Mixed effects models but not t-tests or linear regression detect progression of apathy in Parkinson's disease over time: A comparative analysis.

Mixed effects models but not t-tests or linear regression detect progression of apathy in Parkinson's disease over time: A comparative analysis.

Anne-Marie Hanff[1,2,3,4]*(RN, MSN), Rejko Krüger(Dr)[1,2,5] Christopher McCrum[4](PhD), Christophe Ley[6](PhD), on behalf of NCER-PD.

1 Transversal Translational Medicine, Luxembourg Institute of Health, Strassen, Luxembourg

2 Translational Neurosciences, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

3 Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University Medical Centre+, Maastricht, The Netherlands

4 Department of Nutrition and Movement Sciences, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands

5 Parkinson Research Clinic, Centre Hospitalier du Luxembourg, Luxembourg, Luxembourg

6 Department of Mathematics, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Corresponding author

Anne-Marie Hanff
1A-B, rue Thomas Edison
L-1445 Strassen
Luxembourg
Mail: anne-marie.hanff @ lih.lu
Tel.: 00352 691 857 808

Highlights

- Mixed effects models were the only method able to detect an increase in apathy symptoms
- Paired t-tests and linear regression underestimated longitudinal change
- Researchers should more often consider mixed effects models for longitudinal analysis

Abstract

OBJECTIVE While there is an interest in defining longitudinal change in people with chronic illness like Parkinson’s disease (PD), statistical analysis of longitudinal data is not straightforward for clinical researchers. Here, we aim to demonstrate how the choice of statistical method may influence research outcomes, (e.g., progression in apathy), specifically the size of longitudinal effect estimates, in a cohort.

STUDY DESIGN AND SETTING In this retrospective longitudinal analysis of 802 people with typical Parkinson’s disease in the Luxembourg Parkinson's study, we compared the mean apathy score at visit 1 with the mean apathy score at visit 8 by means of the paired two-sided t-test. Additionally, we analysed the relationship between the visit numbers (all observations) and the apathy score (change in apathy per year) using linear regression and longitudinal two-level mixed effects models.

RESULTS Mixed effects models were the only method able to detect progression of apathy over time. While the effects estimated for the group comparison and the linear regression were smaller with high p-values (+1.016/ 7years, $p = 0.107$, -0.008/ year, $p = 0.897$, respectively), indicating an insignificant change in apathy over time, effect estimates for the mixed effects models were positive with a very small p-value, indicating a significant increase in apathy symptoms per year by +0.335 ($p < 0.001$). We provided evidence for, and theoretical explanations of, how mixed effects models can be used to assess symptoms progression more reliably, as well as the limitations of group comparison and linear regression in the analysis of longitudinal data.

CONCLUSION Mixed effects models can be used to estimate different types of longitudinal effects while the inappropriate use of paired t-tests and linear regression to analyse longitudinal data can lead to underpowered analyses and an underestimation of longitudinal change. Thus, researchers should rather consider mixed effects models for longitudinal

analyses. In case this is not possible, limitations of the analytical approach need to be discussed and taken into account in the interpretation of results of cohort studies.

Plain language summary

WHAT WE DID: We analysed data from a group of people with typical PD up to eight years to understand how a statistical method can affect outcomes.

WHAT WE FOUND: We used different statistical methods to assess if apathy (lack of motivation) in people with PD changed over time. The linear mixed effects models showed a significant increase in apathy each year whereas other methods did not find this increase.

WHY IT MATTERS: Using the most appropriate method is important in studying how symptoms change over time as some methods might underestimate change over time. If this is not possible, scientists should discuss the disadvantages of their methods.

Keywords Cohort Studies, Epidemiology, Disease Progression, Parkinson, Lost to Follow-Up, Statistical model

Running title Statistical assessment of Parkinson's progression

Word count 2007

1. Introduction

Longitudinal cohort studies gather data over time to track changes in health outcomes. When assessing the same individuals over time, the different data points are likely to be more similar to each other than measurements taken from other individuals, and it is important to take this intra-individual change into account when analysing the data. Parkinson's disease (PD) is a heterogeneous neurodegenerative disorder resulting in a wide variety of motor and non-motor symptoms including apathy, defined as a disorder of motivation, characterised by reduced goal-directed behaviour and cognitive activity and blunted affect [1]. Non-motor symptoms like apathy increase over time [2]. Consequently, using data from the Luxembourg Parkinson's study [3, 4], we assess change in apathy and demonstrate how the choice of statistical method may influence research outcomes, e.g., change in apathy, specifically the size and interpretation of longitudinal effect estimates in a cohort. Thus, the findings are intended for illustrative and educational purposes related to the statistical methodology. Using three different statistical approaches, the following questions will be tackled: are apathy scores at visit 8 in the cohort worse than at visit 1 (approach: a paired t-test) and do apathy scores become worse with the number of yearly visits (approaches: a linear regression model and a mixed effects model)? To address these questions, we formulated them under the form of statistical hypotheses as follows:

1. H0 : Mean apathy at visit 1 = Mean apathy at visit 8
   HA : Mean apathy at visit 1 != Mean apathy at visit 8
2. H0 : Regression coefficient for visit number = 0
   HA : Regression coefficient for visit number != 0
3. H0 : Fixed effects for visit number = 0
   HA : Fixed effects for visit number != 0

2. Material and methods

This is a retrospective analysis of data from the Luxembourg Parkinson's study, a nation-wide, monocentric, observational, longitudinal-prospective dynamic cohort [3, 4]. Among the participants are people with typical PD and PD dementia (PDD), living mostly at home in Luxembourg and the Greater Region (geographically close areas of the surrounding countries Belgium, France, and Germany). People with atypical PD were excluded. The sample at the date of dataexport (2023.06.22) consisted of 802 individuals, of which 269 (33.5%) were female. Apathy was measured by the discrete score from the Starkstein apathy scale (0 – 42,

higher = worse) [5], a scale recommended by the Movement Disorders Society [6]. Data used in the preparation of this manuscript were obtained from the National Centre of Excellence in Research on Parkinson's disease (NCER-PD). Ethical approval was provided by the National Ethics Board (CNER Ref: 201407/13). We used data from up to eight visits, which were performed annually between 2015 and 2023.

We conducted data analysis using R version 3.6.3 [7]. The paired two-sided t-test compared the mean apathy score at visit 1 with the mean apathy score at the visit 8 (corresponding to the first hypothesis testing setting). We attract the reader's attention to the fact that this implies a rather small sample size as it includes only those people with data from the first and 8^{th} visit. The linear regression analysed the relationship between the visit number and the apathy score (using the "stats" package [7]). More concretely, we analysed whether there is a linear relationship or not (see second hypothesis). To describe the longitudinal progression (third hypothesis), we performed longitudinal two-level mixed effects models analysis with a random intercept on subject level, a random slope for visit number and the visit number as fixed effect (using the "lmer"-function of the "lme4"-package [8]). The latter two approaches use all available data from all visits while the paired t-test does not when missing data occur. We illustrated the analyses in plots with the function "plot_model" of the R package sjPlot [9]. The R syntax for all analyses is provided on the OSF project page: https://osf.io/nf4yb/ .

As illustrated in the flow chart (Figure S1), the sample analysed from the paired t-test is highly selective: from the 802 participants at visit 1, the t-test only included 63 participants with data from visit 8. This arises from the fact that, first, we analyse the dataset from a dynamic cohort, i.e., the data at visit 1 were not collected at the same time point. Thus, 568 of the 802 participants joined the study less than eight years before, leading to only 234 participants eligible for the eighth yearly visit. Second, after excluding non-participants at visit 8 due to death (n = 41) and other reasons (n = 130), only 63 participants at visit 8 were left. To discuss the selective study population of a paired t-test, we compared the characteristics (age, education, age at diagnosis, apathy at visit 1) of the remaining 63 participants at visit 8 (included in the paired t-test) and the 127 non-participants at visit 8 (excluded from the paired t-test) [10].

3. Results and discussion

Hypothesis 1: Worse apathy scores at the eighth visit than at baseline

Panel A in Figure 1 illustrates the means and standard deviations of apathy for all participants at each visit, while the flow-chart (Figure S1) illustrates the number of participants at each stage. On average, we see lower apathy scores at visit 8 compared to visit 1 (higher score = worse). By definition, the paired t-test analyses pairs, and in this case, only participants with complete apathy scores at visit 1 and visit 8 are included, reducing the total analysed sample to 63 pairs of observations. Consequently, the t-test compares mean apathy scores in a subgroup of participants with data at both visits leading to different observations from Panel A, as illustrated and described in Panel B: the apathy score has increased at visit 8, hence symptoms of apathy have worsened. The outcome of the t-test along with the code is given in Table 1. Interestingly, the effect estimates for the increase in apathy were not statistically significant (+1.016 points, 95%CI: -0.225, 2.257, p = 0.107). A reason for this non-significance is a loss of statistical power due to a small sample size included in the paired t-test. To visualise the loss of information between visit 1 and visit 8, we illustrated the complex individual trajectories of the participants in Figure 2. Moreover, as described in Table S1 in the supplement, the participants at visit 8 (63/190) analysed in the t-test were inherently significantly different compared to the non-participants at visit 8 (127/190): they were younger, had better education, and most importantly their baseline apathy scores were lower. Consequently, those with the better overall situation kept coming back while this was not the case for those with a worse outcome at baseline, which explains the observed (non-significant) increase. This may result in a biased estimation of change in apathy when analysed by a paired t-test.

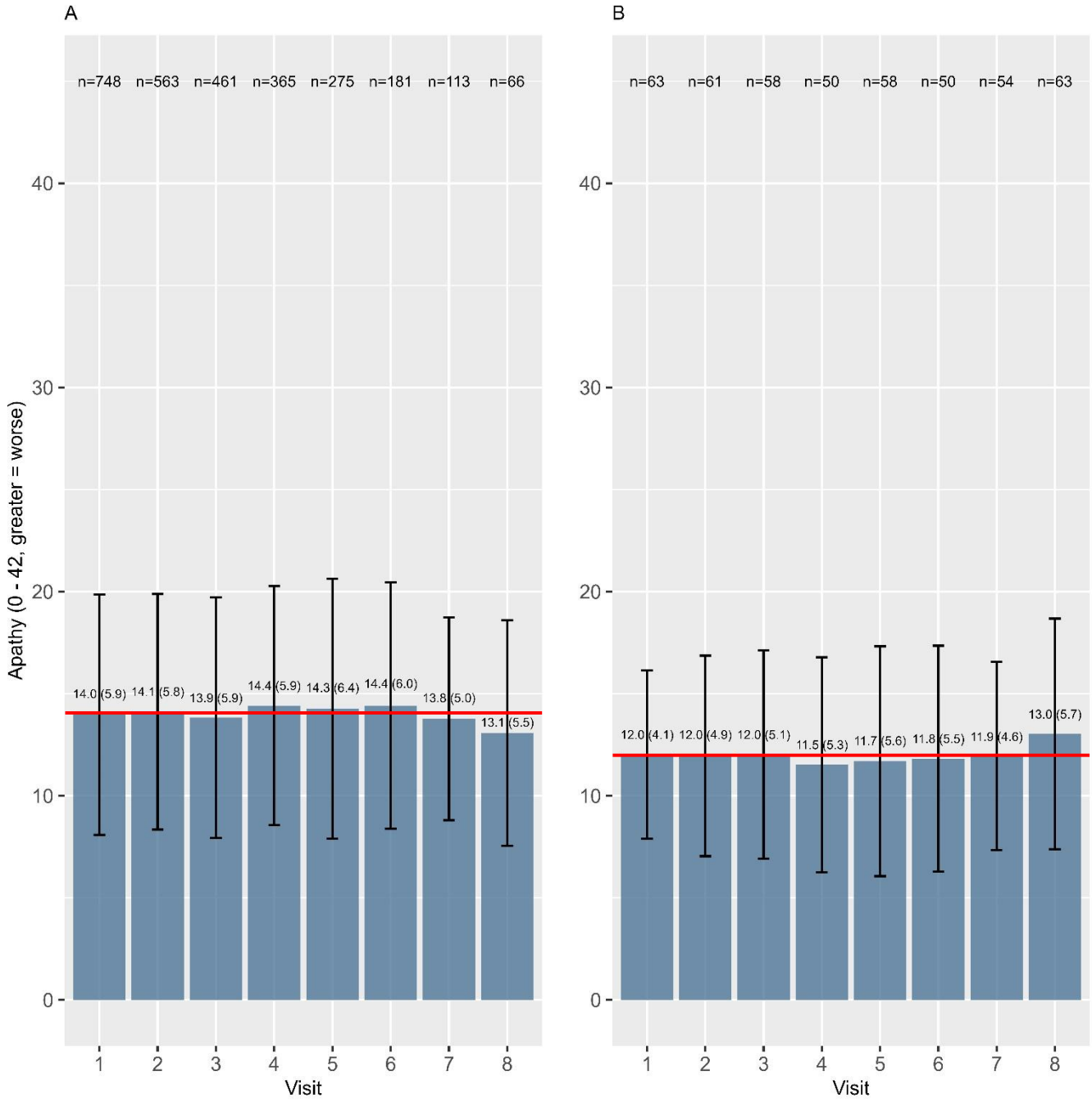

Figure 1: Bar charts illustrating apathy scores (means and standard deviations) per visit (Panel A: all participants, Panel B: subgroup analysed in the t-test). The red line indicates the mean apathy at visit 1

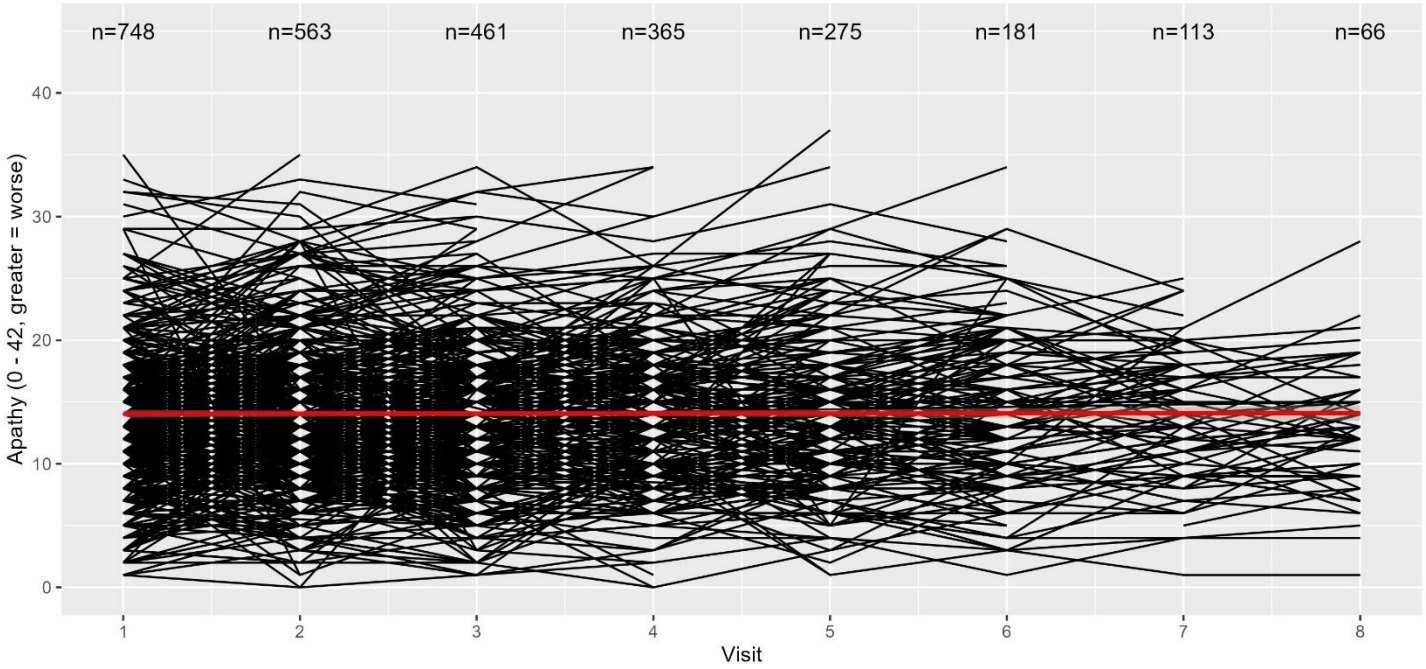

Figure 2: Scatterplot illustrating the individual trajectories. The red line indicates the regression line.

Table 1: Results from the group comparison, the linear regression and the linear mixed models

Hypo-thesis	**Statistical test**	**R-Code**	**Effect**	**95% CI**	**p-value**
1	Paired t-test	Stats::t.test(wide_data$apathy_score_at_visit_8, wide_data$apathy_score_at_visit_1, paired = TRUE)	+1.016/ from visit 1 to visit 8	-0.225, 2.257	0.107
2	Linear regression	Stats::lm(apathy~visit, data = long_data)	-0.008 / year	-0.107, 0.1224	0.897
3	Linear mixed effects models	lme4::lmer(apathy~visit+(1+visit|subject_ID), REML = FALSE, data=long_data)	+0.335 / year	0.235, 0.434	< 0.001

Hypothesis 2 and 3: Increase of apathy symptoms with the number of visits

From the results in Table 1, we see that the linear regression coefficient, representing change in apathy symptoms per year, is not significantly different from zero, indicating no change over time. On the contrary, the effect estimates for the linear mixed effects models indicated a significant increase in apathy symptoms per year by +0.335 points (95%CI: 0. 235, 0.434, $p < 0.001$). Consequently, mixed effects models were the only method able to detect an increase in apathy symptoms over time and choosing mixed effect models for the analysis of longitudinal data reduces the risk of false negative results.

Comparison of the different statistical methods

The effect sizes differed depending on the choice of the statistical method. Thus, the paired t-test and the linear regression resulted in an output that would lead to different interpretations than the mixed effects models. More specifically, compared to the t-test and linear regression (which indicated non-significant changes in apathy of only +1.016 points over seven years and -0.008 points per year, respectively), the linear mixed effects models found an increase of +0.335 points per year on the apathy scale. This increase is more than twice as high (+2.345 points in seven years) as indicated by the t-test and suggests linear mixed models is a more sensitive approach to detect meaningful changes perceived by people with PD over time. The differences in the effect sizes are also reflected in the regression lines in Panel A and B of Figure 3.

Mixed effects models are a valuable tool in longitudinal data analysis as these models expand upon linear regression models by considering the correlation among repeated measurements within the same individuals through the estimation of a random intercept [11-13]. Specifically, to account for correlation between observations, linear mixed effects models use random effects to explicitly model the correlation structure, thus removing correlation from the error term. A random slope in addition to a random intercept allows both the rate of change and the mean value to vary by participant, capturing individual differences. This distinguishes them from group comparisons or standard linear regressions, in which such explicit modelling of correlation is not possible. Thus, the linear regression not considering correlation among the repeated observations leads to an underestimation of longitudinal change, explaining the smaller effect sizes and insignificant results of the regression. By including random effects, linear mixed effects models can better capture the variability within the data.

Missing data in longitudinal studies

Another common challenge in longitudinal studies is missing data. Compared to the paired t-test and regression, the mixed effects models can handle missing data by including also participants with missing data at single visits and by accounting for the individual trajectories of each participant as illustrated in Figure 2 [14]. The mixed effects models provide a valuable alternative to the paired t-test and linear regression as its assumptions are valid when the data is missing at random [15]. As mentioned in relation to the t-test above, characteristics at visit 1 were associated with missing data at visit 8. Thus, further steps may be required to handle missing data [14] when using the compared statistical methods. Note that we do not further elaborate here on this topic since this is a separate issue to statistical method comparison. Further information on can be found elsewhere [15].

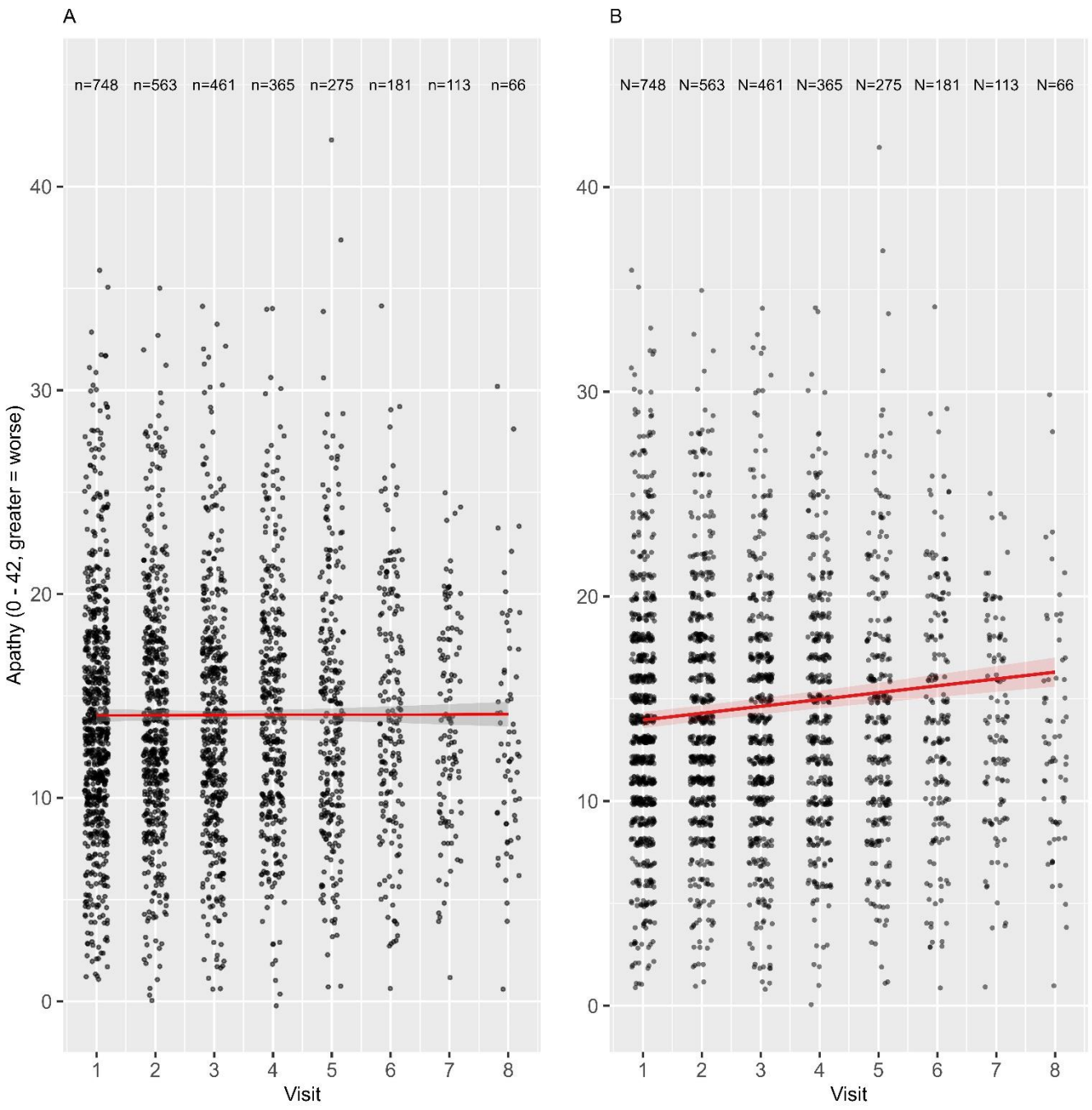

Figure 3: Scatterplot illustrating the relationship between visit number and apathy. Apathy measured by a whole number interval scale, jitter applied on x- and y-axis to illustrate the data points (Panel A: Linear regression, Panel B: Linear mixed effects model). The red line indicates the regression line.

4. Conclusion

Mixed effects models were the only method able to detect progression of apathy over time. While the effects estimated for the group comparison and the linear regression were smaller with high p-values, indicating a statistically insignificant change in apathy over time, effect estimates for the mixed effects models were positive with a very small p-value, indicating a statistically significant increase in apathy symptoms per year in line with clinical expectations. Mixed effects models can be used to estimate different types of longitudinal effects while an inappropriate use of paired t-tests and linear regression to analyse longitudinal data can lead to underpowered analyses and an underestimation of longitudinal change and thus clinical significance. Therefore, researchers should more often consider mixed effects models for longitudinal analysis. In case this is not possible, limitations of the analytical approach need to be discussed and taken into account in the interpretation.

Acknowledgements

We would like to thank all participants of the Luxembourg Parkinson's Study for their important support of our research. Furthermore, we acknowledge the joint effort of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) Consortium members from the partner institutions Luxembourg Centre for Systems Biomedicine, Luxembourg Institute of Health, Centre Hospitalier de Luxembourg, and Laboratoire National de Santé generally contributing to the Luxembourg Parkinson's Study as listed below:

Geeta ACHARYA 2, Gloria AGUAYO 2, Myriam ALEXANDRE 2, Muhammad ALI 1, Wim AMMERLANN 2, Giuseppe ARENA 1, Michele BASSIS 1, Roxane BATUTU 3, Katy BEAUMONT 2, Sibylle BÉCHET 3, Camille BELLORA 2, Guy BERCHEM 3, Alexandre BISDORFF 5, Ibrahim BOUSSAAD 1, David BOUVIER 4, Jessica CALMES 2, Lorieza CASTILLO 2, Gessica CONTESOTTO 2, Nancy DE BREMAEKER 3, Nico DIEDERICH 3, Rene DONDELINGER 5, Nancy E. RAMIA 1, Daniela ESTEVES 2, Katrin FRAUENKNECHT 4, Joëlle FRITZ 2, Manon GANTENBEIN 2, Piotr GAWRON 1, Soumyabrata GHOSH 1, Marijus GIRAITIS 2,3, Enrico GLAAB 1, Martine GOERGEN 3, Elisa GÓMEZ DE LOPE 1, Jérôme GRAAS 2, Mariella GRAZIANO 7, Valentin GROUES 1, Anne GRÜNEWALD 1, Gaël HAMMOT 2, Anne-Marie HANFF 2, 10, 11, Linda HANSEN 3, Michael HENEKA 1, Estelle HENRY 2, Sylvia HERBRINK 3, Sascha HERZINGER 1, Michael HEYMANN 2, Alexander HUNDT 2, Nadine JACOBY 8, Sonja JÓNSDÓTTIR 2,3, Jochen KLUCKEN 1,2,3, Rejko KRÜGER 1,2,3, Pauline LAMBERT 2, Zied LANDOULSI 1, Roseline LENTZ 6, Laura LONGHINO 3, Victoria LORENTZ 2, Paula Cristina LUPU 2, Tainá M. MARQUES 2, Guilherme MARQUES 2, Patricia MARTINS CONDE 1, Patrick MAY 1, Deborah MCINTYRE 2, Chouaib MEDIOUNI 2, Francoise MEISCH 1, Myriam MENSTER 2, Maura MINELLI 2, Michel MITTELBRONN 1, 2, 4, 10, 12, 13, Romain NATI 3, Ulf NEHRBASS 2, Sarah NICKELS 1, Beatrice NICOLAI 3, Jean-Paul NICOLAY 9, Fozia NOOR 2, Clarissa P. C. GOMES 1, Sinthuja PACHCHEK 1, Claire PAULY 2,3, Laure PAULY 2, 10, Lukas PAVELKA 2,3, Magali PERQUIN 2, Rosalina RAMOS LIMA 2, Armin RAUSCHENBERGER 1, Rajesh RAWAL 1, Dheeraj REDDY BOBBILI 1, Kirsten ROOMP 1, Eduardo ROSALES 2, Stefano SAPIENZA 1, Venkata SATAGOPAM 1, Margaux SCHMITT 2, Sabine SCHMITZ 1, Reinhard SCHNEIDER 1, Jens SCHWAMBORN 1, Raquel SEVERINO 2, Amir SHARIFY 2, Ruxandra SOARE 1, Ekaterina SOBOLEVA 1,3, Kate SOKOLOWSKA 2, Hermann THIEN 2, Elodie THIRY 3, Rebecca TING JIIN LOO 1, Johanna TROUET 2, Olena TSURKALENKO 2, Michel VAILLANT 2, Mesele VALENTI 2, Carlos VEGA 2, Liliana VILAS BOAS 3, Paul WILMES 1, Evi WOLLSCHEID-LENGELING 1, Gelani ZELIMKHANOV 2,3

1 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
2 Luxembourg Institute of Health, Strassen, Luxembourg
3 Centre Hospitalier de Luxembourg, Strassen, Luxembourg
4 Laboratoire National de Santé, Dudelange, Luxembourg
5 Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg
6 Parkinson Luxembourg Association, Leudelange, Luxembourg
7 Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg
8 Private practice, Ettelbruck, Luxembourg

9 Private practice, Luxembourg, Luxembourg
10 Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
11 Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre+, Maastricht, the Netherlands
12 Luxembourg Center of Neuropathology, Dudelange, Luxembourg
13 Department of Life Sciences and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Detailed author roles

A-MH: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Project administration, Writing – original draft, Writing – review & editing.

RK: Conceptualization, Methodology, Funding, Resources, Supervision, Writing – review & editing.

CMC & CL: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Disclosures

Funding sources and Conflict of Interest

This work was supported by grants from the Luxembourg National Research Fund (FNR) within the National Centre of Excellence in Research on Parkinson's disease [NCERPD(FNR/NCER13/BM/11264123)].

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethical Compliance Statement

The studies involving human participants were reviewed and obtained a positive opinion from the National Ethics Board (CNER Ref: 201407/13). The patients/participants provided their written informed consent to participate in this study. We confirm that we have read the Journal’s position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

References

1. Levy, R. and B. Dubois, Apathy and the functional anatomy of the prefrontal cortex-basal ganglia circuits. Cereb Cortex, 2006. 16(7): p. 916-28. Doi: 10.1093/cercor/bhj043
2. Poewe, W., et al., Parkinson disease. Nat Rev Dis Primers, 2017. 3: p. 17013. Doi: 10.1038/nrdp.2017.13
3. Hipp, G., et al., The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. Front Aging Neurosci, 2018. 10: p. 326. Doi: 10.3389/fnagi.2018.00326
4. Pavelka, L., et al., Luxembourg Parkinson’s study -comprehensive baseline analysis of Parkinson’s disease and atypical parkinsonism. Front Neurol, 2023. 14. Doi: 10.3389/fneur.2023.1330321
5. Starkstein, S.E., et al., Reliability, validity, and clinical correlates of apathy in Parkinson's disease. J Neuropsychiatry Clin Neurosci, 1992. 4(2): p. 134-9. Doi: 10.1176/jnp.4.2.134

6. Leentjens, A.F., et al., Apathy and anhedonia rating scales in Parkinson's disease: critique and recommendations. Mov Disord, 2008. 23(14): p. 2004-14. Doi: 10.1002/mds.22229
7. R Core Team. R: A language and environment for statistical computing. 2021; Available from: https://www.R-project.org/.
8. Bates, D., et al., Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw, 2015. 67: p. 1-48. Doi: 10.18637/jss.v067.i01
9. Lüdecke, D. sjPlot: Data Visualization for Statistics in Social Science. 2022; R package version 2.8.11]. Available from: https://CRAN.R-project.org/package=sjPlot.
10. Little, R.J.A., A Test of Missing Completely at Random for Multivariate Data with Missing Values. J Amer Statist Assoc, 1988. 83(404). Doi: 10.1080/01621459.1988.10478722
11. Twisk, J.W.R., Applied Multilevel Analysis: A Practical Guide for Medical Researchers. 2006: Cambridge University Press.
12. Twisk, J.W.R., Applied Longitudinal Data Analysis for Epidemiology. A Practical Guide. Vol. 2. 2013: Cambridge University Press.
13. Twisk, J.W.R., Applied Mixed Model Analysis. A Practical Guide. Practical Guides to Biostatistics and Epidemiology. Vol. 2. 2019, New York.
14. Long, D.J., Longitudinal data analysis for the behavioral sciences using R. Vol. 1. 2012, United States of America: SAGE.
15. Twisk, J.W.R., et al., Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J Clin Epidemiol, 2013. 66(9): p. 1022-8. Doi: 10.1016/j.jclinepi.2013.03.017

Supplement

Figure S1: Flow-chart

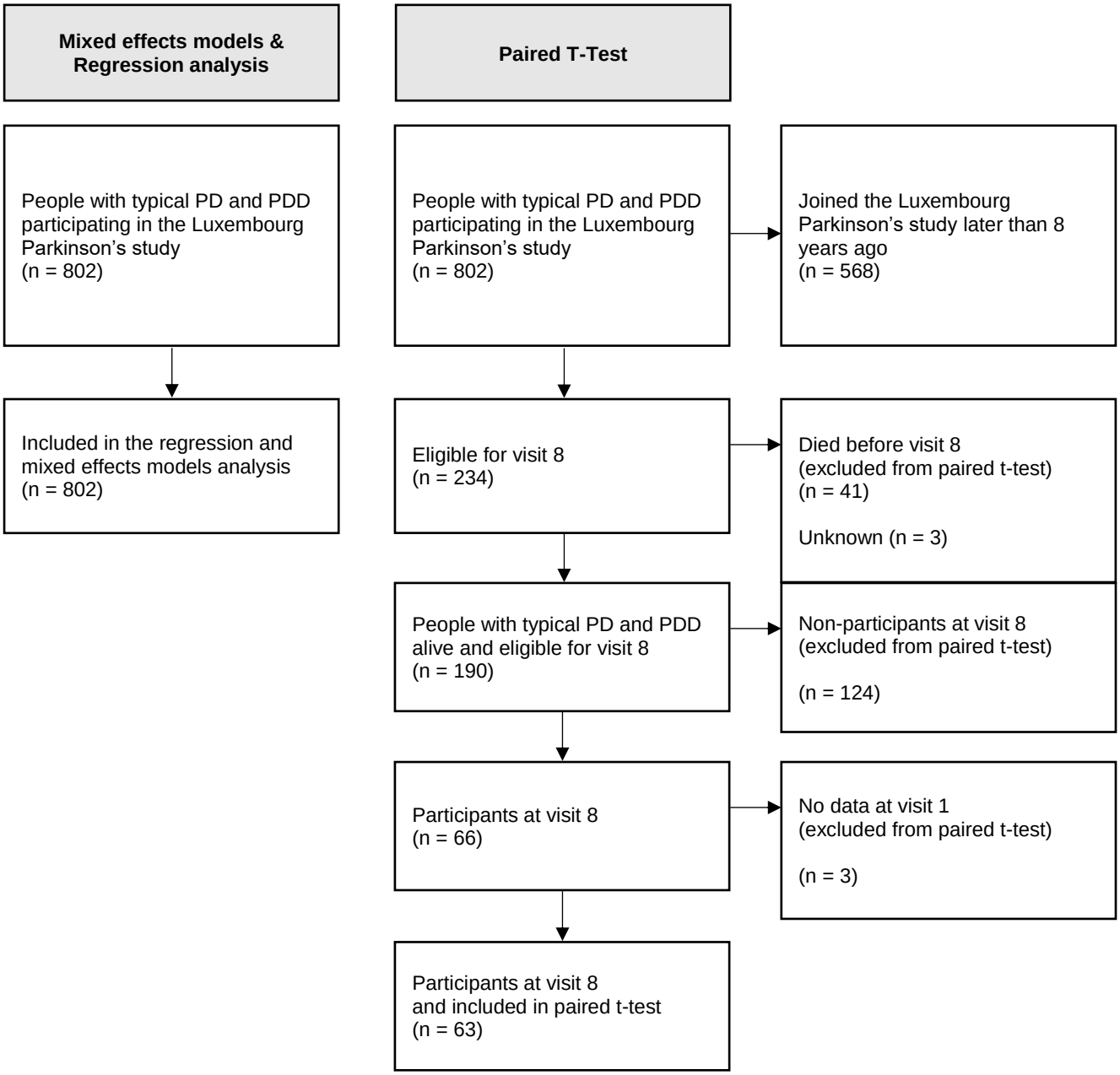

Table S1: Comparison of characteristics between particpants at visit 8 (included in the paired t-test) and the non-participants at visit 8 (excluded from the paired t-test)

Values at baseline	**Participants without visit 8** (N = 127)	**Participants with visit 8** (N = 63)	**p-value**
Apathy Score	14.3 (6.0)	12.0 (4.1)	p = 0.003
Age (y.)	65.8 (11.7)	62.3 (9.6)	p = 0.029
Age at diagnosis (y.)	59.5 (12.8)	58.0 (10.3)	p = 0.367
Years of education	12.6 (3.6)	14.0 (3.5)	p = 0.009