for surveys of young people can be used. Each has advantages and disadvantages. School-based samples are inexpensive but exclusionary and increasingly difficult to implement. RDD sampling is expensive and inefficient but may be more inclusive of dropouts, truants, chronic absentees, and others at risk for problem behaviors. RDD samples, however, may underrepresent lower SES households and some minority groups. List-assisted samples are less expensive and more efficient than RDD. Such samples may be more inclusive than school samples but, as with RDD and other telephone-based samples, may be nonrepresentative.

In terms of survey mode, personal interviews result in very high-quality data but are expensive and not anonymous. They can include relatively complex skip patterns but require data entry. Self-administered surveys have the advantages of being anonymous and inexpensive but can result in lower quality data. It is not feasible to include complex skip patterns in SAQ, particularly those with children or young adolescents as respondents. Data entry and extensive

data cleaning are often required. Telephone interviews can produce high-quality data and are relatively inexpensive. As with face-to-face interviews, they are not anonymous, and response rates are becoming increasingly problematic. CASI generate very high-quality data. They can provide confidentiality and anonymity and allow complex skip patterns and embedded audiovisual stimuli. However, they are initially expensive to implement.

Anonymity seems to be key in obtaining accurate reports of sensitive behaviors such as drinking, drug use, or sexual experience. Self-generated codes are one means of linking data over time while maintaining anonymity. In using such codes, however, it is necessary to compensate for respondent errors. Off-one procedures may be useful in this regard. Mismatching using such procedures is rare (~2%). Off-one matching does not affect reliability, may give closer population prevalence estimates, and does not seem to bias estimates of effect sizes.

THE PITTSBURGH ADHD LONGITUDINAL STUDY: METHODOLOGICAL AND CONCEPTUAL CHALLENGES

Brooke S.G. Molina, William E. Pelham, Jr., Elizabeth M. Gnagy, and Tracey K. Wilson

The Pittsburgh ADHD Longitudinal Study (PALS) is a large longitudinal study of individuals who do and do not have childhood attention-deficit/hyperactivity disorder (ADHD) and are being followed prospectively to study the association between childhood ADHD and alcoholism (principal investigators Molina and Pelham, AA11873). The PALS focuses on the onset, course, and causes of alcoholism with a current emphasis on drinking during the transition to adulthood. The study will also address the development of drug use and drug disorder during this period of development (principal investigators Pelham and Molina, DA12414). The proband sample is the largest of its type, with a wealth of childhood data available to prospectively test hypotheses regarding variability in alcohol and drug disorder vulnerability from childhood characteristics. The comparison sample is demographically similar, and all participants are interviewed annually, which allows us to track the onset, course, and causes of drinking, drug use, and related impairment in the probands as compared with the control subjects.

Sample Ascertainment

The probands initially received a diagnosis of ADHD before their enrollment in the Summer Treatment Program for ADHD (STP) (Pelham and Hoza, 1996), which was conducted between 1987 and 1996 at the University of Pittsburgh Medical Center. Of 519 eligible children, 364 were enrolled into the PALS an average of 8 years later (70% participation rate). Recruitment into the follow-up study began in 1999, and most participants were enrolled within 2 years. Age at wave 1 (the first follow-up assessment) ranged from 11 to 28 with most participants (99%) between 11 and 25 years of age. A subset of these participants also participated in a single follow-up interview that predated the PALS; they were 13 to 18 years of age at that time (e.g., Molina and Pelham, 2003).

Participants without ADHD ($n = 240$) were recruited at the time of follow-up to be demographically similar to the probands but with ADHD as an exclusionary criterion. Most adolescents and four young adults were recruited through several large pediatric practices in Allegheny County (40.8% of sample) that reached a population of patients from diverse socioeconomic backgrounds. The remaining participants were recruited using various advertising and recruitment methods in the greater Pittsburgh area. Groups were matched on age, percentage male (89.2_{ADHD} vs. 88.8_{nonADHD}), and percentage minority (18.2_{ADHD} vs. 15.4_{nonADHD}).

After little or no contact with probands for 3 to 10 years, significant staff and investigator time was required to locate and recruit the probands. For example, we estimate that 9 to 14 attempts to contact probands or their parents were required before they were located and scheduled for an initial interview vs. 3 to 5 attempts for the comparison participants. On average, probands missed their first visit and required rescheduling, despite the use of confirmation letters and telephone calls; most control subjects needed to be scheduled only once. We attribute this difference to the nature of the population that we are attempting to study (disorganization, poor planning and follow-through, increased incentives needed to complete tasks not intrinsically rewarding, etc.).

Methods

In childhood, a number of instruments were used to gather baseline information for the probands. These included a diagnostic interview with parents that included developmental, educational, and treatment history; multiple standardized measures of parent and teacher report of behavior; symptoms, and impairment; individual IQ and achievement testing; and a host of objective observational data gathered over an 8-week STP.

PALS interviews are conducted at the ADD Program; home visits, telephone interviews, and mailed paper-and-pencil questionnaires are used as need dictates. Visits are conducted annually to avoid missing the rapid changes that occur in alcohol consumption during adolescence and young adulthood. Parent participation is solicited regardless of the age of the participant because of likely under-reporting of impairment in the probands (Barkley et al., 2002; Hoza et al., 2002). Many measures are administered via computer to enable efficient data entry, and care has to be taken to ensure that information is entered accurately (e.g., that probands understand and follow the entry format) and to ensure a back-up plan in case of computer failure (e.g., paper and pencil). Data collected include parent, teacher (for adolescents), and self-reports of domains selected for their theoretical relevance to the development of alcohol disorder: alcohol, tobacco, and other drug use; personality and psychopathology; risky or otherwise delinquent behaviors; other domains of current functioning such as work history; intelligence and achievement; attitudes; cognitions; beliefs and values such as alcohol expectancies; parent-child, peer, and romantic relationships; parental psychopathology and family history of alcohol and drug problems; life events and environment, including parenting variables; and marital satisfaction.

Wave 1 interviews required ~5 hr of participant time. Subsequent annual interviews require less time because we trimmed the battery of variables unlikely to change over time (e.g., IQ, personality). Exceptions are based on anticipated variability through adolescence or need for standardized assessments once adolescents enter young adulthood (e.g., the WAIS+ is administered when adolescents turn 17). Consistency of measures between adolescence and adulthood was sought, although exceptions were required [e.g., DISC Ver. 3.0 in adolescence; Structured Clinical Interview for DSM-IV Axis 1 Disorders–Nonpatient Edition in adulthood (First et al., 1996)]. Participant payments are tied to anticipated length of visit; measure order is randomized, and measures were ultimately prioritized scientifically in the event that a briefer interview is necessary to retain the participant in the study (e.g., telephone interview for participant living overseas).

Sample Retention and Scheduling Procedures

To date, retention has been high (>90% retention at waves 2 and 3; wave 4 is ongoing). We attribute the high rate of retention to a number of factors that include a mostly full-time core group of recruitment and interviewing staff, whose skills have developed over time to be resourceful, creative, and persistent yet not coercive. By using a mix of full-time and part-time staff to maximize flexibility, interviews are offered 7 days a week into the evening to

accommodate the wide range of schedules and living arrangements in our sample. Probands require between 22 and 26 attempts to contact on average (including leaving voicemail messages) vs. 7 to 10 attempts for the comparison participants. The need for a high rate of documented communication among staff regarding confirmation success and rescheduling efforts has been paramount. Visits are scheduled within 3 months of the previous visit anniversary date; attempts to schedule continue if the window is missed. Statistical analyses allow age, rather than wave, to be the unit of measurement. Finally, we use a variety of retention procedures common in longitudinal studies (Stouthamer-Loeber et al., 1992), such as collecting contact information for family and friends; sending birthday cards, holiday cards, and newsletters; and using electronic search databases. We also provide referrals and requested information to health care providers and universities (in the case of educational accommodations being requested), return telephone calls promptly and enthusiastically, and in general make ourselves available to the participants. Finally, all of the probands had been treated in our STP, which is a highly valued experience to them (Pelham and Hoza, 1996). The high rates of parent and child satisfaction with that experience in our clinic no doubt contributes to the families' willingness to become and remain involved in the follow-up. To date, only seven individuals have permanently refused further participation in the PALS.

Assessment of Functioning in School

Teacher reports of functioning are critical in the assessment of children, in particular children with ADHD (e.g., Mannuzza et al., 2002). Methods that we find helpful to successfully collect standardized ratings of behavior from secondary education teachers and reports of discipline, grades, and attendance from guidance counselors include (1) dedicating one staff person to have primary responsibility for this task; (2) choosing one teacher; (3) calling the school if measures are not returned within 2 to 3 weeks; (4) payment of \$10 to each teacher/guidance counselor providing information; (5) carbon copy notification to principals; and (6) providing a brief justification for the ratings to teachers, including the importance of their response. These procedures typically result in a return rate of >90%.

Assessment of Interim Functioning

Because functioning between initial treatment and beginning of follow-up is likely to influence the outcomes of interest (e.g., treatment, school functioning), we have developed several measures that enable collection of data on interim functioning. For example, lifetime history of school discipline is collected from schools, parents, and probands; lifetime history of stimulant medication is collected from clinic parents and validated through records of prescribing physicians. Had we thought of the importance of such data before funding, we would have made sure that we collected at least minimal information inexpensively but in an ongoing manner from our original STP participants.

Data Management and Analytic Issues

We have found success with a two-site model of data collection (University of Pittsburgh) and management (SUNY Buffalo). At the data collection level, a computer-administered battery with secure internet link (currently using SPSS Data Entry Web Server software) decreases data entry error and facilitates rapid access to data for analysis. Paper-and-pencil measures are used as need dictates; these data are subsequently entered at the Buffalo site. Data queries from the Buffalo staff are delivered to the Pittsburgh staff for response. The two sites are linked to two local servers, enabling rapid posting of data sets and results for simultaneous viewing and discussion during telephone conferencing. A separate site devoted to data management has meant that one site can focus exclusively on subject retention and running while ensuring rapid checking of data and rapid follow-up of missing data, as well as minimal lag between data collection and data entry.

Funding Issues

The cost of the PALS has increased significantly over time, largely as a result of the increasing cost of subject retention and staff retention. Once staff became sufficiently skilled in the basic responsibilities of the PALS, promotions have followed, involving independent functioning and assumption of primary responsibility for specific projects (e.g., new staff training, collection of school data, monitoring recruitment progress, supervision of data management). We have found it helpful to seek the involvement of multiple agencies (e.g., National Institute on Drug Abuse funds the R01 that supported the addition of measures specifically related to drug abuse). In addition, it has been crucial to recognize over time (1) feasibility of research plans (do not promise more than you can deliver) and (2) hidden or incidental costs associated with a large longitudinal study (e.g., flexibility in participant payments, software needs) of grown children with ADHD.

LESSONS LEARNED IN CONDUCTING RESEARCH ON ALCOHOL INVOLVEMENT: “IF ONLY I HAD KNOWN BEFOREHAND!”

Kristina M. Jackson and Kenneth J. Sher

Preparation of this paper was supported by National Institute on Alcohol Abuse and Alcoholism Grants R01 AA07231 to Kenneth J. Sher, P50 AA11998 to Andrew C. Heath, and R37 AA13987 to Kenneth J. Sher.

Longitudinal, prospective research on alcohol use, abuse, and dependence has increasingly been used, with growing recognition of the limitations of cross-sectional data for characterizing and explaining variation in the course of alcohol involvement and for resolving the potential causal role of various hypothesized risk and protective factors. Our discussion below is a brief survey of salient issues surrounding methods of design, fieldwork, logistics, and analysis.

Design

A number of considerations should be evaluated with respect to the issue of the nature and number of measurement occasions. For example, one consideration is whether the proposed statistical model is statistically identified with the proposed number of time points. Another issue is whether a broader age span can be covered with fewer years by using an accelerated (i.e., cohort-sequential) design. It is also important to cover key developmental periods/events and to sample the respondents frequently enough to resolve temporal processes. The length of the time period to be surveyed should be considered in terms of whether there is sufficient time to observe anticipated changes. Finally, the stability of study constructs should be considered when planning the frequency in which study constructs are reassessed.

Shorter follow-up intervals permit more precision and less retrospection, allow for better resolution of short-term time-bound functional relations, and permit the researcher to be more flexible with content (e.g., addressing with greater frequency only those things that are changing more frequently). Depending on the frequency, they allow one to examine periodicity (e.g., days/week or season); in addition, shorter intervals can be aggregated into longer epochs. However, shorter intervals may require more frequent assessments, which can increase subject burden and cost. Although there are advantages to shorter intervals, longer intervals may be useful for efficiently measuring certain variables that occur with low frequency (e.g., marriage, school entry, serious but rare alcohol consequences such as injury or arrest), again reducing subject burden.

Choosing the rating periods to be used (e.g., lifetime, past year, 30 days, past week) requires considerable thought. With increasing intervals, there is less precision with respect to specific occasions but perhaps greater typicality. It is also important to consider the developmental

period because it has implications for which covariates and outcome measures are used. For example, both proximal and distal nature of covariates may differ as a function of development: parental use/supervision would be more proximal to children/adolescents and more distal to adults. Similarly, developmental tasks and roles may vary depending on the developmental period (e.g., school transition, job entry, marriage). Also, the appropriateness of the measures may change over time, in terms of ecological relevance (e.g., alcohol exposure versus dependence for young adolescents versus older adults). Finally, the stability of drinking changes as a function of age, and there is considerable change in drinking through adolescence/young adulthood but relative stability later in life. If one is interested in charting trajectories, then he or she should focus on those periods when heterogeneity in stability and change is most pronounced.

Data may take the form of surveys (questionnaires, interviews) or more event-based data, each of which has merits. Surveys are well suited for studying large samples, start-up costs can be low, and they have a relatively low and time-delimited subject burden. Ecological data (event level) are collected in the natural environment, can help to resolve within-subject processes, hold “individual difference” third variables constant, and are especially useful for exploring causal directionality and describing context effects. Interviews permit “control” of the assessment and allow probing and detection of motivational issues but require extensive training and quality control and are difficult to conduct in “bursts” (e.g., collecting data over short time periods to prevent confounding with seasonality or holiday effects).

When designing a prospective study, clarity surrounding study goals is paramount. For example, different studies would be designed on the basis of whether the intent is to predict alcohol involvement (etiology) or examine the consequences of alcohol involvement. It is also important to select the outcome measure(s) of interest (e.g., alcohol use, abuse, dependence, subjective effects, acute consequences, long-term consequences). Likewise, whether the researcher is interested in modeling population trends or individual variability is important to recognize initially.

Logistics

Consistency of measurement over time is a major challenge. For example, researchers who conduct prospective clinical research are faced with changing DSM definitions of alcohol abuse and dependence over time and must balance needs for consistency versus currency. Also, participants can be affected by continuing participation and can anticipate questions, perhaps biasing their response. They may also develop a sense of acquaintanceship with interviewers and may be embarrassed to reveal continuing problems. Also, evaluation prompted by surveys can lead to changes in behavior.

The extent to which ostensibly “fixed” covariates should be reassessed is another logistical concern. Some measurements change more than is commonly known (e.g., family history, personality “traits”). Family history can change because of incidence in a close relative or perhaps because individuals learn over time that they are adopted and do not know their biological parents. Larkins and Sher (2003) found that personality changes more than would be expected (stability $r \leq 0.50$) and that personality exhibits normative decreases over young adulthood. Even variables such as age of onset can change systematically as a function of age (Parra et al., in press).

Tracking and maintaining a panel of subjects can be challenging, particularly as secular changes relating to solicitation, junk mail, spam, and caller ID have made contact more difficult in recent years. It is useful to have regular but not too frequent contacts with participants between follow-up occasions (e.g., a regular newsletter that includes tracking information). Having

different sources of tracking information (multiple informants), updated regularly, is extremely important.

The complexity and importance of good data management of prospective research is often underappreciated. Variable names, data set structure, and scoring programs must take into account the nature of the prospective design. Documenting the history of data editing and cleaning is essential. Data dictionaries and variable definitions need to withstand the test of time and need to be comprehensible and thorough for future investigators.

Finally, maintaining staff morale (and staff) over extended periods of time can be difficult. Prospective studies present challenges regarding workflow and changing tasks over time. It is helpful to have staff continuity, but prospective studies pose challenges to job security as work can span multiple funding cycles.

Data Analysis

Prospective studies bring a wealth of complex and challenging statistical issues with respect to how best to model consistency and change over time, how best to account for missing data (whether planned or the result of attrition), and how cohort effects should be taken into account. The choice of approach involves numerous considerations. The data analyst must consider the expected form of the data (e.g., categorical or continuous), type of data (e.g., are covariates time-varying or fixed?), number of observation points, hypotheses to be tested, and sample size (note that power increases with repeated observations). If a researcher is interested in progression and regression, then he or she might use onset/persistence analyses or latent transition analysis. If he or she is interested in identifying risk factors, then the best approach might be an autoregressive model or a latent growth model or, if trait processes are assumed, a state-trait model. Also note that planned missing data can be advantageous for prospective designs: the number of measurement occasions can be reduced without sacrificing power. Finally, given important secular trends in the use of alcohol, cohort can be controlled for by treating it as a fixed covariate or as a grouping variable (i.e., a potential moderator) by using a multiple-group structural equation modeling procedure.

Concluding Comments

All research design is challenging. Prospective research adds multiple layers of new challenges. However, the challenges are usually worth it because the findings permit greater resolution of key descriptive research questions as well as provide critical data for addressing issues surrounding etiology and consequences of alcohol involvement.

DISCUSSION

Nancy L. Day

This research was supported by grants AA06666 and AA14215 (Nancy Day, principal investigator) from the National Institute on Alcohol Abuse and Alcoholism.

Dr. Faden put together a remarkable symposium, and it was a pleasure to hear each of the speakers discuss various aspects of longitudinal research. There is no question about the importance of longitudinal studies. They allow us to understand the natural history of a disease, to chart the relations between a risk factor and the development of a disorder, to look at patterns of development, to evaluate the long-term effects of an exposure, and many other aims that could be identified. The talks have illustrated a number of these uses.

What is it that these studies have in common? All longitudinal studies, however disparate in their specific aims, have one underlying phenomenon in common: they all use change as an

intrinsic part of their research design. It is the best approximation to an experimental design for those of us who do human research. Hypotheses can be framed around the concept of change in a vast number of ways to answer a research question. It is limited only by the imagination of the researchers. Furthermore, understanding the effects of change on a person, a population, or a disease is one way that we can demonstrate causality.

The study discussed by Dr. Molina capitalizes on the change from adolescence to young adulthood and studies the effects of this change on alcohol use among those who have ADHD. Dr. Windle's study also uses developmental change to define trajectories to health outcomes. Another way to use change is to identify people with and without a risk factor and follow them across time to measure the relative incidence of disorder; the research reported by Dr. Jackson was an example of this design. These are very different models; one uses change as a stressor, another as part of the natural history, while a third uses change in time to follow the development of a disorder.

The other side of change, nonchange, is also an element in study design. We might hypothesize that the effect of a risk factor will be the same across different developmental stages or across racial or ethnic groups. Failure to change can also be a risk factor, as when a developmental stage is blocked.

Change, however, can also play havoc with a carefully designed protocol. Some of the problems and some excellent solutions to the effects of change on a study design were illustrated in detail and with great angst by the presenters. As each has noted, to conduct a successful longitudinal study, it is critical to anticipate potential problems ahead of time and plan for and around them. It is equally important to recognize that, despite your best planning, change happens. It will disrupt your research, and you have to be quick on your feet when it occurs and design into your protocol some amount of flexibility to be able to adjust for surprises.

Anticipated change was illustrated by each of the talks. Over time, the cohort will age. This leads to problems in measurement, for example, as an instrument that is appropriate at one time may not be applicable at a different age. It also leads to attrition as members of the cohort are lost from death or illness. In addition, the longer the interval, the more likely it is that recall will be less accurate, necessitating the development of instruments that help trigger memory. Moreover, people feel free to move, change their names, disappear, and even refuse to participate in the succeeding phases of a project. On the positive side, however, people also reappear and often will agree to participate in a later phase. Each of these problems leads to a loss of internal and external validity in the study design, and, as a result, a considerable amount of effort is spent by investigators in minimizing them.

Unanticipated change arises in all kinds of ways. A change in treatment can affect the outcomes that you are following; political changes, such as the Health Insurance Portability and Accountability Act, may affect your ability to recruit subjects; social changes may make the behavior in which you are interested more or less prevalent or different altogether. The development of new diagnostic criteria can render your data obsolete. Dr. Grube illustrated some of these problems in his talk. Indeed, as Dr. Jackson reported, characteristics that we have thought were stable, such as temperament and sex, now also seem to be changeable. On the positive side, though, there are often new developments in methods, instruments, and data analysis techniques that can be incorporated into an ongoing study.

So, in a longitudinal study, change is a key element. With the creative use of change, researchers have been able to test hypotheses and demonstrate causality, define the effects of risk factors, monitor the natural history of an outcome, and evaluate the long-term effects of change on the outcome of interest. No other study design enables these kinds of research questions. At the same time, however, change creates major problems in longitudinal studies, and most of our

efforts are spent trying to prevent change that is not part of the research design from affecting the validity of the data (Davies and Windle, 1997, 2000, 2001; Davies et al., 1999; Domenico and Windle, 1993; Parra and Sher, 2001; Tubman et al., 1996; Windle, 1992, 1994a, 1994b, 1996b, 2000a, 2000b; Windle and Windle, 1996, 1997, 2001; Windle and Davies, 1999).

References

- Barkley RA, Fischer M, Smallish L, Fletcher KE. The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. *J Abnorm Psychol* 2002;111:279–289. [PubMed: 12003449]
- Davies P, Dumenci L, Windle M. The interplay between maternal depressive symptoms and marital distress in the prediction of adolescent adjustment problems. *J Marriage Fam* 1999;61:238–254.
- Davies P, Windle M. Interparental discord and adolescent adjustment trajectories: the potentiating and protective role of intrapersonal attributes. *Child Dev* 2001;72:1163–1178. [PubMed: 11480940]
- Davies PT, Windle M. Gender-specific pathways between maternal depressive symptoms, family discord, and adolescent adjustment. *Dev Psychol* 1997;33:657–668. [PubMed: 9232381]
- Davies PT, Windle M. Temporal linkages between dating relationship pathways and trajectories of psychosocial adjustment among middle adolescents. *Merrill-Palmer Q* 2000;46:90–118.
- Domenico D, Windle M. Intrapersonal and interpersonal functioning among middle-aged female adult children of alcoholics. *J Consult Clin Psychol* 1993;61:659–666. [PubMed: 8370862]
- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV Axis I Disorders–Nonpatient Edition (SCID-1/NP, Version 2.0). Biometrics Research Department, New York State Psychiatric Institute; New York: 1996.
- Grube JW, Morgan M, Kearney KA. Using self-generated identification codes to match questionnaires in panel studies of adolescent substance use. *Addict Behav* 1989;14:159–171. [PubMed: 2786325]
- Hoza B, Pelham WE, Dobbs J, Owens JS, Pillow DR. Do boys with attention-deficit/hyperactivity disorder have positive illusory self-concepts? *J Abnorm Psychol* 2002;111:268–278. [PubMed: 12004833]
- Kearney KA, Hopkins RH, Mauss AL, Weisheit RA. Self-generated identification codes for anonymous collection of longitudinal questionnaire data. *Public Opin Q* 1984;48:370–378. [PubMed: 10265980]
- Larkins JM, Sher KJ. Stability of family history-related personality differences over the third decade of life (abstract). *Alcohol Clin Exp Res* 2001;25(suppl 5):44a.
- Mannuzza S, Klein RG, Moulton JL III. Young adult outcome of children with “situational” hyperactivity: a prospective, controlled follow-up study. *J Abnorm Child Psychol* 2002;30:191–198. [PubMed: 12008657]
- Molina BSG, Pelham WEJ. Childhood predictors of adolescent substance use in a longitudinal study of children with ADHD. *J Abnorm Psychol* 2003;112:497–507. [PubMed: 12943028]
- Parra GR, O’Neill SE, Sher KJ. Reliability of self-reported age of substance involvement onset. *Psychol Addict Behav* 2003;17(3):211–218. [PubMed: 14498815]
- Pelham, WE.; Hoza, B. Intensive treatment: a summer treatment program for children with ADHD. In: Hibbs, E.; Jensen, P., editors. *Psychosocial Treatments for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*. APA Press; New York: 1996. p. 311-340.
- Stouthamer-Loeber M, van Kammen W, Loeber R. The nuts and bolts of implementing large-scale longitudinal studies. *Violence Victims* 1992;7:63–78.
- Tubman J, Windle M, Windle R. The onset and cross-temporal patterning of sexual intercourse in middle adolescence: prospective relations with behavioral and emotional problems. *Child Dev* 1996;67:327–343. [PubMed: 8625716]
- Windle M. A longitudinal study of stress buffering for adolescent problem behaviors. *Dev Psychol* 1992;28:522–530.
- Windle M. A study of friendship characteristics and problem behaviors among middle adolescents. *Child Dev* 1994a;65:1764–1777. [PubMed: 7859553]

- Windle, M. Coexisting problems and alcoholic family risk among adolescents. In: Babor, TF.; Hesselbrock, V.; Meyer, RE.; Shoemaker, W., editors. *Types of Alcoholics*. New York Academy of Sciences; New York: 1994b. p. 157-164.
- Windle M. An alcohol involvement typology for adolescents: convergent validity and longitudinal stability. *J Stud Alcohol* 1996;57:627-637. [PubMed: 8913994]
- Windle M. A test of the adolescent maturational hypotheses of delinquency using latent growth curve modeling. *Appl Dev Sci* 2000a;4:193-207.
- Windle M. Parental, sibling, and peer influences on adolescent alcohol use and alcohol problems. *Appl Dev Sci* 2000b;4:98-110.
- Windle M, Davies PT. Depression and heavy alcohol use among adolescents: concurrent and prospective relations. *Dev Psychopathol* 1999;11:823-844. [PubMed: 10624728]
- Windle M, Windle RC. Coping strategies, drinking motives, and stressful life events among adolescents: associations with emotional and behavioral problems, and academic functioning. *J Abnorm Psychol* 1996;105:551-560. [PubMed: 8952188]
- Windle M, Windle RC. Depressive symptoms and cigarette smoking among middle adolescents: prospective associations, and intrapersonal and interpersonal influences. *J Consult Clin Psychol* 2001;69:215-226. [PubMed: 11393599]
- Windle RC, Windle M. An investigation of adolescents' substance use behaviors, depressed affect, and suicidal behaviors. *Am J Child Psychol Psychiatry* 1997;38:921-929.