Published in final edited form as:

Newborn Infant Nurs Rev. 2010 March 1; 10(1): 10–18. doi:10.1053/j.nainr.2009.12.003.

Potential Pitfalls in Collecting and Analyzing Longitudinal Data from Chronically Ill Populations

Diane Holditch-Davis, PhD, RN, FAAN [Marcus Hobbs Distinguished Professor of Nursing; Associate Dean for Research Affairs]

Duke University School of Nursing Box 3322 DUMC Durham, NC 27710

Janet Levy, PhD [Assistant Research Professor]

Duke University School of Nursing Box 3322 DUMC Durham, NC 27710 janet.levy@duke.edu

Abstract

Research on the prevention and management of chronic illnesses involves understanding changes in complex and interrelated aspects of each individual. To capture these changes or to control for them, nursing and health research needs to be longitudinal. However, there are a number of potential pitfalls in analyzing longitudinal data from a chronically ill population. This paper will examine four major pitfalls: selection of time points, measurement, choosing appropriate statistical procedures, and missing values. Although the frequency of data collection is often driven primarily by practical concerns, it will affect the results. In addition, outcome measures may capture different constructs at different points in times. Traditional analysis techniques often have assumptions about data characteristics that are violated in clinical populations. Missing values are common in research with chronically ill individuals because of problems of subject retention and because individuals have frequent medical complications. Solutions to these pitfalls are also discussed.

Keywords

Longitudinal research; Data analysis; Data collection; Measurement; Missing values

Preventing and managing chronic illness is the focus of much of nursing and health care research. Chronic illnesses, including health problems that result in prolonged alterations in functioning (e.g., prematurity, stroke), are the major causes of morbidity and mortality in the United States.^{1–7} Chronic illnesses affect all aspects of a person's life--physical, psychological, social, and economic.^{8,9} Chronic illnesses change over time, and individuals with the same illness have very different health trajectories.^{10–13} Individual responses to chronic illness are affected by trajectories in other areas of human functioning, including physiological and psychological development^{14–18} and transitions between sites of care.^{19,20} Most aspects of preventing and managing chronic illness occur over time, including activities to follow the medical regimen,^{21,22} deal with the emotional aspects of the illness,^{23–26} minimize symptoms,^{27–29} prevent complications,^{16,30,31} and limit disease progression.^{9,32} Thus, longitudinal research is needed to understand chronic illnesses.

© 2009 Elsevier Inc. All rights reserved

diane.hd@duke.edu 919-684-8862 voicemail; 919-681-8899 fax.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Recent reviews of nursing research have cited many instances in which longitudinal studies would enhance the understanding of individuals' chronic illness experiences and lead to more effective interventions.^{33, 34} However, analyzing longitudinal data from a population whose conditions and activities are changing can result in a number of potential pitfalls. For example, a review of programs of research on preterm infants found that the researchers often ignored the longitudinal nature of their data in analyses, compared data from only two time points using correlations, or identified developmental patterns only descriptively.³⁵ This paper will examine four major pitfalls--selection of time points, measurement, choosing appropriate statistical procedures, and missing values--and offer solutions to them. To illustrate these problems, we will draw upon data collected by the first author in several studies of the outcomes of preterm infants.^{e.g., 15,16,25}

Selection of Time Points

The first potential pitfall of longitudinal research is inappropriate selection of the specific time points to be studied. Longitudinal studies are done for one of two reasons: to identify the presence and direction of change or to describe the shape of change. Studies focused on the presence and direction of change require only two time points (a third would be desirable): one at baseline and one or two after the major amount of change has occurred. On the other hand, studies of the shape of change require that sampling be done frequently enough (three or more time points) to capture the shape of the change process and natural turning points and transitions. More time points will be needed during periods of rapid change than during stable periods and to describe phenomena with complex shapes. If too few time points are selected or they are spaced too far apart, the study will be unable to identify the underlying shape of the change process. For example, if a researcher wants to show that an intervention has had an effect, only baseline and a single time point when the intervention's effects are strongest are needed, but if the researcher wants to show how the intervention's effects change over time, multiple time points will be required to determine when the intervention first has an effect, when the intervention has the strongest effect, whether the intervention has any delayed effects, and how rapidly the intervention's effects decline.³⁶

However, selecting the appropriate number and frequency of contacts is difficult in studies of chronic illnesses. Phenomena differ on how quickly they change. When Holditch-Davis and colleagues studied the development of sleep and waking states in preterm infants, weekly observations of sleep were required to show the developmental patterns: active sleep decreased with age while quiet waking, active waking, and quiet sleep increased.³⁷ However, examining changes in mothers' depressive symptoms after their preterm infants were discharged from the hospital required contacts only every 2–6 months to show that depressive symptoms were highest during hospitalization, decreased until 6 months after discharge, and then remained stable.³⁸

Also, some phenomena such as pain or temperature may show fluctuations or cycles in addition to the systematic changes that interest the researcher. Testing for cyclicity may require 50 or more time points.³⁹ If the researcher is only interested in longer term systematic changes, fewer time points can be used but only if they consistently occur at the same point in cycle. When measures are taken more or less often than the cycle length, data collection will occur at different points in the cycle, and the cycle may be mistaken for systematic change.³⁹ Unfortunately, researchers often do not know the exact cycle length so they will need to use more time points.

Yet another factor affecting the timing of contacts is the characteristics of the measures. More variable measures require more time points to get a stable estimate of an individual's values than do stable measures. An example of a variable measure is the behavior of mothers and

young children during their typical daily activities. Although there are undoubtedly underlying patterns of behaviors, mother and child interactive behaviors vary dramatically from day to day. Thus, when Holditch-Davis studies maternal and child interactive behaviors, the behaviors are measured every 10 seconds over at least 45 minutes.^{e.g., 40, 41} The percentage of 10-second periods out of the total observation spent in each variable is calculated. Even then, the behaviors are variable from week to week. Table 1 shows the correlations of a few mother and 3-year-old prematurely born child behaviors obtained from two 2-hour observations conducted 1 week apart.⁴⁰ All but one of the 14 variables in Table 1--child touch--was significantly correlated between the two observations, but none of the correlations were as high as .70, the usual indicator of a stable measure. Correlated *t*-tests were also calculated to determine whether the values for the two observations differed (see Table 1). Three variables--child look, child touch, and child play--differed significantly between observations. The children looked at and touched their mothers more during the first observation and played more during the second one probably as a result of greater comfort in the observational situation on the second observation. To see systematic changes over time in unstable variables such as these behaviors, contacts need to be obtained more frequently than for more stable variables.

One of the reasons that selecting the specific time points to be studied is so difficult is that these decisions cannot be made for purely methodological reasons. They are affected by practical considerations such as attrition and study costs. Also, longer time periods between contacts result in a greater chance that events unrelated to the phenomenon of interest may become confounded with systematic change, the focus of longitudinal research. For example, in Holditch-Davis's first study of sleep-wake state development, the nursery population changed half way through the study to include a greater percentage of seriously ill infants.⁴² As a result, adding a covariate for illness severity was necessary. These events are usually called history or period effects and confound the underlying pattern of change, which is called maturation. There is no way to prevent history effects, but researchers need to be aware of the potential for these effects to occur, and add appropriate covariates if needed, as Holditch-Davis did. In addition, shorter follow-up periods reduce the likelihood that history effects will occur.

On the other hand, shorter times between contacts increases the burden on study participants and the chance that they will remember the instruments, a situation called instrument reactivity.^{43–44} Instrument reactivity is a potential pitfall for longitudinal research because participants' memories of an instrument and their previous answers may affect their current answers, thereby distorting the findings. Grouping multiple instruments together so that contacts are longer and less frequent also increases burden of each contact for participants but decreases overall study costs.

Therefore, selection of the specific time points to be studied in a longitudinal study is an important research decision that always requires balancing theoretical considerations of the goals of the study, the nature of the underlying phenomenon of interest, and characteristics of the measures used with practical considerations such as the potential of subject attrition, history effects, instrument reactivity, and study costs.

Measurement in Longitudinal Research

Inaccurate measurement is another potential pitfall in longitudinal research. Measures used in longitudinal studies need to be sensitive to change, especially change within individuals; be reliable over time; and exhibit minimal maturational transformation in either the instrument or the underlying construct.^{43–45} A measure's sensitivity to change is its ability to detect changes in individuals over time. Thus, instruments designed to measure stable constructs (e.g., trait anxiety) are not appropriate for use as outcomes in longitudinal studies. Instead, instruments

that are able to capture the expected changes over time in the phenomenon of interest need to be selected.

Longitudinal research also increases the likelihood that instruments may become invalid over time. Longitudinal study involves repeated administration of instruments, increasing the chance of memory of the instrument and of influencing participant answers through the study's ongoing attention to a phenomenon. Pitfalls due to memory of instruments can be reduced by spacing contacts far enough apart to minimize participant memory of instruments, providing instruction to participants to answer the questions as they are currently feeling without regard to previous answers, and by using alternative forms of the instrument when available. Problems due to participant familiarity with topic are more difficult to deal with. The major approaches to minimizing this problem are researcher awareness of the possibility and, when the design permits, starting participants at different points in the process under study so that the responses of participants with different experience with the instruments can be compared. For example, if a researcher is interested in studying symptoms during chemotherapy, some participants can be enrolled before starting chemotherapy, some after one or two sessions, and others after several. Symptom responses of these groups of participants can be compared at similar points in the chemotherapy process, such as after several sessions and at the last session. If their responses are different, memory of the instruments is likely to be affecting the participant responses.

Further, most instruments require the participants to report on past events, whether it is what they ate yesterday, felt in the past week, or their activity in the past month. Thus, choosing measures that are minimally subject to errors of memory, or recall bias, is important. For example, it is easier to remember specific events over short periods of time. Participants are more likely to remember accurately what they ate yesterday than what they typically ate over the past year.

Finally, the instruments for longitudinal studies need to be free of maturational transformation, or changes in what the instrument is measuring at different time points. One cause of maturational transformation is a construct that is subject to change over time (such as motor abilities in infants) because these types of constructs cannot be accurately measured with the same instrument over time. For example, in one of Holditch-Davis's studies of the developmental outcomes of preterm infants,^{16,37,41} the Bayley Scales of Infant Development⁴⁶ were administered when the infants were 9 and 27 months of age corrected for prematurity. Table 2 shows the means, standard deviations, and correlations of these data. The means were lower at 27 than 9 months and the standard deviations at 27 months were larger. Correlations between the two sub-scales at the same age were as high as the correlations between same sub-scales at 9 and 27 months. Yet the Mental Development Index (MDI) and Psychomotor Development Index (PDI) are supposed to be measuring different constructs: cognitive development for the MDI and motor development for the PDI. If each sub-scale was really measuring the same construct at the two ages, correlations would be expected to be higher for the same sub-scale at the two ages than between different sub-scales at the same age. These findings indicate that the Bayley Scales, despite being a well-validated instrument, were measuring different constructs in this sample at the two ages.

Another form of maturational transformation occurs when the measurement characteristics of an instrument change over time. The assumption that an instrument is not subject to maturational transformation is frequently violated in longitudinal studies though it is usually a greater issue for studies of children or frail older people, because of their more rapid development, than for studies of adults. Evidence of maturational transformation is found in changes over time in instrument psychometrics (internal consistency, variances, factor analysis). This form of maturational transformation can also occur over time in non-age-related

phenomena. Table 3 shows results from an ongoing study by Holditch-Davis and colleagues. In this study, mothers of preterm infants were assigned to provide interventions to their infants while they were in the neonatal intensive care unit and for 2 months after hospital discharge. Study nurses taught the interventions and contacted the mothers weekly during hospitalization and biweekly after discharge. At hospital discharge and when the infants was 2 months of age corrected for prematurity, the mothers rated their satisfaction with their intervention on an investigator-designed questionnaire with two sub-scales: the first rated the mother's satisfaction with the intervention and the second her view of the helpfulness of the study nurse. Although the mothers rated the intervention positively at both time points and the two sub-scales were significantly correlated at each time point, there were no correlations between the two time points (see Table 3). In addition, when the individual items were factor analyzed, the results were different at the two time points. Although five factors were found at each age, the items making up the factors differed. For example, Factor 1 consisted of items 9, 13–19, and 21–23 at discharge but of items 4 and 14 at 2 months. These findings suggest that this questionnaire was measuring something different at the two time points and perhaps that maternal satisfaction with the intervention was influenced by different factors during hospitalization and 2 months later.

Therefore, selection of instruments for a longitudinal study is an important research decision that requires careful consideration before the study and evaluation of instrument characteristics once data has been collected. Inaccurate measurement can occur because of lack of instrument sensitivity to change, unreliability of instruments over time, recall biases in the participants, or maturational transformation in the instruments or underlying constructs being measured. When instruments are unreliable or invalid over time, the researcher will be unable to model phenomena longitudinally.

Selection of Appropriate Longitudinal Analysis Techniques

Another potential pitfall of longitudinal research is the selection of appropriate analysis techniques. The selection of a particular statistical model is both an art and a science, though increasingly more the latter. Furthermore, the selection of an inappropriate model can often lead to inaccurate results. Fortunately for researchers, statistical science has an impressive array of tools with which to analyze longitudinal data, even when considering only the case of the normally distributed dependent variable.^{47–52} The task is one of determining whether the statistical model is a reasonable fit for the processes generating the data and can simultaneously provide information about the research question. Because we can never be entirely certain about the former, researchers must rely on experience and intuition in making these choices. In this paper, we will limit our discussion to statistical models for continuous and normally distributed dependent variables.

Many research questions in the area of chronic illness examine the nature of a trajectory of measurements obtained over time on a group of participants. For example, with data on preterm infants,^{15,16,37,41} at least three processes play a large role in the selection of the statistical model. First is the presence of “unbalanced” data because infants are often measured at different points in time. This occurs because infants are born at different gestational ages, become healthy enough to study at different chronological ages, and are discharged home at different post-menstrual ages.

Secondly, as the phenomenon under study “emerges”, it may do so at different rates in different infants. For example, preterm infants once off mechanical ventilation have irregular respiratory patterns due to frequent apnea and brief respiratory pauses that decrease over the preterm period. However, the amount of apnea and respiratory pauses and the rates at which these respiratory variables disappear vary greatly among infants with some infants showing minimal

apnea even at early post-menstrual ages and others having significant amounts of respiratory pauses even at term age.⁵³ As a result, the variability in measures increases explosively or decreases as time progresses. This phenomenon is clearly illustrated in Table 4, which shows data based on the weights of a group of African American preterm infants recorded at six points in time from birth through 24 months of age corrected for prematurity.²⁵ On average, variability in weights increased over time reflecting the fact that infants gained weight at different rates over the first 2 years of life.

Third, the extent to which measures are related to one another over time is often complex. Observations collected at time points closer together are often more highly correlated than those collected at time points further apart. Alternatively, as the infant matures, the correlations between observations collected at relatively close time points increases, reflecting the emergence of stable phenomena. This is again shown in the preterm infant weight gain over the first 2 years of life (see Table 4). For this sample of infants, the correlation between weight at birth and 2 months of age was .29, between weights at 2 and 6 months the correlation jumps to .70, and between 18 and 24 months it is .88, but the correlation between birthweight and weight at 24 months was only .25.

We will review three classes of statistical models that might be used to answer longitudinal research questions with a normally distributed dependent variable: independent *t*-tests, repeated measures analysis of variance (RMANOVA), and the general linear mixed model (mixed model). The pitfalls of utilizing each class of model for answering longitudinal research questions with chronic illness data that frequently manifest the three processes described above will be discussed.

A series of independent *t*-tests, examining group differences at each point in time separately, has been used with longitudinal data, especially to test an intervention.^{e.g., 54} This approach does not directly test hypotheses about trajectories over time. The best it can do is to support inferences regarding the similarity or differences in two trajectories, by determining whether or not groups differ at each time point. This is the most glaring pitfall of using this approach with chronic illness data. Because this approach is bivariate rather than multivariate, the three issues of unbalanced data, differing variances, and complex correlations over time do not affect it.

Repeated measures analysis of variance, a variant of the classical analysis of variance, has historically been a very popular model for the analysis of longitudinal data.^{e.g., 49,51,55} Because of its multivariate nature, this class of statistical models allows for the testing of hypotheses about trajectories and, through interaction terms, to determine how different those trajectories are conditioned on levels of another factor. The major pitfall in using RMANOVA with chronic illness data lies in its requirement that there be no missing data. If missing data occur, researchers must either eliminate any case that includes missing data at even one point in time, or use a variety of techniques to impute the missing data. The former can result in a pronounced loss of power or severe bias.^{40,56,57} The latter is more feasible given the proliferation of methodologies for this purpose^{58–62} but adds an additional layer of complexity to any analysis.

RMANOVA relaxes the assumption that observations are independent to accommodate the obvious dependence between observations over time collected from each participant. This model assumes that correlations between time points are all of the same size and direction regardless of how far apart in time they have been collected.⁵⁹ As the growth data in Table 4 shows, much longitudinal data fails to meet these assumptions. Similarly, the model assumes equal variance at each time point.⁵⁹ When processes generating the data do not fulfill these assumptions, the resulting tests of statistical significance are in error.⁵⁹ This is another pitfall in the use of these models for longitudinal data in chronically ill individuals. For example,

because preterm infants develop at varying rates over the neonatal time period, one expects to see increasing variability in measurements as a particular phenomenon emerges or decreases as it disappears, as previously discussed. Similarly, we would expect the covariances between measurements over time to change.

The linear mixed model, also called hierarchical linear models or mixed models, is an extension of the general linear model that allows specification of fixed and random effects related to the “independent variables.”^{47–50} Three components are estimated: fixed effects, random effects, and within-subjects covariation in measures over time. The fixed effects are similar to the multiple regression parameters and estimate population regression lines. The random effects specify each subject's regression line as an increment in intercept and slope between his or her regression line and the population regression line. This class of models allows specification of how measurements vary over time within subjects. A particular form of the linear mixed effects model is actually called the “growth curve model.”^{63–65} The increasing or decreasing variability between subjects across time is easily specified as are more complex types of within subject correlations over time. Because of their ability to accommodate a variety of processes likely to generate longitudinal data, mixed models generally will yield more accurate hypothesis tests than RMANOVAs.^{58,59} That is, the hypothesis tests on the fixed portion of the model specifying the *overall* (also referred to as the *marginal*) relationship between time and the dependent variable of interest are more accurate. This particular aspect of mixed linear models will be discussed again in the section on missing data.

Finally, mixed models are able to answer research questions regarding trajectories over time for a particular group and whether or not those trajectories are different conditioned on levels of another variable (for example subgroups). A defining feature of mixed longitudinal models is the modeling of variability in individual trajectories over time, rather than the trajectories themselves. Individual trajectories can be examined graphically. However, statistical inferences about them are not typically performed due to the small number of data points usually available and large amount of measurement error present in the data for a single subject.⁵⁹ Mixed models also allow independent variables to vary over time, which is important because the trajectories of chronic illness are embedded in other trajectories such as physical and psychological development, which also vary over time. For example, in a study of the development of sleep-wake states in preterm infants, Holditch-Davis and colleagues found that respiration was more regular in active sleep when infants were being treated with a methylxanthine, a variable that varied from week to week.³⁷

Potential pitfalls of these models are primarily related to the degree of understanding required to execute them. Researchers must be familiar with both the underlying statistical model as well as the particular software package chosen to fit them. Additionally, if small sample sizes are utilized with these models, parameter estimates may lose stability.^{56,57}

Missing Data in Longitudinal Research

The fourth potential pitfall of longitudinal research is missing data. In many studies of chronically ill populations, there are large amounts of missing data. Attrition is always a problem because of the individual nature of chronic illness trajectories, acute health crises, unpredictable hospitalizations, and relatively high mortality rates.^{e.g., 17, 37} A longer time between contacts provides more time for voluntary and involuntary attrition and more chance for historical events to become confounded with systematic change. If contacts are spaced too far apart, participants may forget they are in a study and decline further involvement when contacted. For example, Holditch-Davis has learned that attrition rates increase in studies of premature infants and their mothers when contacts are further apart than 6 months.

In addition, individuals are often available for study at only certain points in time (or at certain ages). To obtain observations across a relatively wide range of ages or time points, one may have to observe many participants, each one providing information at only a fraction of the total number of time points. For example, Holditch-Davis and colleagues examined the development of sleeping and waking states in 134 preterm infants.³⁷ Weekly observations were obtained between 28 and 55 weeks post-menstrual age. But no infant provided more than 9 weeks of data, and 80% provided 5 weeks or fewer. The reason for this was that infants were observed in the hospital only after their health stabilized, and they were followed until discharge or 43 weeks post-menstrual age. An additional observation was obtained about 2 months after discharge. As a result, relatively few observations were available for weeks 28, 29, and 30, the earliest ages that infants were available for study as their health stabilized after birth. Similarly, there were relatively few observations for infants at 40, 42, and 43 weeks as the infants were being discharged. If all of the time points on which a given infant lacked sleep and waking state data were considered to be “missing,” then it would not have been possible to analyze this data with older methodologies such as RMANOVA.

Much statistical literature focuses on the treatment of missing data in longitudinal research.^{66–69} Most discussions have been heavily influenced by Rubin's categorization of missing data as one of three types: Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR).⁶⁹ These three situations are defined based on the extent to which the presence or absence of data is related to the values of the missing data themselves. Data are MCAR if their presence or absence is completely unrelated to any data in the study or the values of the missing data itself. An example of missing data completely at random is if the presence or absence of temperature data in a longitudinal study were due to very infrequent and random surges in electricity that damaged one or two temperature readings per hour.

On the other hand, data are MNAR when its presence or absence is directly related to the values of the missing data itself. For example, when studying rapid eye movements (REMs) in infant sleep states, the observer may be only able to discern the largest REMS because of poor visibility, so small REMs are not recorded. Missing at Random is the situation where the presence or absence of missing data is related to the dependent variable of interest only through its relationship with other variables measured in the study. For example, in a longitudinal study of preterm infant temperature, the presence or absence of temperature data may be due to the frequency of clinical procedures that the infant undergoes.

In the case of the study of sleep-wake state development in preterm infants,³⁷ we would argue that the presence or absence of sleep state data is related to the medical health of the infant. Data are more likely to be missing at earlier ages for infants who take longer to stabilize, and data are more likely to be missing near ages 40, 41, and 42 for infants who are healthier and have been discharged. However, the availability of sleeping and waking information on a child is not directly related to *particular values* of sleeping and waking states that are missing. Just because we know that infants who have more severe medical problems had less sleep states recorded at 28, 29 and 30 weeks post-menstrual age, does not mean that we can identify particular values of those sleep states that are missing (e.g., all values less than 20%). For these reasons, we argue that in this study the sleep and waking states were Missing at Random, or MAR.³⁷

A variety of statistical methodologies can handle missing values in order to generalize to populations from which the observed and missing values were generated. First, if the data are missing at random or completely at random, then hypothesis tests via maximum likelihood estimation provide valid results.^{60,67} For longitudinal data in which the dependent variable is assumed to be normally distributed, this means fitting the general linear mixed model, and

testing hypotheses about model parameters with likelihood ratio tests. Theoretically, when data are missing at random or completely at random, statistical hypotheses about parameters in the linear mixed model using likelihood ratio tests are accurate (in that they provide accurate probabilities). With the likelihood ratio approach, missing data values are not estimated directly; the test statistics are simply based on all available data. It should be noted that other methods of testing hypotheses are *not* valid under the missing at random assumption. That is, they do not provide valid inferences to the population of missing and non-missing data values (the probability values yielded by hypothesis tests will be in error to an unknown degree). Examples of these methodologies are Least Squares Estimation and Generalized Estimating Equations.^{59, 66}

If, on the other hand, the data are missing not at random, and the researcher still wishes to infer to the population which generated both the missing and non missing data, the missing data must be directly estimated prior to modeling. As stated above, many methodologies are available with which to do this,^{58–62} but description of these methods is beyond the scope of this article.

Regardless of the statistical methodology or model ultimately chosen, good scientific practice dictates that the researcher investigate and understand the likely reasons for the missing values. A thorough understanding of why and how data are missing in each study allows researcher to design future studies to minimize missing data.

Conclusions

In conclusion, longitudinal research is the ideal strategy for nurses and other health care providers to understand the prevention and management of chronic illness. Longitudinal studies are needed to enhance the understanding of individuals' chronic illness experiences and lead to more effective interventions. However, there are a number of potential pitfalls in longitudinal studies conducted using a population, such as individuals with a chronic illness, whose conditions and activities are changing. In particular, researchers need to be careful to select an adequate number of time points to capture the phenomenon under study while not including so many time points as to increase subject burden or study expense unmanageably. Inaccurate measurement can occur in a longitudinal study because of lack of instrument sensitivity to change, unreliability of instruments over time, recall biases in the participants, or maturational transformation in the instruments or in the underlying construct being measured. Selection of the correct statistical techniques is an important issue for longitudinal research because unbalanced data, differing variances, and complex correlations over time are common in chronically ill populations and can easily affect statistical results. Recently developed statistical procedures, particularly the general linear mixed model, can deliver appropriate results despite these problems. Finally, missing values are common problems in studies of chronically ill individuals because of the individual nature of illness trajectories, acute health crises, unpredictable hospitalizations, and relatively high mortality rates. For longitudinal data in which the dependent variable is normally distributed, statistical hypotheses on parameters of the general linear mixed model using likelihood ratio tests will provide valid inferences when the missing data are missing at random or completely at random.

Acknowledgments

The preparation of this paper was partially supported by grants R01 NR009418, R01 NR01894, and R01 NR05263 from the National Institute for Nursing Research, NIH to the first author.

We wish to acknowledge John Boling's technical assistance and the contributions of Michael Belyea, Lloyd Edwards, Todd Schwartz, and Mark Weaver to the analyses in this manuscript.

References

1. Gallagher DM. Current clinical issues impacting the lives of patients living with HIV/AIDS. *J Assoc Nurs AIDS Care* 2007;18(1 Suppl):S11–16.
2. Hwang W, Weller W, Ireys H, Anderson G. Out-of-pocket medical spending for care of chronic conditions. *Health Aff (Millwood)* 2001;20(6):267–278. [PubMed: 11816667]
3. Martin JA, Kung HC, Mathews TJ, et al. Annual summary of vital statistics: 2006. *Pediatrics* 2007;121(4):788–801. [PubMed: 18381544]
4. Mathews TJ, MacDorman MF. Infant mortality statistics from the 2003 period linked birth/infant death data set. *Natl Vital Stat Rep* 2006;54(16):1–29. [PubMed: 16711376]
5. Menotti A, Kromhout D, Blackburn H, Jacobs D, Lanti M. Forty-year mortality from cardiovascular diseases and all causes of death in the US Railroad cohort of the Seven Countries Study. *Eur J Epidemiol* 2004;19(5):417–424. [PubMed: 15233313]
6. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008;371(9608):261–269. [PubMed: 18207020]
7. Suthers K, Kim JK, Crimmins E. Life expectancy with cognitive impairment in the older population of the United States. *J Gerontol B Psychol Sci Soc Sci* 2003;58(3):S179–S186. [PubMed: 12730319]
8. Funk, SG.; Tornquist, EM.; Leeman, J.; Miles, MS.; Harrell, J., editors. Key aspects of preventing and managing chronic illness. Springer; New York: 2001.
9. Thorne S, Paterson B, Russell C. The structure of everyday self-care decision making in chronic illness. *Qual Health Res* 2003;13(10):1337–1352. [PubMed: 14658350]
10. Paterson BL. The shifting perspectives model of chronic illness. *J Nurs Scholarsh* 2001;33(1):21–26. [PubMed: 11253576]
11. Thorne SE, Paterson BL. Two decades of insider research: What we know and don't know about chronic illness experience. *Annu Rev Nurs Res* 2000;18:3–25. [PubMed: 10918930]
12. Baldewicz TT, Leserman J, Silva SG, et al. Changes in neuropsychological functioning with progression of HIV-1 infection: Results of an 8-year longitudinal investigation. *AIDS Behav* 2004;8(3):345–355. [PubMed: 15475681]
13. Teno JM, Weitzen S, Fennell ML, Mor V. Dying trajectory in the last year of life: Does cancer trajectory fit other diseases? *J Palliat Med* 2001;4(4):457–464. [PubMed: 11798477]
14. Beyene Y, Gilliss C, Lee K. “I take the good with the bad, and I moisturize”: Defying middle age in the new millennium. *Menopause* 2007;14(4):734–741. [PubMed: 17279059]
15. Holditch-Davis D, Docherty S, Miles MS, Burchinal M. Developmental outcomes of infants with bronchopulmonary dysplasia: Comparison with other medically fragile infants. *Res Nurs Health* 2001;24:181–193. [PubMed: 11526617]
16. Holditch-Davis D, Merrill P, Schwartz T, Scher M. Predictors of wheezing in prematurely born children. *J Obstetr Gynecol Neonat Nurs* 2008;37(3):262–273.
17. McConnell ES, Branch LG, Sloane RJ, Pieper CF. Natural history of change in physical function among long-stay nursing home residents. *Nurs Res* 2003;52(2):119–126. [PubMed: 12657987]
18. Black BP, Holditch-Davis D, Miles MS. Life Course Theory as a framework to examine becoming a mother of a medically fragile preterm infant. *Res Nurs Health* 2009;32(1):38–49. [PubMed: 18726937]
19. Davis LL, Buckwalter K. Family caregiving after nursing home admission. *J Mental Health Aging* 2001;7(3):361–379.
20. Hendrix CC, Ray C. Informal caregiver training on home care and cancer symptom management prior to hospital discharge: A feasibility study. *Oncol Nurs Forum* 2006;33(4):793–798. [PubMed: 16858461]
21. Bosworth HB, Dudley T, Olsen MK, et al. Racial differences in blood pressure control: Potential explanatory factors. *Am J Med* 2006;119(1):70.e9–e15. [PubMed: 16431192]
22. Trivedi RB, Ayotte B, Edelman D, Bosworth HB. The association of emotional well-being and marital status with treatment adherence among patients with hypertension. *J Behav Med* 2008;31(6):489–497. [PubMed: 18780175]

23. Bailey D, Albala D, Clipp E, et al. Psychosocial trajectories of men monitoring PSA levels following treatment for prostate cancer. *Oncol Nurs Forum* 2007;34(1):186.
24. Davis LL, Weaver M, Zamrini E, Stevens A, Kang DH, Parker CR Jr. Biopsychological markers of distress in informal caregivers. *Biol Res Nurs* 2004;6(2):90–99. [PubMed: 15388906]
25. Holditch-Davis D, Miles MS, Weaver MA, et al. Patterns of distress in African-American mothers of preterm infants. *J Behav Dev Pediatr* 2009;30(3):193–205.
26. Sandelowski M, Lambe C, Barroso J. Stigma in HIV-positive women. *J Nurs Schol* 2004;36(2):122–128.
27. Barroso J, Preisser JS, Leserman J, Gaynes BN, Golden RN, Evans DN. Predicting fatigue and depression in HIV-positive gay men. *Psychosomatics* 2002;43(4):317–325. [PubMed: 12189258]
28. Hudson AL, Taylor D, Lee KA, Gilliss CL. Symptom experience and self-care strategies among healthy, midlife African-American women. *J Natl Black Nurses Assoc* 2005;16(2):6–14. [PubMed: 16570641]
29. Keefe FJ, Lipkus I, Lefebvre JC, et al. The social context of gastrointestinal cancer pain: a preliminary study examining the relation of patient pain catastrophizing to patient perceptions of social support and caregiver stress and negative responses. *Pain* 2003;103(1–2):151–156. [PubMed: 12749969]
30. Lyles KW, Colón-Emeric CS, et al. for the HORIZON Recurrent Fracture Trial. Zoledronic acid in reducing clinical fracture and mortality after hip fracture. *N Engl J Med* 2007;35 nihpa40967.
31. Mosher CE, Lipkus IM, Sloane R, et al. Cancer survivors' health worries and associations with lifestyle practices. *J Health Psychol* 2008;13(8):1105–1112. [PubMed: 18987083]
32. Goetzel RZ, Ozminkowski RJ, Villagra VG, Duffy J. Return on investment in disease management: A review. *Health Care Financ Rev* 2005;26(4):1–19. [PubMed: 17288065]
33. Fitzpatrick, J.; Fulmer, T.; Wallace, M.; Flaherty, E. Geriatric nursing research digest. Springer; New York: 2000.
34. Miles, MS.; Holditch-Davis, D. Enhancing nursing research with children and families using a developmental science perspective. In: Fitzpatrick, JJ.; Miles, MS.; Holditch-Davis, D., editors. *Annual Review of Nursing Research*. Vol. vol. 21. Springer; New York: 2003. p. 1-20.
35. Holditch-Davis, D.; Black, B. Care of preterm infants: Programs of research and their relationship to developmental science. In: Fitzpatrick, JJ.; Miles, MS.; Holditch-Davis, D., editors. *Annual Review of Nursing Research*. Vol. vol. 21. Springer; New York: 2003. p. 23-60.
36. Kelly, JR.; McGrath, JE. On time and method. Sage; Newbury Park, CA: 1988. p. 9-28.
37. Holditch-Davis D, Scher M, Schwartz T, Hudson-Barr D. Sleeping and waking state development in preterm infants. *Early Hum Dev* 2004;80(1):43–64. 2004. [PubMed: 15363838]
38. Miles MS, Holditch-Davis D, Scher M, Schwartz T. A longitudinal study of depressive symptoms in mothers of prematurely born infants. *J Dev Behav Pediatr* 2007;28(1):36–44. [PubMed: 17353730]
39. Lentz MJ. Time series--issues in sampling. *West J Nurs Res* 1990;12(1):123–127. [PubMed: 2301167]
40. Holditch-Davis D, Bartlett TR, Belyea M. Developmental problems and the interactions between mothers and their three-year-old prematurely born children. *J Pediatr Nurs* 2000;15(3):157–167. [PubMed: 10897553]
41. Holditch-Davis D, Schwartz T, Black B, Scher M. Correlates of mother-premature infant interactions. *Res Nurs Health* 2007;30:333–346. [PubMed: 17514707]
42. Holditch-Davis D. The development of sleeping and waking states in high-risk preterm infants. *Infant Behav Dev* 1990;13:513–531.
43. Menard, S. Longitudinal research. 2nd ed.. Sage; Newbury Park, CA: 1991. p. 35-38, p. 42-48.
44. Taris, TW. A primer in longitudinal data analysis. Sage; London: 2000. p. 39-54.
45. Collins, LM. Measurement in longitudinal research. In: Collins, LM.; Horn, JL., editors. *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. American Psychological Assoc.; Washington, DC: 1991. p. 137-148.
46. Bayley, N. Manual for the Bayley Scales of Infant Development-Second Edition. Psychological Corp; San Antonio: 1993.
47. Bryk, T.; Raudenbush, S. Hierarchical linear models: Applications and data analysis methods. Sage; Newbury Park, CA: 1992.

48. Holditch-Davis D, Edwards L, Helms R. Modeling development of sleep-wake behaviors: I. Using the mixed general linear model. *Physiol Behav* 1998;63(3):311–318. [PubMed: 9469721]
49. Krueger C, Tian LA. Comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biol Res Nurs* 2004;6(2):151–157. [PubMed: 15388912]
50. Singer, JD.; Willet, JB. *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University; Oxford, UK: 2003.
51. Winer, B. *Statistical principles in experimental design*. 2nd ed.. McGraw-Hill; New York: 1971.
52. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: A generalized estimating equation approach. *Biometrics* 1988;44:1049–1060. [PubMed: 3233245]
53. Holditch-Davis D, Scher M, Schwartz T. The development of respiration in preterm infants. *J Perinat* 2004;24(10):631–639.
54. Artinian NT, Flack JM, Nordstrom CK, et al. Effects of nurse-managed telemonitoring on blood pressure at 12-month follow-up among urban African Americans. *Nurs Res* 2007;56(5):312–322. [PubMed: 17846552]
55. Morrison, DF. *Applied linear statistical methods*. Prentice-Hall, Inc; Englewood Cliffs: 1983.
56. Gueorguieva R, Krystal JH. Move over ANOVA: Progress in analyzing repeated measures data and its reflection in papers published in the Archives of General Psychiatry. *Arch Gen Psychiatr* 2004;61:310–317. [PubMed: 14993119]
57. Delucchi K, Bostrom A. Small sample longitudinal clinical trials with missing data: A comparison of analytic methods. *Psychol Methods* 1999;4(2):158–172.
58. Verbeke, G.; Molenberghs, G. *Linear mixed models for longitudinal data*. Springer; New York: 2000.
59. Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, SL. *Analysis of longitudinal data*, second edition. Oxford University; Oxford: 2002.
60. Diggle PJ, Kenward MG. Informative drop-out in longitudinal data analysis (with discussion). *Appl Stat* 1994;43:49–73.
61. Rubin DB. Multiple Imputation after 18+ years. *J Amer Stat Assoc* 1996;91:473–489.
62. Schafer JL, Olsen MK. Multiple Imputation for multivariate missing-data problems: A data analyst's perspective. *Multivar Behav Res* 1998;33:545–571.
63. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:968–974.
64. Verbyla AP. Conditioning in the growth curve model. *Biometrika* 1986;73:475–483.
65. Verbyla AP, Venables WN. An extension of the growth curve model. *Biometrika* 1988;75:129–138.
66. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 2002;7(2): 147–177. [PubMed: 12090408]
67. Little, RJ.; Rubin, DB. *Statistical analysis with missing data*. Wiley; New York: 1987.
68. Schafer, JL. *Analysis of incomplete multivariate data*. Chapman & Hall; London: 1997.
69. Rubin DB. Inference and missing data. *Biometrika* 1976;63:581–592.

Table 1

Correlations between the 3-Year Observations for the 50 Subjects and Correlated *t*-tests for Differences between the Observations.

Variable	Correlations	t(49)
Mother Talk	.64***	-0.25
Mother Gesture	.51***	1.13
Mother Touch	.49***	-1.41
Mother Look at Child	.58***	0.30
Mother Positive	.59***	-0.59
Mother Negative	.45**	-1.26
Mother Play with Child	.63***	-0.02
Child Talk	.63***	-1.38
Child Gesture	.38**	-0.80
Child Touch	.12	-2.48*
Child Look at Mother	.60***	-2.31*
Child Positive	.41**	-0.85
Child Negative	.51***	-1.65
Child Play	.50***	2.22*

*
p < .05.

**
p < .01.

p < .001.

Table 2

Bayley Scores of Premature Infants at 9 and 27 Months Corrected Age

Variables	9 Months	27 Months	Correlations Over Ages
	N Mean (SD)	N Mean (SD)	
Bayley MDI	96 97.6 (10.3)	77 87.7 (17.3)	$r(61) = .25^*$
Bayley PDI	96 91.2 (14.0)	73 83.6 (17.3)	$r(61) = .39^{**}$
Correlations of MDI & PDI at the Same Age	$r(94) = .41^{***}$	$r(71) = .39^{***}$	

*
p < .05.**
p < .01.***
p < .001

Table 3

Perceptions of 64 Mothers of the Helpfulness of Intervention at Infant Discharge and 2 Months Corrected Age

Variables	Satisfaction	Nurse Helpfulness	
	2 M	Discharge	2 M
Satisfaction with Study at Discharge	-.15	.44***	-.21
Satisfaction with Study at 2 M		-.04	.57***
Nurse Helpfulness at Discharge			-.02
Nurse Helpfulness at 2 M			

p < .001

Table 4

Correlations and Variances of 189 African American Preterm Infant Weights (in Kilograms) Across Six Ages Corrected for Prematurity*

	Infant Age in Months					
	Birth	2 Months	6 Months	12 Months	18 Months	24 Months
Birth	0.154 (189)					
2 Months	.294 (57)	1.342 (57)				
6 Months	.319 (119)	.703 (41)	1.943 (119)			
12 Months	.331 (141)	.573 (49)	.762 (103)	2.021 (141)		
18 Months	.436 (106)	.457 (43)	.727 (85)	.809 (94)	4.065 (107)	
24 Months	.246 (128)	.395 (50)	.684 (99)	.834 (119)	.883 (89)	2.683 (128)

* Variances are shown on the diagonal and are bolded, correlations on the off diagonal, and sample sizes in parentheses.