



Published in final edited form as:

Curr Neurol Neurosci Rep. 2017 February ; 17(2): 14. doi:10.1007/s11910-017-0723-4.

Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington's Disease as a Model

Tanya P. Garcia¹ and Karen Marder²

¹Department of Epidemiology and Biostatistics, Texas A&M University, TAMU 1266, College Station, TX 77843-1266, USA

²Department of Neurology, Columbia University, 630 West 168th Street, New York, NY 10032, USA

Abstract

Understanding the overall progression of neurodegenerative diseases is critical to the timing of therapeutic interventions and design of effective clinical trials. Disease progression can be assessed with longitudinal study designs in which outcomes are measured repeatedly over time and are assessed with respect to risk factors, either measured repeatedly or at baseline. Longitudinal data allows researchers to assess temporal disease aspects, but the analysis is complicated by complex correlation structures, irregularly spaced visits, missing data, and mixtures of time-varying and static covariate effects. We review modern statistical methods designed for these challenges. Among all methods, the mixed effect model most flexibly accommodates the challenges and is preferred by the FDA for observational and clinical studies. Examples from Huntington's disease studies are used for clarification, but the methods apply to neurodegenerative diseases in general, particularly as the identification of prodromal forms of neurodegenerative disease through sensitive biomarkers is increasing.

Keywords

Generalized estimating equations; Longitudinal study; Missing data; Mixed effect models; Time-varying effects

Introduction

Understanding the overall progression of neurodegenerative diseases is critical to the timing of therapeutic interventions and design of effective clinical trials. Disease progression can be assessed via longitudinal studies that measure outcomes repeatedly over time in relation to risk factors. In Huntington's disease (HD), for example, longitudinal studies have assessed

Correspondence to: Tanya P. Garcia.

This article is part of the Topical Collection on *Dementia*

Compliance with Ethical Standards

Conflict of Interest Tanya P. Garcia declares that she has no conflict of interest. Karen Marder reports grants from the Huntington's Disease Society of America, CHDI, TEVA, 1UL1 RR024156-01, and non-financial support from Raptor Pharmaceutical.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

the effect of medication use on performance of motor, cognitive, and neuropsychiatric function over time [1]. Longitudinal assessments of motor and cognitive impairments have also revealed insights into the natural progression of HD [2, 3].

Longitudinal data allow researchers to assess multiple disease aspects: changes of outcome(s) over time in relation to associated risk factors, timing of disease onset, and individual and group patterns over time. Assessing longitudinal temporal changes is central to learning specific time patterns of clinical impairments that could be missed otherwise [4]. Moreover, compared to cross-sectional studies, longitudinal studies often have less variability and increased statistical power [5].

Analyzing longitudinal data is complicated, however, by practical and theoretical issues. These include data that are missing, correlated, and collected at irregularly spaced visits. Modern statistical methods handle these complications, but hindrances are knowing when to use these methods, verifying their assumptions, and interpreting their output correctly. Confusion on these points can lead to inappropriate and inaccurate analysis.

In this paper, we review statistical techniques for analyzing longitudinal data for neurodegenerative diseases. Among all methods discussed, the mixed effect regression model (“Mixed effects regression (MER)” Section) is most flexible and designed to handle multiples challenges of longitudinal data. As such, it is recommended by the FDA in analysis of observational studies and clinical trials. Throughout, methods are described using examples from HD, a progressive, primarily single-gene disorder with complete penetrance that can be genetically diagnosed years before clinical symptoms onset. Compared to Alzheimer’s and Parkinson’s diseases, HD is less complicated, in that the genetic cause of HD absolutely predicts whether or not the person will develop HD and the CAG repeat length is correlated with age at onset. While the examples are in the context of HD, methods presented are relevant to neurodegenerative diseases in general, particularly as the identification of prodromal forms of neurodegenerative disease through sensitive biomarkers is increasing.

To clarify key points throughout, we use fictional, simple examples described in the text below and refer to them as cases 1, 2, and 3. A comparison of all methods mentioned is in Table 1.

Study of Total Motor Scores. One way of assessing HD progression is through clinical evaluations using the Unified Huntington’s Disease Rating Scale (UHDRS [6]). The UHDRS includes components that rate motor, cognitive, functional, and behavioral performance. In all cases, our outcome of interest is the total motor score (TMS), a component of the UHDRS that assesses the subject’s overall motor performance from 0 (no impairment) to 124 (high impairment).

Case 1: Single site study of TMS values collected at two time points. Suppose study participants from a single site are divided into three disease categories: “low,” “medium,” and “high” corresponding to the likeliness of being diagnosed with HD based on motor signs in the next 5 years. Inclusion into a specific disease category is based on percentile cut-offs of the calculated CAG-Age Product (CAP) formula [7]: age at baseline \times (CAG repeats—33.66). In general, the upper end of the “low” disease category is the 25th–40th percentile, and the lower end of the “high” disease category is the 60th–75th percentile. Exact cutoffs are based on an algorithm [7] applied to study data. For each participant, we collect TMS values at the beginning and end of the study.

Case 2: Single site study of TMS values collected at multiple time points. Similar to case 1, except now we collect TMS values on each participant annually over 10 years.

Case 3: Multiple site study of TMS values collected at multiple time points. Similar to case 2, except now participants come from multiple sites.

In hierarchical modeling terms, cases 1 and 2 are two-level models: level 1 represents the repeated TMS values over time for each participant (“within-subject”) model and level 2 represents the TMS values between participants (“between-subject” model). Case 3 is a three-level model which extends the two-level model with level 3 representing the TMS values among sites.

Challenges of Longitudinal Neurodegenerative Disease Studies

Correlated Data

Measurements in longitudinal studies are correlated by design. Correlation exists between repeated measures on the same individual (e.g., cases 1, 2) or from clustering of individuals across sites (e.g., case 3). Correlation within a site exists because subjects from the same site may have similar responses due to the site investigator, study protocol variations or equipment (e.g., MRI scanners). Attempts are always made to standardize assessments through training and use of phantoms in the case of scanners (i.e., a specially designed object that helps to evaluate and tune a scanner for reliability purposes). Ignoring the different sources of correlation in longitudinal studies has severe consequences: higher false positive rates and invalid confidence intervals from underestimated standard errors [8].

Another concern of longitudinal studies, particularly with multi-site studies, is handling vastly different numbers of participants across sites. Unequal sample sizes between sites have three key consequences. First, it risks violating the constant variance assumption of ANOVA-based methods (“Starter methods for longitudinal data analysis” section) that is not an issue for more advanced modeling approaches (“Modern methods for longitudinal data” section). Second, power is affected and is determined by the site with the smallest sample size. Third, even in more advanced modeling approaches, some effects may not be detectable. For example, if one site has a large number of “high” disease category participants and another site has a very small number of “low” disease category participants, then the effect of disease category may not be easily identified.

Irregularly Timed Data

Longitudinal studies generally encourage regularly occurring visits for data collection. But study participation frequency, and total study visits vary due to scheduling limitations and dropout. In our TMS example, individuals in the “high” or “medium” disease category may have limited mobility because they have more severe disease as the study progresses and may miss scheduled visits.

Missing Data

Missing data is the most problematic issue as there is no universally accepted correction, and inappropriate ones can have negative consequences.

Impact of Missing Data—Missing data can decrease the study’s statistical power and increase bias. Statistical power improves whenever the study’s sample size increases or variability of the study’s outcome measure (e.g., total motor score) is accurate.

Unfortunately, missing data negatively impacts the sample size and variability. First, analyses that exclude participants with missing values inadvertently reduce the study's sample size, potentially reducing the statistical power. Second, when participants who would have had extreme data values drop out (e.g., participants with very high or very low TMS), the variability of the study's outcome measure is incorrectly underestimated.

Missingness Mechanisms—Proper analysis of missing data requires understanding the “missingness mechanism” which describes why missing data occur [9]. Three mechanisms exist: missing completely at random, missing at random, and missing not at random. We provide examples of each mechanism using cases 1, 2, and 3 where the outcome variable is the total motor score (TMS). Table 1 provides a summary of how the methods discussed in this paper behave under each missingness mechanism.

Missing Completely at Random (MCAR): MCAR is when the missingness of the outcome variable is completely unsystematic. For example, consider case 1 where TMS is measured on two occasions. Suppose budget cuts force the investigator to reduce the number of subjects assessed at the second evaluation. If the investigator randomly samples among those participants initially evaluated, then missingness at the second time point is MCAR. This is because the subsample is random and not related to any other variable in the study.

Verifying MCAR can be achieved with Little's test [9] which examines group characteristics (e.g., mean) of participants with and without missing data. If characteristics are not equal for both groups, MCAR does not hold.

Missing at Random (MAR): MAR is when the probability that an outcome is missing is related to some other fully observed variable in the model, but not the variable with the missing value itself. In case 1, for example, suppose family history information is additionally collected on all participants at the first visit. If participants with at least two family members who have HD are less likely to return for the second evaluation, then the missingness is MAR. This is because the likeliness of missing data depends on the observed family history information.

Testing between MAR versus MCAR can be achieved with the SPSS missing data module. The general rule, however, is to assume the missingness mechanism is MAR unless there are strong reasons to assume MCAR.

Missing Not at Random (MNAR): MNAR is when the missingness depends on the missing values themselves. In Case 1, for example, suppose TMS values are fully observed at the first evaluation, but some are missing at the second, and that no family history information was collected at the first visit. If the missing values are from participants who have at least two family members with HD, then this is MNAR because the missingness depends on the unobserved family history information. To better understand this, note the distinction between our examples from MAR and MNAR. For MAR, the missingness at the second evaluation depends on observed family history information, whereas missingness in the MNAR example depends on unobserved family history information.

It is impossible to distinguish MNAR from MAR because doing so involves comparisons with unobserved missing data. When missingness is suspected to be MNAR, it is important to consult with a statistician to develop an appropriate model that accounts for this missingness mechanism. The usual approach is a joint model where the missingness model is varied and tested using sensitivity analyses [8].

Non-recommended Practices for Missing Data—Several simple remedies have been proposed for missing data, but are not generally recommended.

Complete-Case Analysis: This analyzes data only from those participants whose data is observed throughout the entire study. When missingness is MCAR, complete-case analysis yields unbiased parameter estimates. Otherwise, it yields biased, less precise estimates.

Last Observation Carried Forward (LOCF): LOCF replaces a participant's missing values with the last observed one. Assuming a participant will maintain that last observed value is unrealistic in most neurodegenerative disease studies.

Imputation Methods: Simple mean imputation replaces missing observations with the mean for that variable. Conditional mean imputation (or regression) replaces missing observations with predictions from regressing the outcome on other completely observed variables.

Despite their simplicity, both methods impute missing values only once and thus disregard the uncertainty of the imputed values. Such single imputation biases standard errors downward, leading to artificially narrow confidence intervals that give a false view of the estimate's precision.

A remedy is multiple imputation where multiple copies of the original data are generated and missing values are replaced from an appropriate stochastic model. The copies are analyzed as complete data sets and parameter estimates from each set are combined to produce a single estimate. Standard errors take into account the uncertainty of the imputation process. Despite their advantages over single imputation, multiple imputation is still not recommended by the FDA.

Recommended Practices for Missing Data—The FDA recommends approaches that account for the missing data mechanism such as generalized estimating equations (“Generalized estimating equations (GEE)” section) and mixed random effect models (“Mixed effects regression (MER)” section), with the latter being most preferred because of its ability to handle a more general missingness mechanism (MAR compared to MCAR for generalized estimating equations).

Starter Methods for Longitudinal Data Analysis

Change Score Analysis

When there are only two time points in the study (e.g., case 1), a straightforward approach is analyzing the change score: the differences between the measures at each time point. For

case 1, a change score is the changes in TMS measured at the start and end of the study. To compare change scores between the “low,” “medium,” and “high” disease categories, one could use a one-way analysis of variance (ANOVA). A one-way ANOVA is valid here because we are analyzing change scores, not repeated measures individually (hence the correlation problem is removed).

Analyzing change scores has been widely used in neurodegenerative disease research. For HD, Sturrock and colleagues [10] used the approach to evaluate longitudinal in vivo brain metabolite profiles in HD over a 24-month period. Poudel and colleagues [11] assessed longitudinal changes in white matter microstructure in HD over an 18-month period.

ANOVA Approaches

ANOVA approaches for longitudinal data include a repeated measures ANOVA and multivariate ANOVA (MANOVA). Both focus on comparing group means (e.g., the TMS scores between “low,” “medium,” and “high” disease categories), but neither informs about subject-specific trends over time.

ANOVA approaches are limited in handling irregularly timed and missing data. Repeated measures ANOVA requires all participants be measured at the same number of time points, and MANOVA requires fully complete data. Applying ANOVA methods to data with missing observations yields biased parameter estimates [12].

Repeated Measures ANOVA—Repeated measures ANOVA assesses group differences over time. Group sizes may be different, but subjects must be measured at the same number of time points. A repeated measures ANOVA is appropriate for case 2, and we describe the model in terms of this example.

The approach uses two main factors (time and disease category in case 2) and an interaction term (time \times disease category) to assess group differences over time. For case 2, the time main effect tests if TMS significantly changes over time averaged across disease categories. The disease category main effect tests whether, on average, one disease group has higher TMS than another. The interaction term, when statistically significant, indicates that the effect of time varies between disease categories. This variation can be observed by plotting the sample means of TMS over time by disease category: one may observe if TMS for one disease category increases (or decreases) over time compared to another.

A downside of repeated measures ANOVA is it assumes the measured outcomes have equal variances and covariances over time. This may be unrealistic since variances tend to increase with time and covariances decrease with increasing intervals in time. The MANOVA model, in comparison, makes more flexible variance-covariance assumptions as discussed next.

MANOVA—MANOVA models treat repeated observations as a vector (i.e., observations are multivariate). For example, in case 2, for each person in each disease category, the multivariate observations are 10-dimensional vectors of the TMS scores measured over 10 years.

MANOVA makes no assumptions about the variance-covariance structure of the repeated measures, and thus removes misspecification concerns. Despite this flexibility, MANOVA requires complete data. Subjects with incomplete data are either removed from the analysis or have missing values imputed, both of which are disadvantageous (“Non-recommended practices for missing data” section). Furthermore, MANOVA models do not allow time-varying predictors which is critical to modeling disease dynamics.

The limitations of ANOVA approaches lead toward the use of modern approaches that robustly handle challenges of longitudinal studies as discussed next.

Modern Methods for Longitudinal Data

Two preferred methods for longitudinal data are generalized estimating equations model (GEE) [13] and mixed effects regression (MER) [14]. Both allow time-invariant predictors that never change (e.g., gender, genotype) and time-varying predictors (e.g., age), and handle irregularly timed and missing data without the need for explicit imputation.

Generalized Estimating Equations (GEE)

Overview—A GEE model is designed for analyzing the regression relationship between covariates and repeated responses, but not the correlation structure of the repeated responses. If the latter is of interest, a GEE is inappropriate and one should consider MER (“Mixed effects regression (MER)” section). In estimating the regression parameters, the correlation structure in a GEE is represented using a working, potentially incorrect model (see “Modeling correlation” section). Even when the working model is incorrect, however, the GEE approach yields unbiased parameter estimates. Traditionally, GEEs are intended for two-level hierarchical data (e.g., cases 1 and 2), but recent work [15] has allowed extensions to three levels (e.g., case 3).

GEEs have been widely used in the neurodegenerative disease literature. For HD, Maroof and colleagues [16] used GEEs to model trajectories of cognitive scores (repeated response) in relation to time, education and baseline age. Keogh and colleagues [1] used GEEs to separately assess longitudinal performance of motor, cognitive and neuropsychiatric functions (repeated response) in relation to medication use.

Advantages and Limitations—Two primary advantages of GEEs are its robustness to misspecification of the repeated measures’ correlation structure and its computational simplicity. Estimation in GEEs uses a working correlation structure that may be inconsistent with the observed correlations of the repeated measures. Regardless, the regression parameter estimates are consistent, but associated standard errors are incorrect when the working correlation structure is wrong. Standard errors of time-dependent covariates are generally overestimated and time-independent covariates are underestimated. See [17] for recommended corrections to standard error estimates.

The ability to yield valid estimates even when the correlation structure is not correctly modeled is a similar benefit to that for MANOVA models (“MANOVA” section), but GEEs are more advantageous in that they do not disregard participants with incomplete data.

Finally, estimation in GEEs is carried out with quasi-likelihood methods which is computationally easier than full-likelihood methods (as done for MER).

Limitations of GEEs are threefold. First, GEEs assume missing data are MCAR which may not hold for neurodegenerative disease studies. Extensions to the more flexible MAR assumption have been proposed including a weighted-estimating equations approach [8]. Second, one cannot perform hypothesis testing on correlation parameters since these are not directly estimated. Third, usual methods (e.g., likelihood ratio tests, Akaike/Bayesian Information Criteria) cannot be used to test and compare model fits because the focus is solely on regression parameters, not all model parameters (i.e., regression and correlation parameters). All of these limitations are handled by MER models.

Modeling Correlation—Estimation of regression parameters in a GEE is carried out under a working correlation structure for the repeated measures, meaning that a (possibly incorrect) model is chosen to represent the correlation observed between repeated measures. The working structure is selected at the beginning of the analysis, and we recommended that it resemble the observed correlations for better estimation of standard errors. However, even if the working structure is incorrect, regression parameter estimates remain consistent. We describe next four common working structures and provide guidance on each.

The independent correlation structure assumes there is no correlation between repeated measures. This is a simple, yet unrealistic choice for longitudinal data, and one that results in large efficiency loss for time-varying covariates [18]. It is a fair choice for initial analyses to quickly assess the regression relationship between covariates and repeated responses.

The exchangeable correlation assumes correlations within a cluster are equal. In our example, consider TMS at baseline for all participants clustered by disease category (or by sites). An exchangeable correlation structure assumes that the correlation between TMS values of any two participants within a disease category (or within a site) is the same regardless of which participants are chosen. That is, participants are *exchangeable* within a disease category (or within a site). The correlation between participants from different disease categories (or different sites) is zero.

An example where exchangeable correlation is unreasonable is case 2 where clusters are the participant's TMS values over 10 years. Assuming exchangeable correlation means that the correlation between TMS values in years 1 and 2 is the same as the correlation of TMS values between years 1 and 10. This is unrealistic since TMS values closer in time (years 1 and 2) are more likely to have higher correlation than those farther apart (years 1 and 10). In practice, an exchangeable correlation is reasonable when “objects” in a cluster can be moved without impact; e.g., participants in the same disease category or site, but not measures over time. An autoregressive correlation is more appropriate for case 2 as described next.

The autoregressive correlation accounts for time-varying correlation by assuming that measurements taken closer in time are more highly correlated than measurements taken farther apart. In practice, this structure is identified using an autocorrelation plot which

displays the correlation by time-lag (e.g., ACFPLOT in SAS). A steadily decreasing plot is indicative of autoregressive correlation.

An unstructured correlation makes no assumptions about the correlation form and uses different parameters for each correlation component (i.e., for n time points, there are $n(n-1)/2$ components). Though flexible, this model is computationally costly. In case 2, with 10 time points, there are $10(10-1)/2=45$ separate correlations to be estimated. The large number of computations decreases accuracy of parameter estimates and may even lead to model fitting failure. In practice, an unstructured correlation is recommended when there are few time points.

Mixed Effects Regression (MER)

Overview—MER models provide information regarding the regression relationship between covariates and repeated responses, and about the correlation structure of the repeated response. It captures correlations of repeated measures using “random effects” that serve to describe cluster-specific trends over time. In case 2 where clusters are individuals, random effects can serve to describe each participant’s trend over time, and in case 3, an additional random effect can serve to differentiate sites. Random effects allow estimation of cluster-specific effects useful for understanding interindividual variability in longitudinal responses and cluster-specific predictions.

MERs have been widely used in neurodegenerative disease studies. For HD, Tabrizi and colleagues [19] used linear MERs to assess the longitudinal changes of different outcomes: clinical, cognitive, quantitative motor, neuropsychiatric assessments and MRI measures of the brain over a 36-month period. Each outcome was separately modeled using MERs and clusters corresponded to each person’s annual measures over the 36-month period. Long and colleagues [20] used linear MERs to estimate the timing of motor impairments and Collins and colleagues used it to assess finger tapping as a longitudinal marker of HD progression.

Advantages and Limitations—A MER model is advantageous over GEEs in that (i) it allows multi-level hierarchical models that allow predictions for each data hierarchy level. (ii) One may perform hypothesis testing on correlation parameters since they are directly estimated. (iii) Usual methods (e.g., likelihood ratio tests, Akaike/Bayesian Information Criteria) can be used to test and compare model fits because all model parameters (i.e., regression and correlation parameters) are estimated. (iv) It is more robust to missing data and assumes missingness is MAR which is more general than the MCAR assumption of GEEs.

A primary limitation of MER models is their computational complexity over GEEs particularly with nonlinear MER as it involves time-consuming numerical integration over the random effects. A second limitation is the reliance on correct specification of the mean and correlation structure of the repeated responses for valid hypothesis testing conclusions. We discuss next the impact of misspecification.

Modeling Correlation—Correlation in MERs is captured through random effects and their associated distributions. In theory, correctly estimating model parameters requires

accurately specifying the random effect distribution (the standard assumption is a normal distribution) [21]. But in practice, an incorrect distribution may not have severe consequences.

When the random effects distribution is specified incorrectly, but the covariates and random effects are independent, then parameter estimates and associated standard errors are valid [22•]. Otherwise, when the covariates and random effects are dependent, then bias is incurred [23, 24].

Covariates and random effects depend on each other when, for example, the variability of the random effect depends on the patient's disease category (case 1) or site location (case 2). Testing for this dependence can be done using the Hausman chi-squared test [24]. If there is no evidence of dependence, then we recommend applying the MER assuming random effects are normally distributed. Otherwise, the investigator should consult with a statistician and use a procedure that makes no modeling assumptions about the random effect distribution [22•].

Time-Varying Predictors

GEEs and MERs can model time-varying predictors useful for understanding disease progression. For example, changing medication usage (yes/no response) or changing medication dosage, or changing blood pressure and weight.

Time-varying predictors are typically modeled using linear combinations of splines which are flexible curves that connect two or more points [25]. Spline modeling involves two decisions: (i) the choice of the spline functions and (ii) the number of splines used. These decisions impact how precisely and smoothly (wiggleness) the time-varying effects are captured. Fortunately, these decisions have been well-studied, and the recommended approach is using P-spline functions with the number of splines automatically selected from a criterion that maximizes accuracy and minimizes wiggleness [26]. This approach is available in R (*mgcv*) and SAS (*PROC GAM*).

Conclusions

We discussed challenges of longitudinal data from neurodegenerative disease studies (data that are correlated, irregularly timed and/or missing) and major techniques that handle them (GEEs and MERs). Simpler ANOVA-based approaches cannot handle irregularly timed and missing data. It resorts to modeling complete-cases or imputing missing values, and the focus rests on comparing group means rather than subject-specific trends over time.

GEEs and MERs overcome these challenges, the former providing estimates that are population averaged and the latter providing subject-specific estimates. These two estimates agree only for continuous normal outcomes with the identity link. When the missing data are MCAR, GEE and MER models produce unbiased parameter estimates. But when the missing data are MAR, GEE does not perform well, but MER models do as long as mean and variance-covariance structure are modeled properly. The greater flexibility of MERs

lends preference to using them over GEEs for longitudinal data, and is recommended by the FDA for observational studies and clinical trials.

MERs have become a standard in studies of HD for properly modeling correlated longitudinal data. It has been frequently used in analyses of prospective, observational, multi-center longitudinal studies such as COHORT [27], PHAROS [3], PREDICT [2•], and TRACK-HD [19]. Dorsey and colleagues [27] used a MER model to reveal a monotonic decline of movement, cognition, behavior and function using data from COHORT consisting of measures from participants and controls who had at least 3 consecutive years of longitudinal data. For PHAROS [3], Biglan and colleagues used a MER model, adjusted for age and sex, to differentiate linear trends of motor, cognitive, psychiatric, and functional decline between individuals with and without the HD mutation. Paulsen and colleagues [2•] used a MER model on PREDICT data to reveal that imaging variables based on regional brain volumes had the largest effect sizes in detecting differences between premanifest HD participants and controls. Tabrizi and colleagues [19] also used a MER model to compare phenotypic differences between controls, premanifest HD, and early HD participants.

Each analysis encountered different challenges, particularly in dealing with missingness and timing of data collection. COHORT, PHAROS, and PREDICT were at least 6 year studies, whereas TRACK-HD was only a 3-year study. A challenge, thus, in analyzing TRACK-HD was dealing with weak statistical power because of few HD converters. The analysis of COHORT also had issues of missing data. Follow-up in COHORT was intermittent, and of the 1514 participants, only 366 had at least 3 consecutive years of longitudinal data, meaning that the analysis was a type of complete-case analysis which may be improved upon using data from all 1514 participants. The analyses of PHAROS and PREDICT had fewer issues with missing data, having used all longitudinal data collected, and having dropout rates less than 5%. Follow-up times in data collection also differed: COHORT, PREDICT, and TRACK-HD had 1-year follow-ups and PHAROS had 9-month follow-ups. Despite the regularity of these observations, more frequent observations could help to minimize missing data and more accurately detect rates of decline. A modern push towards more frequent data collection is the use of sensors, microelectronics, and telecommunications that now provide inexpensive, wearable systems to track HD impairments more frequently, even at the convenience of a patient's home setting [28]. Techniques discussed in this paper can serve as a starting point for analyzing the more frequently collected sensor data, but more advanced techniques [29] are recommended as the volume of data increases and validation against clinically collected UHDRS data should be considered for validation.

Acknowledgments

The authors would like to give a special thank you to Dr. Susan Fox for taking the time to review this manuscript.

Funding This work is supported in part by the National Institute Of Neurological Disorders And Stroke of the National Institutes of Health under Award Number K01NS099343, the Huntington's Disease Society of America Human Biology Project Fellowship, Texas A&M School of Public Health Research Enhancement and Development Initiative (REDI-23-202059-36000), and National Center for Advancing Translational Sciences (2UL1RR024156-06).

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. Keogh R, Frost C, Owen G, Daniel RM, Langbehn DR, Leavitt B, et al. Medication use in early-HD participants in TRACK-HD: An investigation of its effects on clinical performance. *PLoS Curr*. 2016; 8:1.
2. Paulsen JS, Long JD, Johnson HJ, Aylward EH, Ross C, Williams JK, et al. Clinical and biomarker changes in premanifest Huntington disease show trial feasibility: a decade of the PREDICT-HD study. *Front Aging Neurosci*. 2014; 6:78. Provides examples of longitudinal study issues and analyses of current neurodegenerative disease study. [PubMed: 24795630]
3. Biglan KM, Shoulson I, Kieburtz K, Oakes D, Kayson E, Shinaman MA, et al. Clinical-genetic associations in the prospective huntington at risk observational study (PHAROS). *JAMA Neurol*. 2015; 14620:1.
4. Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. *Psychol Methods*. 2012; 17(1):61–77. [PubMed: 22103434]
5. Zeger SL, Liang KY. An overview of methods for the analysis of longitudinal data. *Stat Med*. 1992; 11(14–15):1825–39. [PubMed: 1480876]
6. H. S. Group. Unified Huntington's disease rating scale: reliability and consistency. *Mov Disord*. 1996; 11(2):136–42. [PubMed: 8684382]
7. Zhang Y, Long JD, Mills JA, Warner JH, Lu W, Paulsen JS. Indexing disease progression at study entry with individuals at-risk for Huntington disease. *Am J Med Genet B Neuropsychiatr Genet*. 2011; 156B(7):751–63. [PubMed: 21858921]
8. Gibbons RD, Hedeker D, DuToit S. Advances in analysis of longitudinal data. *Annu Rev Clin Psychol*. 2010; 6:79–107. [PubMed: 20192796]
9. Little, RJA., Rubin, DB. *Statistical analysis with missing data*. 2. New York: John Wiley and Sons; 2002.
10. Sturrock A, Laule C, Wyper K, Milner RA, Decolongon J, Santos RD, et al. A longitudinal study of magnetic resonance spectroscopy Huntington's disease biomarkers. *Mov Disord*. 2015; 30(3): 393–401. [PubMed: 25690257]
11. Poudel GR, Stout JC, Domínguez DJF, Churchyard A, Chua P, Egan GF, et al. Longitudinal change in white matter microstructure in Huntington's disease: the IMAGE-HD study. *Neurobiol Dis*. 2015; 74:406–12. [PubMed: 25497085]
12. Shaw RG, Mitchell-Olds T. ANOVA for unbalanced data: an overview. *Ecology*. 1993; 74(6): 1638–45.
13. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73(1):13–22.
14. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38(4):963–74. [PubMed: 7168798]
15. Teerenstra S, Lu B, Preisser JS, Van Achterberg T, Borm GF. Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics*. 2010; 66(4):1230–7. [PubMed: 20070297]
16. Maroof DA, Gross AL, Brandt J. Modeling longitudinal change in motor and cognitive processing speed in presymptomatic Huntington's disease. *J Clin Exp Neuropsychol*. 2011; 33(8):901–9. [PubMed: 21644140]
17. Agresti, A. *Categorical data analysis*. 3. New York: John Wiley and Sons; 2013.
18. Fitzmaurice, G., Molenberghs, G. Advances in longitudinal data analysis: a historical perspective. In: Geert, V.Davidian, MarieGarrett, F., Geert, M., editors. *Longitudinal data analysis*. Boca Raton: Chapman and Hall/CRC; 2008. p. 3-27.
19. Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-

- HD study: analysis of 36-month observational data. *Lancet Neurol.* 2013; 12(7):637–49. [PubMed: 23664844]
20. Long JD, Paulsen JS, Marder K, Zhang Y, Kim JI, Mills JA, et al. Tracking motor impairments in the progression of Huntington’s disease. *Mov Disord.* 2014; 29(3):311–9. [PubMed: 24150908]
 21. Litière S, Alonso A, Molenberghs G. Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics.* 2007; 63(4):1038–44. [PubMed: 17425642]
 - 22•. Garcia TP, Ma Y. Optimal Estimator for Logistic Model with Distribution-free Random Intercept. *Scand J Stat.* 2016; 43(1):156–71. Provides details for how to test if random effects and covariates are dependent or not for mixed effect regression.
 23. McCulloch CE, Neuhaus JM. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. *Stat Sci.* 2011; 26(3):388–402.
 24. Hausman JA. Specification tests in econometrics. *Econometrica.* 1978; 46(6):1251–71.
 25. de Boor, C. A practical guide to splines. New York: Springer-Verlag; 2001.
 26. Wood SN. Generalized additive models : an introduction with R. *Texts Stat Sci.* 2006; 392:xvii.
 27. Dorsey ER. Characterization of a large group of individuals with huntington disease and their relatives enrolled in the COHORT study. *PLoS One.* 2012; 7(2):e29522. [PubMed: 22359536]
 28. Andrzejewski KL, Dowling AV, Stamler D, Felong TJ, Harris DA, Wong C, et al. Wearable sensors in Huntington disease: a pilot study. *J Huntingtons Dis.* 2016; 5(2):199–206. [PubMed: 27341134]
 29. Sørensen H, Goldsmith J, Sangalli LM. An introduction with medical applications to functional data analysis. *Stat Med.* 2013; 32(30):5222–40. [PubMed: 24114808]

Summary of methods for analyzing longitudinal study data

Table 1

	Change score analysis	Repeated measures ANOVA	Multivariate ANOVA (MANOVA)	Generalized estimating equations (GEE)	Mixed effect regression (MER) ^d
Description	Analyzes differences between outcomes measured at two time points.	Uses two main factors, and an interaction term to assess group differences over time	Repeated responses over time treated as multivariate observations.	Designed for analyzing the regression relationship between covariates and repeated responses. GEEs do not allow inference on correlation structure of the repeated responses, but MERs do.	
Number of time points	Only 2.	Multiple.	Multiple.	Multiple.	Multiple.
Irregularly timed data	No.	No.	No.	Yes.	Yes.
Time-varying predictors	Not allowed.	Time treated as classification variable.	Time treated as classification variable.	Allowed.	Allowed.
How correlation between repeated responses modeled	Not applicable.	Assumes outcomes have equal variances and covariances over time.	No specific assumption.	Working models ^b that may or may not resemble observed correlations.	Random effects that quantify variation among units and serve to describe cluster-specific trends over time.
Missing data	Requires complete data. Analysis based on complete-cases or imputed missing values. ^c				Handles missing data without explicit imputation needed. GEEs assume missingness is MCAR. Mixed effects regression assume missingness is MAR.
	1 Method yields unbiased parameter estimates and standard errors for (a) complete-case analysis when missingness is MCAR, (b) multiple imputation when missingness is MCAR or MAR.				
	2 Method yields only unbiased parameter estimates for (a) single mean imputation when missingness is MCAR, (b) conditional mean imputation when missingness is MCAR or MAR.				
	3 Method yields biased estimates for (a) complete-case analysis when missingness is MAR, (b) last observation carried forward when missingness is MCAR or MAR, (c) single mean imputation when missingness is MAR.				
Computation	Group differences of change scores analyzed with one-way ANOVA.	ANOVA implementation in standard software (SAS, SPSS, R).	MANOVA implementation in standard software (SAS, SPSS, R).	Quasi-likelihood methods; PROC GENMOD in SAS.	Likelihood methods; PROC MIXED in SAS.

^aPreferred FDA method for incomplete longitudinal data

^bWorking models are typically one of four choices: independent, exchangeable, autocorrelation, unstructured (“Modeling correlation” section). Even when the working model is incorrect, regression parameter estimates are consistent, but associated standard errors are not. Agresti [17] has recommendations for correcting standard error estimates

^cMCAR missing completely at random, MAR missing at random