

Thinking Non-Linearly

Olivia Belbin*

Memory Unit, Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain

*Corresponding author: Olivia Belbin, Memory Unit, Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain, E-mail: OBelbin@santpau.cat

Received date: February 18, 2016; Accepted date: February 19, 2016; Published date: February 26, 2016

Copyright: © 2016 Belbin O. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited

Editorial

Neurodegeneration is a multifaceted process that can lead to complex phenotypes. Consequently, research into the cause and progression of neurodegenerative diseases requires sophisticated statistical models. While studies that use dichotomous traits (e.g. affected versus unaffected) can be valuable for identifying disease-causing / risk-modifying factors, they do not consider the contribution of factors that modify the progression or severity of the disease. For this reason, quantitative traits are more appealing especially when combined with repeated measures over time (longitudinal studies) as they not only inform about disease modification but can greatly increase statistical power. For example, > 80% power to detect differences in rate of change of the quantitative trait can be achieved with relatively small sample sizes (< 100 measurements, which can equate to < 50 subjects). In comparison, dichotomous studies can require > 5000 subjects in order to achieve 80% power to detect group difference unless the effect size is substantial (OR > 2). Perhaps more importantly, longitudinal studies also avoid issues surrounding the application of unsubstantiated / arbitrary cut-offs when defining diagnostic groups. The benefit of longitudinal quantitative traits is evident in that biomarker (cerebrospinal fluid, brain imaging and neuropsychological assessments) studies are increasingly demonstrating that changes in the biomarker over the disease course are more informative than comparing cross-sectional differences in mean values across diagnostic groups. Neuropsychological assessments (e.g. cognitive score tests) and brain imaging (e.g. brain atrophy or cerebral blood flow) represent great examples of non-invasive correlates for disease progression (e.g. Alzheimer's disease) and severity that can be measured quantitatively over the course of a neurodegenerative disease.

An important issue that comes into play when analysing “noisy” biological data is the inherent underlying heterogeneity. Firstly, baseline values of the variable in question often vary greatly between individuals as can the rate and direction of change. While tangible factors (fixed effects) that may contribute to these changes (e.g. inter-individual differences in age, body weight, sex or genetic background) can be included in statistical models, intangibles (random effects) such as differences in disease stage, degree of underlying pathology are oftentimes immeasurable but can nevertheless create substantial heterogeneity and therefore should also be considered. A good example of an intangible factor comes from longitudinal studies of brain atrophy where the equilibrium between cortical thinning due to atrophy versus cortical thickening due to pathology-induced inflammation can be a confounding factor when assessing atrophy rates but difficult to measure quantitatively. Statistical models that allow for intangibles (random effects) are therefore more appropriate for complex disease processes.

A common method for analysing longitudinal studies is linear mixed modelling (LMM). LMM predicts the progression of a biomarker over time taking into consideration both fixed and random effects. While this is a relatively sophisticated and somewhat standard approach, the underlying assumptions of this model should not be ignored. Namely that 1) the random components of the model are Gaussian, 2) that the quantitative trait is continuous (i.e., does not have a floor or ceiling threshold) and 3) that a unit change in any variable is associated with a constant fixed change in the trait. These latter two assumptions are particularly important when considering cognitive decline. Firstly, as with most biological measures, cognitive score tests have minimum (floor) and maximum (ceiling) scores. This means that nonsensical predictions that lie outside of the feasible range can be obtained when using LMM. Moreover, the rate of cognitive decline is not linear over the entire time-course of the study, particularly when close to the floor and ceiling thresholds. Conversely, non-linear mixed effects (NLME) models allow for both fixed and random effects and use a logistic (S-shaped) function in which the quantitative trait starts at the ceiling value followed by a linear decline phase and plateau at the floor value. As such, this approach has been validated as a more robust method for modelling biological data and is increasingly being employed in analyses of genetic quantitative trait loci, longitudinal studies of cognitive decline and pharmacokinetic studies. As an example of its utility, NLME can model both the extent (age at which patients reach 50% baseline score) and rate (age at which patients reach 50% baseline score) of cognitive decline between individuals or groups and can be implemented using freely available packages such as R and Mat lab.

A common limitation of longitudinal studies is the number of repeated measures available for each patient. The higher the density of data-points, the more accurate the statistical model but equally important is full coverage of the disease time-course from start point, through the linear phase of change through to the end point. The collection of such data is itself limited by the fact that inclusion criteria / patient referral for clinical studies often require that the underlying process has already taken hold. As a result, the relative lack of data available for early disease stages will result in limited power to detect modifying factors early in the disease. Another factor that should be considered is the resolution of the measure in question. For example, in the case of cognitive decline, a common test is the minimal state examination, which involves a 30-point questionnaire. In comparison, the CAMCOG is a 10⁷ point questionnaire. Differences in the suitability of these two tests for neuropsychological assessment aside, it is apparent that a 10⁷-point system will have greater resolution to detect subtle changes in cognition than a 30-point system when applied in NLME modelling.

In conclusion, the complexity that underlies neurodegenerative diseases dictates that the use of simplistic statistical models is inappropriate in most instances. As increasingly sophisticated and robust non-linear models become available, the limiting factor for such

studies may be access to quantitative and informative clinical data. Thus the importance of collecting a full battery of clinical, pathological data and genetic data over various time-points should not be underplayed when recruiting patient samples.