Original Article

JHSMR Journal of Health Science and Medical Research

Application of Longitudinal Data the Multilevel Models Approach on Diabetes Mellitus

Wudneh Ketema Moges, Ph.D.,Tena Manaye Endalamaw, M.Sc.

Department of Statistics, Debre Berhan University, Debre Berhan 445, Ethiopia.

Received 23 July 2022 ● Revised 28 October 2022 ● Accepted 2 December 2022 ● Published online 8 March 2023

Abstract:

Objective: Diabetes mellitus is a metabolic disorder that develops over time and affects the cardiovascular system, eyes, kidneys, nerves, and blood sugar levels. The aim of this investigation was to determine the prevalence of diabetic mellitus patients, identify the associating risk factors using a multilevel longitudinal model, and understand the multilevel model changes for the level-1 and level-2 models.

Material and Methods: We examined such types of scenarios using multilevel longitudinal models such as the simple random intercept multilevel model, the random coefficient model, and the null model.

Results: There were 248 individuals with diabetes mellitus enrolled in the study for follow-up measurements over 4 time points, among these 248 individuals, 211 had complete data for all four time points. Based on the intraclass correlation coefficient, much of the variability (88.35%) in diabetes mellitus patients was accounted for by the follow-up time in this study, whereas 11.65% of the variability could not be accounted for by the follow-up time. Moreover, the data analysis suggested that sex had a significant effect on diabetes mellitus patients with the progression of time.

Conclusion: Based on the results of our study, sex, baseline fasting and educational status had a significant effect on diabetes mellitus patients over time. The educational status of diabetes mellitus patients was found to have a significant effect throughout the follow-up time; this shows that when treating diabetes mellitus patients, the physician should beware of the nature of the disease and how to management diabetes requires a high level of awareness and motivation on part of the patients regarding self-care.

Keywords: ANOVA, diabetes, longitudinal, MACOVA, multilevel

Contact: Wudneh Ketema Moges, Ph.D.
Department of Statistics, Debre Berhan University, Debre Berhan 445, Ethiopia.
E-mail: wudnehketema@gmail.com

J Health Sci Med Res 2023;41(3):e2023936
doi: 10.31584/jhsmr.2023936
www.jhsmr.org

Introduction

Diabetes mellitus is a metabolic disorder that develops over time and affects the cardiovascular system, eyes, kidneys, nerves, and blood sugar levels. Diabetes mellitus is associated with elevated blood glucose levels. Type 2 diabetes mellitus is the most common type, affecting mostly adults, and develops when the body either stops producing enough insulin or becomes resistant to it. Type 2 diabetes mellitus has become significantly more common during the past three decades in nations of all income levels. Diabetes mellitus type 1 is a long-term condition in which the pancreas produces little or no insulin on its own. Previously, it was known as juvenile diabetes or insulin-dependent diabetes mellitus. Access to cheap medical care, such as insulin injections, is essential for diabetes mellitus patients to survive[1].

Diabetes mellitus screening can help identify people, who have undiagnosed, asymptomatic type 2 diabetes mellitus or prediabetes and are eligible for evidence-based interventions to delay the progression of type 2 diabetes mellitus and its complications. The American Diabetes Association (ADA) has developed national screening guidelines[2], and the United States Preventive Services Task Force (USPSTF)[3] has offered recommendations in order to aid clinical practice in relation to diabetes mellitus patients. Despite the screening guidelines and clinical practice recommendations in place, there are over 7 million type 2 diabetic adults in the United States and 74 million type 1 diabetics. According to nationally representative data in the USA, there are significant gaps between screening eligibility and screening completion, with only half of individuals meeting the ADA and USPSTF guidelines reporting a completed screening test[3,4].

In the presence of follow-up measurements, it has been observed that response variables change over time; therefore, longitudinal modeling techniques are required to appropriately account for the dependence that exists among repeated measurements[5]. Using cross-sectional models, which are appropriate only for data with a single measurement per subject, while dealing with longitudinal data is inappropriate and will result in unreliable conclusions. Due to the complexity of such studies and the correlated nature of longitudinal data, more sophisticated models are required to account for the interdependence of multiple outcome values observed for each subject at different time points[6].

Multi-level models (MLM), also known as hierarchical linear models[7], random coefficient models[8], and mixed-effect models[9], have grown in popularity in their application to analyze multiwave longitudinal data. Although multilevel models have been widely used in educational research for more than two decades[10], they are still relatively new to rehabilitation psychology researchers. As multilevel models as well as hierarchical and individual growth models increase in popularity, the need for reliable and flexible software that can be used to fit them to their corresponding data sets also increases.

The primary purpose of longitudinal data analyses is the observation of the effects of covariates on the response changes over time, which also brings to light the difference between cross-sectional and repeated measurements. Furthermore, longitudinal data can be used to examine the variation within and between response levels over time.

Objective

The aim of our investigation was to determine the prevalence of diabetic mellitus patients, identify the risk factors of diabetes mellitus using a multilevel longitudinal model and understand the multilevel model changes for level-1 and level-2 models.

Material and Methods

Data description: In this work, diabetes mellitus patients with both type I and type II, who were prescribed insulin and metformin, were monitored at Debre Berhan Referral Hospital between September 1, 2012 and August 30, 2015 via four time points during the follow-up period.

Patients were assessed four times:

- When the program was initiated, which served as the baseline (time T_0)
- At 6 months (time T_1)
- At 12 months (time T_2)
- At 18 months (time T_3)

Multilevel model

Multilevel modeling is the process of examining data that include variables assessed at several hierarchy levels. Subjects on the same level or cluster of a multilevel data set or subjects in a particular level or cluster may be more similar to one another compared to subjects belonging to other levels or clusters[4].

Simple random intercept model

The intercept can differ between several clusters using a straightforward random intercept model. There is a simple linear regression model a single intercept shared by all observations in a data collection.

In this section, we'll start with the null models, which are simple random intercept models. Null models are those that have no independent variables.

The following are the two levels of a null multilevel model:

Level 1: $y_{ij}=\beta_{oj}+\epsilon_{ij}$

Level 2: $\beta_{oj}=\gamma_{00}+u_{oj}$

Random slope coefficient model

Simple intercept-only models are used to create random slope coefficient models. To do this, level-1 independent predictor variables are added to straightforward intercept-only models. Equations for the two levels of the random slope coefficient model can be expressed as:

Level 1: $y_{ij}=\beta_{oj}+\beta_1 x+\epsilon_{ij}$

Level 2: $\beta_{oj}=\gamma_{00}+u_{oj}$

$\beta_{1j}=\gamma_{10}+u_{1j}$

Longitudinal data multilevel models

Multilevel models are used when variables are measured at many levels of the hierarchy. Multilevel models are useful for analyzing within- and between-person changes in longitudinal data because they distinguish between how people change over time and how these changes differ between individuals[11]. This study examined longitudinal diabetic patient data from a multilevel perspective.

Multilevel model for longitudinal data

For analyzing hierarchically structured data, multilevel models were developed. Students nested within schools and employees nested within companies are two examples of hierarchically structured data. As a result, a hierarchy is made up of lower-level observations (individual-level data) that are nested within higher levels (group-level data). Multilevel models are analyses of models that contain variables measured at different levels of the hierarchy.

Multilevel models are useful in the analysis of within-person and between-person changes in longitudinal data by distinguishing two things: how individuals change over time and how these changes vary across individuals[12].

Level 1: Within-person variation (WP): that means "INTRA-individual[13] differences" – time-varying.

Level 2: Between-person variation (BP): "INTER-individual differences" – time-invariant.

Ethics approval and informed consent

The participants of previous studies provided informed consent for the use of the study data, and the

initial study protocol was approved by the Ethics Committee of Debre Berhan Referral Hospital. The ethics review committee approved the use of the secondary data in the current investigation.

Results and Discussion

There were 248 individuals with diabetes mellitus enrolled in the study, who were assessmed at least at one time, point. Among these 248 individuals, 211 had complete data for all four time points.

The descriptive statistics of the given samples are presented in Table 1. We conducted an analysis on all socio-demographic variables. The findings regarding the functional status of diabetes mellitus patients revealed that about 236 (95.16%) had working diabetes mellitus whereas 12 (4.86%) had ambulatory diabetes mellitus; this indicates that the majority of our diabetes mellitus patients could be categorized under the working status diabetes mellitus. Concerning the clinical diagnosis of the diabetes mellitus patients investigated in this study about 159 (64.01%) were type 1 and 89(35.99%) were type 2 diabetics.

As can be seen from Table 2, the patients' average fasting blood sugar levels showed an increasing trend during the first three follow-up time points (179.53 mg/dL, 189.31 mg/dL, and 190.38 mg/dL at baseline, time point 1, and time point 2, respectively); however, they decreased at the end of the study (184.87 mg/dL at time point 3). The highest standard deviation values were found at the baseline time point compared to the other time points. Therefore, this indicates that the measurement values of the fasting blood sugar level both increased and decreased over the follow-up times, indicating that, in regard to this response, the data were characterized by both intermittent and dropout or missing observations; the missing values increased over time.

Table 1 Socio-demographic characteristics of diabetes mellitus patients at Debre Berhan Referral Hospital

Variables		Patient number	Percentage
Sex	Male	137	55.24%
	Female	111	44.76%
Functional status	Working	236	95.16%
	Ambulatory	12	4.86%
Marital status	Married	194	78.23%
	Single	54	21.77%
Residence setting	Urban	159	64.11%
	Rural	89	35.89%
Educational status	Illiterate	87	35.08%
	Primary	67	27.02%
	Secondary	44	17.74%
	Tertiary	50	20.16%
Occupation	Full-time employment	137	55.24%
	Not working	111	54.76%
Clinical diagnosis	Type I	159	64.01%
	Type II	89	35.99%

Table 2 Descriptive statistics for fasting glucose level at time point

Time	Baseline	Time 1	Time 2	Time 3
N	248 (100%)	232 (93.5%)	239 (96.41%)	236 (95.2%)
Mean fasting blood glucose level (mg/dL)	179.53	189.31	190.38	184.87
S.D. (mg/dL)	±94.800	±92.950	±96.855	±89.998
Max. value (mg/dL)	600	574	600	587
Min. value (mg/dL)	22	22	29	31
95% CL	(167.68, 191.34)	(177.28, 201.33)	(178.04, 202.73)	(173.33, 196.41)
Missing data	0	16	9	12

N=number of patient, 95% CL=95% confidence level

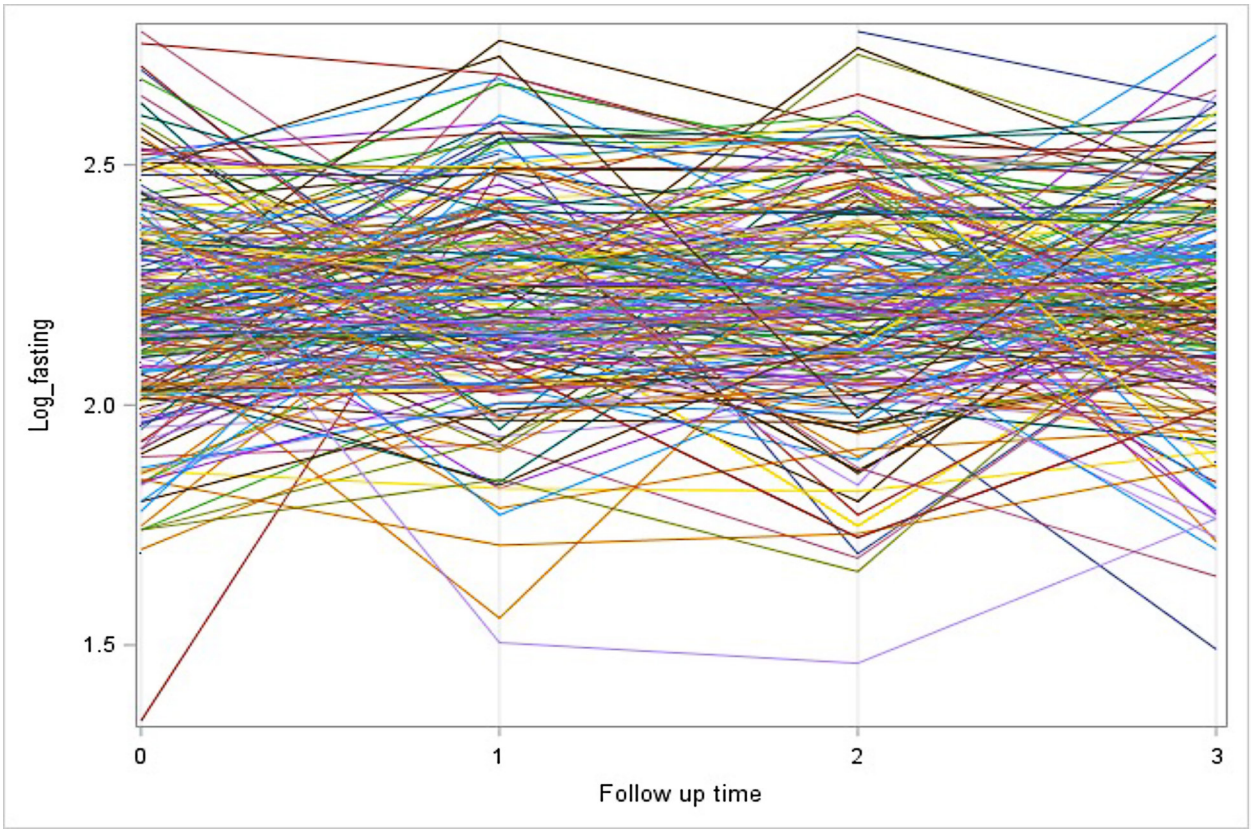

Figure 1 Predicted individual growth curves random intercept model

When considering the 95% confidence level (CL) for a given time point, the narrowest confidence interval was observed between time point 3 and the baseline time point, whereas the widest 95% CL was found among time point 2 and time point 1. Based on the results, the smallest variation in fasting glucose levels among our diabetes mellitus patients was detected between the values of time point 3 and the baseline time point.

Figure 1 depicts how the shape of the response-profile plot is approximately the same for all subjects at each follow-up time point. However, the profile plots are shifted up or down to better match the subjects' individual profiles. Generally, there was little variability at the beginning compared to the end of the study.

SAS results

```
Proc mixed data=dm covtset;
Model log fasting=/solution;
Random intercept/subject=class type=un;
Run;
```

The "contest" is required to calculate the standard errors of variance in component estimates and to specify the type of the unstructured covariance matrix[14].

Based on the results obtained using SAS software, we notice that the estimated grand mean, is 2.217, this indicates the mean fasting glucose level across all patients. The population mean is different from significantly (T-value of 318.82) which indicates that it should be included in the model shown model (Table 3).

The "within-person" variance ($\delta^{\wedge 2}$) was 0.046, and the "between-person" variance (δ^2_0) was 0.351. This proc mixed syntax generated the covariance parameter estimates, which are shown in Table 4. We were also able to compute the intraclass correlation coefficient (ICC) that indicates how much of the total variation in patient care is accounted for by time[15].

The equation for calculating the ICC is provided below. In this equation, '$\sigma^{\wedge 2}$time' refers to the covariance estimates for the intercept, and '$\sigma^{\wedge}$error' refers to the covariance estimate for the residual.

$$ICC = \frac{\sigma^2 time}{\sigma^2 time + \sigma^2}$$

Thus, based on the output above, we calculated the ICC as

$$ICC = \frac{0.3505}{0.3505+0.04619} = 0.8835$$

This indicated that 88.35% of the variability in the diabetes mellitus patients in this study was accounted for by follow-up time, whereas 11.65% of the variability could not to be accounted for by follow-up time. These results answered our first research question and provided a two-level model. The high proportion of variation in the follow-up time of diabetes mellitus patients emphasized the significance of accounting for the data's hierarchical structure.

Broadly speaking, we can conclude that the 88.35% variation observed among our diabetes mellitus patients according to their fasting blood glucose levels can be attributed to individual differences.

Our results indicated that the fixed effect, $\gamma^{\wedge}_{00}$, was 2.211; this represents the average mean log fasting sugar level across all individuals with a baseline fasting value. Similarly, $\gamma^{\wedge}_{10}$ was 0.0048; this describes the rate at which Y_{ij}, the individual i's situation, changes when the predictor's variable is included as a level-I predictor within the group.

The "within-person" variance ($\delta^{\wedge 2}_{\varepsilon}$) is 0.039, As a result, this figure represents the average scatter of an individual's observed outcome values around his or her own true change trajectory of diabetes mellitus patients, therefore this finding is line up with[15]. $\delta^{\wedge 2}_{\varepsilon}$ is still significantly smaller for all diabetes mellitus unconditional growth model than what we obtained for null model.

Table 3 Null model

Effect	Estimate	Standard error	DF	t-value	p-value
Intercept	2.217	0.007	3	318.82	<0.0001

DF=degree of freedom

Table 4 Covariance parameter estimates

Cov Parm	Subject	Estimate	Standard error	z-value	p-value
Intercept	Time	0.351	0.002	3.84	<0.0001
Residual		0.0462	0.002	21.84	<0.0001

Table 5 Unconditional growth model

Cov Parm	Estimate	Standard error	z-value	p-value
Time	0.002	0.0004	4.00	<0.0001
Residual	0.039	0.002	18.83	<0.0001

Table 6 Solution for fixed effects

Effect	Estimate	Standard error	DF	t-value	p-value
Intercept	2.211	0.011	707	206.52	<0.0001
Time	0.005	0.006	246	0.75	0.4512

The value of the between group variability was $\delta^{\wedge 2}_{0}$ 4.885 (2.211^2), therefore this indicates that much higher than within person patients.

As shown in Table 7, the intercept represents the average fasting blood sugar level counts. The fact that the male diabetes patients estimate (-0.008)=0.99 was higher than that of the reference group (p-value=0.017) indicated that sex exerted a significant effect on diabetes mellitus patients with the progression of time[7].

The educational status of diabetes mellitus patients is significant effect through the follow up time, so this showed that when an individual have had a diabetes mellitus patients, a physician should beware of the nature of the disease and how to manage diabetes requires a high level of awareness and motivation on part of the patients regarding self-care, so this investigation contradict with the theory[16].

The software requires values for the duration of the study, frequency of observations, level -1 variance,

Table 7 Conditional growth model (full model)

Variables		Estimate	Standard error	DF	t-value	p-value
Intercept		1.768	0.031	898	57.13	<0.0001
Age		-0.00003	0.0002	939	-0.16	0.873
Sex	Male	-0.008	0.006	939	-1.36	0.017
	Female	0	.	.	.	.
Weight		0.000	0.0004	939	0.78	0.4353
Baseline_F		-9.78E-6	0.00002	939	-0.38	0.7016
Fasting_B		0.002		939	10	<0.0001
F_status	Working	0.006	0.0002 0.011	939	0.52	0.601
	Ambulatory	0	.	.	.	.
M_status	Married	-0.005	0.007	939	-0.69	0.491
	Single	0	.	.	.	.
Residence setting	Urban	-0.005	0.006	939	-0.75	0.455
	Rural	0	.	.	.	.
Edu_L	Illiterate	0.017	0.008	939	1.97	0.049
	Primary	0.021	0.008	939	2.75	0.006
	Secondary	0.013	0.008	939	1.68	0.094
	Tertiary	0	.	.	.	.
Occupation	Full-time	0.003	0.006	939	0.50	0.620
	Part-time	0	.	.	.	.
BMI		0.0007	0.001	939	0.58	0.564
Clinical_D	Type I	0.009	0.008	939	1.12	0.261
	Type II	0	.	.	.	.

between-person variability in the parameter of interest, and an estimate the effect size for a two-level multilevel model, so this indicates that there is a variability under level-1 and level-2 for diabetes mellitus patients , therefore this study line up with the theory[17].

In addition, we examined the intra-class correlation coefficient (ICC) of each model after adding the level-1 predictors. However, when we included the level-2 predictors, the ICC sharply dropped to a value that was even lower than that of the unconditional model. This is due to a decrease in the unexplained level-2 variation, i.e., the random intercept term. When everyone is evaluated on the same amount of events that are evenly spread throughout time, a multilevel model can be utilized. The multilevel longitudinal model, however, can also be used when the number of measurement waves is not uniform across individuals and when measurement point spacing is not uniform across individuals (for instance, the interval between diabetes mellitus patient screenings may vary across participants)[18].

In summary, multilevel longitudinal models address all of the research questions that repeated measures ANOVA/MANOVA tests are designed to answer without being constrained by the latter's rigid assumptions[13].

Conclusion

This paper briefly describes the characteristics and the application areas of multilevel modeling with the use of longitudinal data.

We demonstrated how to analyze multilevel longitudinal data and obtain some basic descriptive statistics for such data, e.g., means, variances, and standard deviation. Furthermore, we used spaghetti plots to detect any potential trends in the data over time. Based on the likelihood ratio test findings, the unstructured covariance matrix fit the given data better.

According to the results of the study, sex, baseline fasting glucose level, and educational status had a significant effect on diabetes mellitus patients as time passed.

The educational status of diabetes mellitus patients has a significant effect over time, indicating that when an individual has diabetes, the physician should beware of the nature of the disease and management of diabetes requires a high level of awareness and motivation on part of the patients regarding self-care.

Acknowledgements

We would like to thank Debre Berhan Referral Hospital in Ethiopia for providing the data that made this research possible as well as the patients, who agreed to participate in the study.

Conflict of interests

The authors report no conflicts of interest related to this work.

References

1. American diabetes association. Classification and diagnosis of diabetes: standards of medical care in diabetes. Diabetes Care 2019;42 (Suppl 1):S13–28.
2. National center for chronic disease preventive and health promotion, division of diabetes translation. National Diabetes Statistics Report 2017.
3. Bullard KM, Saydah SH, Imperatore G, Cowie CC, Gregg EW, Geiss LS, et al. Secular changes in U.S. prediabetes prevalence defined by hemoglobin a1c and fasting plasma glucose: national health and nutrition examination surveys, 1999–2010. Diabetes Care 2013;36:2286–93.
4. Kiefer MM, Silverman JB, Young BA, Nelson KM. National patterns in diabetes screening data from the National Health and Nutrition Examination Survey (NHANES) 2005–2012. J Gen Intern Med 2015;30:612–8.
5. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
6. Siu AL, U S preventive service task force. Screening for abnormal blood glucose and type diabetes mellitus: U.S. preventive services task force recommendation statement. Ann Intern Med 2015;163:861–8.
7. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods. 2nd ed. California: Sage Publications; 2002.
8. Langford NT. Random coefficient models. New York: Oxford University Press; 1993.
9. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberber O. SAS for Mixed Models. 2nd ed. North Carolina: SAS Institute; 2006.
10. Raudenbush SW. Educational applications of hierarchical models: a review. J Educ Stat 1988;13:85–116.
11. Singer JD, Willett JB. Applied longitudinal data analysis: modeling change and event occurrence. New York: Oxford University Press; 2003.
12. Kreft IG, de Leeuw J. Introducing multilevel modeling. California: Sage; 1999.
13. Kwok O, West SG, Green SB. The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: a Monte Carlo study. Multivariate Behav Res 2007;42:557–92.
14. Bell BA, Ene M, Smiley W, Schoeneberger JA. A multilevel model primer using SAS® PROC mixed. Stat Data Anal 2013;433,1–19.
15. Su R, Cai L, Cui W, He J, You D, Golden A. Multilevel analysis of socioeconomic determinants on diabetes prevalence,

awareness, treatment and self-management in ethnic minorities of Yunnan Province, China. Int J Environ Res Public Health 2016:13;751.

16. The department of statistics and data sciences, The University of Texas at Austin. Multilevel modeling tutorial: using SAS, Stata, HLM, R, SPSS, and Mplus. Texas: The University of Texas at Austin; 2015.
17. Tantular B, Jaya IG, Zulhanif. Multilevel models approach on longitudinal studies for disease data. J Phys Conf Ser 2018;1025;012108.
18. Moges WK, Muralidharan AR, Tadesse HZ. Gaussian longitudinal analysis of progression of diabetes mellitus patients using fasting blood sugar level: a case of Debre Berhan Referral Hospital, Ethiopia. Am J Theor Appl Stat 2017;7:21-8.