Archives of Epidemiology
Twisk J, et al. Arch Epidemiol 7: 151.
www.doi.org/10.29011/2577-2252.100151.
www.gavinpublishers.com

GAVIN PUBLISHERS

Research Article

Multivariate Multilevel Analysis: An Elegant Alternative for the Separate Analysis of Correlated Outcomes

Jos Twisk[1*], Adriaan Hoogendoorn[2], Wieke de Vente[3]

[1]Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, Netherlands

[2]Department of Psychiatry, Amsterdam UMC Amsterdam, Amsterdam Public Health, Mental Health Program and Methodology Program, Amsterdam, Netherlands

[3]Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, Netherlands

***Corresponding author:** Jos Twisk, Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, Netherlands

Citation: Twisk J, Hoogendoorn A, de Vente W (2024) Multivariate Multilevel Analysis: An Elegant Alternative for the Separate Analysis of Correlated Outcomes. Arch Epidemiol 7: 151. DOI: 10.29011/2577-2252.100151

Received Date: 29 January 2024; **Accepted Date:** 13 February 2024; **Published Date:** 16 February 2024.

Abstract

Introduction: Surprisingly, multivariate multilevel analyses, which is often used to analyse longitudinal data, is seldom used to analyse multiple, correlated outcomes, such as questionnaire subscales or biological risk factors. Therefore, the aim of the study was to compare multivariate multilevel analysis with separate, per outcome, analyses in data with correlated outcomes retrieved from the same individual, in both cross sectional and longitudinal studies. **Method:** To compare the results of multivariate multilevel analysis with the separate analyses of each outcome, both real life example datasets and simulated data in which various characteristics are systematically varied, were used. **Results:** Multivariate multilevel analyses produced substantially more accurate estimates and standard errors than separate analyses for each outcome in incomplete datasets and when analysing longitudinal relations with time-dependent covariates. In complete datasets, for time-independent covariates (e.g., baseline characteristics) and when the development over time is analysed, results of the two analytical approaches were highly similar. It was also found that similarity of standard deviations between the outcomes is needed for a proper estimation of the standard error of the regression coefficient in a multivariate multilevel analysis. Similarity of standard deviations can be obtained by using z-scores. Furthermore, before using longitudinal multivariate multilevel analyses it has to be evaluated which type of clustering provides the best model fit. **Conclusion:** Multivariate multilevel analysis is an elegant way to analyse multiple correlated outcomes, which is superior to separate analyses for each outcome in case of missing data and/or when time-dependent covariates are analysed.

Keywords: Multivariate Multilevel Modelling; Cross-sectional data; Longitudinal data; Correlated outcomes; Real life data; Simulated data.

List of abbreviations

DASS	Depression Anxiety Stress Scales
AGHLS	Amsterdam Growth and Health Longitudinal Study
WHO-5	World Health Organization Well-Being Index-5
NOCDA	Netherlands Obsessive Compulsive Disorder Association study
OCD	Obsessive-compulsive disorder
BMI	Body mass index
MCAR	Missing completely at random
MAR	Missing at random
ICC	Intraclass correlation coefficient
BIC	Bayesian Information Criteria
ANOVA	Analyses of variance

Introduction

In epidemiology, patient-related outcomes such as quality of life, wellbeing, or psychological health are usually measured using questionnaires. Mostly, these constructs encompass various subscales of the questionnaires. For instance the WHO-5 Well Being Index consists of subscales for positive mood, vitality, and general interest [1]. and the Depression Anxiety Stress Scales (DASS) contains a depression, an anxiety and a stress subscale [2]. Researchers typically report the following when using questionnaires with subscales: 1) the results of the analysis with the total score, and 2) the results of separate analyses performed for the different subscales. However, according to statistical theory this analytical approach is problematic, since subscales scores are correlated. In other words, when analysing the different subscales of a questionnaire in separate analyses, it is ignored that the subscales are filled in by the same subject and thus that subscale scores are not independent of each other. The widespread use of this problematic analytical approach is surprising, because multivariate multilevel analysis is a relatively simple alternative in which the dependency of observations is taken into account. In multivariate multilevel analysis, subscales of a particular questionnaire can be analysed in a single model, taking into account the correlation between subscales measured in the same subject. Although examples of medical studies using multivariate multilevel analysis for subscales exist [3-5], in most medical studies the dependency of the outcomes is ignored. Therefore, the present paper illustrates the use of multivariate multilevel analysis to analyse multiple correlated outcomes.

It should be noted that the use of multivariate multilevel analysis is not limited to questionnaires with different subscales. There are many more examples of studies in which multiple outcomes are measured in the same subjects (e.g., pathology in multiple brain regions [6], recovery of different segments of the heart after myocardial infarction [7], etc.). Because these outcomes are not independent of each other, also for analysing these type of data, multivariate multilevel analyses is highly suitable (see one of the examples used in this paper).

Multilevel analysis has been specifically developed to deal with correlated observations in educational research to take into account the dependency of the observations made on students who are in the same class (or even in the same school) [8]. Nowadays, multilevel analysis is used in many different research areas. For instance, in social epidemiology, multilevel analysis is used to take into account the dependency of the observations on subjects living in the same neighbourhood [9]. In clinical studies, multilevel analysis is, for instance, used for multicentre studies to taken into account the dependency of the observations on patients from the same centre [10,11]. Apart from suitability for cross-sectional studies with correlated observations, multilevel analysis is also highly suitable for longitudinal studies in order to take into account the fact that the repeated observations within a subject are (highly) correlated [12]. Of note, the use of multilevel analysis for longitudinal data is actually the application of a multivariate multilevel analysis, as more than one outcome per subject is analysed in one model. In the literature, longitudinal data is always analysed as a multivariate outcome and separate analyses of the

different time-points are very rare. On the other hand, different outcomes from the same subject measured through for example subscales of a questionnaire are hardly ever analysed as a multivariate outcome. This is a rather peculiar phenomenon, because both longitudinal designs and designs with otherwise multiple outcomes per subject lead to highly correlated data, and this correlation should be taken into account in the analyses.

In sum, in the present paper the value of multivariate multilevel analyses for correlated outcomes was assessed by comparing multivariate multilevel analysis with separate analyses per outcome using both cross-sectional and longitudinal designs. Besides real life example datasets also simulated data was used in which various characteristics were systematically varied, such as the strength of the correlation between the multiple outcomes, the amount of missing data, and in longitudinal designs, the type of the covariate used in the analysis (i.e., time dependent versus time-independent).

Methods

In a multivariate multilevel analysis, a level for the different outcomes is present below the level of the subject (Figure 1). Repeatedly collecting multiple outcomes from one subject in a longitudinal study creates a three-level data structure. For this three-level structure basically two clustering options are possible: 1) The outcomes are clustered within the time-points and the time-points are clustered within the subjects or 2) The repeated measures are clustered within the outcomes and the outcomes are clustered within the subjects (Figure 2A & 2B).

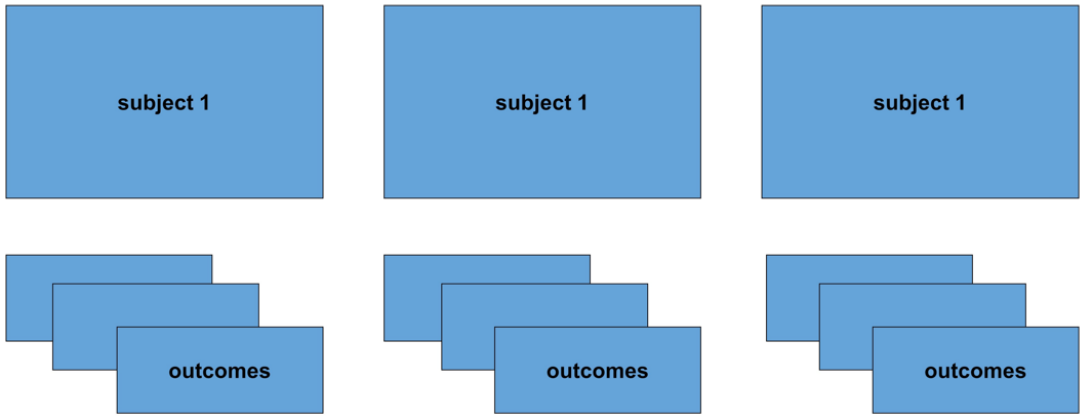

Figure 1: A two-level multivariate multilevel model. Outcomes are clustered within subjects.

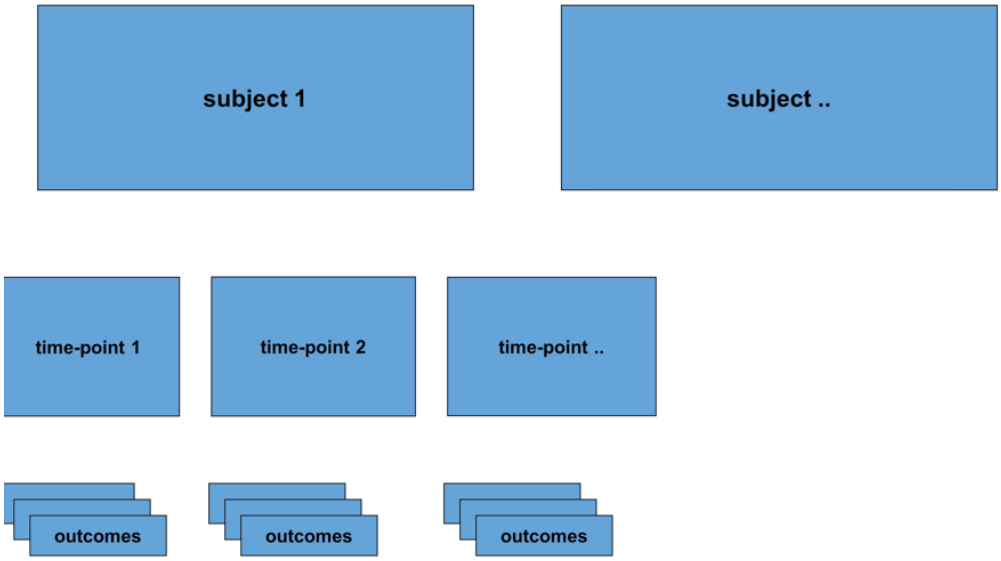

Figure 2A: A three-level multivariate longitudinal multilevel model. Outcomes are clustered within time-points, while the time-points are clustered within subjects.

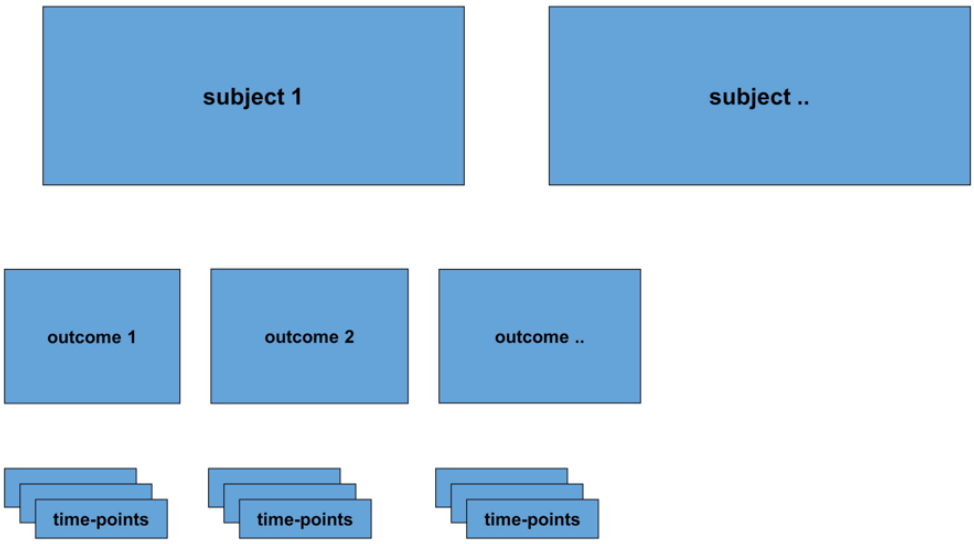

Figure 2B: A three-level multivariate longitudinal multilevel model. Time-points are clustered within outcomes, while outcomes are clustered within subjects.

Cross-sectional real example dataset

The cross-sectional example dataset was taken from a single measurement round of a longitudinal observational cohort study, i.e., the Amsterdam Growth and Health Longitudinal Study (AGHLS) [13]. The aim of this longitudinal study was to follow up the natural growth, health, and lifestyle in a representative sample of 698 Dutch adolescents [13]. In total, ten measurement rounds were performed between 1976 and 2006. The cross-sectional example dataset was taken from the last measurement round in which subjective psychological well-being was measured with the 5-item World Health Organization Well-Being Index (WHO-5). The WHO-5 is focused on subjective quality of life based on three subscales: 1) positive mood, 2) vitality, and 3) general interest [1]. In the example, the difference between males and females for the three different subscales of wellbeing was investigated. The multivariate cross-sectional

multilevel model contained two-levels, i.e., the three subscales of well-being were clustered within the subjects (Figure 1).

Longitudinal real example datasets

The first longitudinal example dataset was taken from the Netherlands Obsessive Compulsive Disorder Association (NOCDA) study. NOCDA is a longitudinal cohort study investigating the long-term course of obsessive-compulsive disorder (OCD) in patients referred to mental health care centres [14]. In this study, 419 patients were followed for a period of 6 years. Baseline measurements were performed between 2005 and 2009 and after that, three follow-up measurements were performed after 2, 4 and 6 years. In the present example, data from the three follow-up measurements were used on two related outcomes: depression and anxiety. The second longitudinal example dataset is taken from the last three measurement rounds of the AGHLS (see above) in which the correlated outcomes were three risk factors for cardiovascular disease (i.e., triglycerides, body mass index (BMI), and mean arterial blood pressure). Because the risk factors were measured on totally different scales, z-scores were used for the analyses (z-scores were taken for each outcome separately, across measurements, to allow for detection of both between subject and within subject [i.e., time] effects).

In both longitudinal example datasets, three analyses were performed. 1) the development over time, 2) the longitudinal relationship with a time-independent covariate (i.e., sex in both examples), and 3) the longitudinal relationship with a time-dependent covariate (i.e., symptom severity in the first example and the personality characteristic dominance in the second example). For the analyses, two multivariate longitudinal multilevel models were used, both containing three levels: 1) the outcomes were clustered within the three time-points, while the three time-points were clustered within the subjects, 2) the repeated measures were clustered within the outcome and the outcomes were clustered within the subject (see Figure 2a and 2b respectively). In all examples, the results of the multivariate multilevel analyses were compared to the results of the analyses for the outcomes separately. All analyses were performed with STATA (version 17).

Simulations

For each simulation scenario, 1000 samples were generated with a continuous normally distributed outcome variable. In the cross-sectional simulated dataset, three continuous outcome variables mimicked a questionnaire with three subscales. Furthermore, a dichotomous covariate was created equally divided over the subjects. In the cross-sectional analysis the relationship between the dichotomous covariate and the different subscales of the outcome questionnaire was analysed. This was done with 1) a multivariate cross-sectional multilevel analysis with two levels: subscales were clustered within the subject and 2) three linear regression analyses for the three subscales separately. Different scenarios were created: different standard deviations for the three subscales, different correlations between the three subscales and different number of subjects, which was either 100 or 200 (resulting in a total number of 300 and 600 observations, respectively). Lastly, several datasets with increasing amount of missing data (missing data percentages ranging between 10% and 40%) and different types of missing data (missing completely at random [MCAR] and missing at random [MAR]) in some of the subscales were simulated.

The longitudinal simulated dataset consisted of 100 subjects who were measured at three time-points. In this dataset, the continuous outcome variable were z-scores of three risk factors (i.e., comparable to the second longitudinal real example dataset, so z-scores were created for each outcome separately, across measurements). This led to a total number of 900 observations. Three analyses were performed: 1) the development over time, 2) the relationship with a time-independent covariate and 3) the relationship with a time-dependent covariate. All three were analysed with 1) a longitudinal multivariate multilevel analysis with three levels: outcomes clustered within the time-point and subsequently the repeated measures clustered within the subject 2) a longitudinal multivariate multilevel analysis with three levels: repeated measures clustered within the outcome and subsequently the outcomes clustered within subjects and 3) three separate longitudinal multilevel analyses with two levels: the three repeated measures clustered within the subject. The development over time and the relationships with the time-dependent and time-independent covariates were simulated to be different for the three outcomes. In the longitudinal simulations, different scenarios were created with different correlations between the outcomes and different correlations within an outcome over time (Table 1). Furthermore, several datasets with increasing amount of missing data at the repeated measures (i.e., 25% missing at the second measurement and 50% missing at the third measurement) and with different types of missing data (i.e., MCAR and MAR)) were simulated. All simulations were performed with STATA (version 17). For both the cross-sectional and longitudinal simulations for MCAR, missing data was randomly drawn from the whole population, while for MAR, missing data was randomly drawn from the lower 75% of the distribution of the outcome. It should be noted that the aim of the simulation study is only to compare the different methods with each other under changing circumstances.

	Correlation over time	Correlation between outcomes			
		1	2	3	4
Scenario 1	0.6	0.3	0.4	0.5	0.6
Scenario 2	0.6	0.4/0.3/0.2	0.5/0.4/0.3	0.6/0.5/0.4	0.7/0.6/0.5
Scenario 3	0.7/0.6/0.5	0.3	0.4	0.5	0.6
Scenario 4	0.7/0.6/0.5	0.4/0.3/0.2	0.5/0.4/0.3	0.6/0.5/0.4	0.7/0.6/0.5

Table 1: Different correlations used for the longitudinal simulations.

(Tables 2 and 3) show the data structure needed for both the cross-sectional and the longitudinal multivariate multilevel analysis. (Box 1 and Box 2) show examples of the STATA code for both the cross-sectional and longitudinal simulations.

Id	Subscale	Outcome	Covariate
1	1	7	5
1	2	3	5
1	3	10	5
2	1	5	7
2	2	4	7
2	3	8	7
..	..	..	..

Table 2: Data structure used for the cross-sectional multivariate multilevel analysis.

Id	Time	Subscale/ Risk factor	Outcome	Time-independent covariate	Time-dependent covariate
1	1	1	7	0	5
1	1	2	3	0	5
1	1	3	10	0	5
1	2	1	6	0	2
1	2	2	2	0	2
1	2	3	9	0	2
1	3	1	6	0	6
1	3	2	5	0	6
1	3	3	10	0	6
..	..	..	..	..	..

Table 3: Data structure used for the longitudinal multivariate multilevel analysis.

```
program test_1
drop _all
matrix c = (1, //
        0.70, 1, ///
        0.70, 0.70, 1, ///
        0.10, 0.20, 0.05, 1)
matrix m = (3.2,3.2,3.6,0.5)
matrix sd = (1,1,1,0.5)
drawnorm outcome1 outcome2 outcome3 sex, n(100) corr(c) cstorage(lower) means(m) sds(sd)
gen id = _n
reshape long outcome, i(id) j(subscale)
replace sex=0 if sex <0.5
replace sex=1 if sex >0.5
mixed outcome i.sex##i.subscale ||id:
end
simulate _b _se, seed(54321) reps(1000): test_1
```

Box 1: Example of the STATA code for one of the cross-sectional simulations.

```
program test_2
drop _all
matrix c = ///
  (1, ///
   0.70, 1, ///
   0.70, 0.70, 1, ///
   0.70, 0.60, 0.50, 1, ///
   0.60, 0.70, 0.60, 0.60, 1, ///
   0.50, 0.60, 0.70, 0.60, 0.60, 1, ///
   0.70, 0.60, 0.50, 0.70, 0.60, 0.50, 1, ///
   0.60, 0.70, 0.60, 0.50, 0.70, 0.60, 0.50, 1, ///
   0.50, 0.60, 0.70, 0.50, 0.60, 0.70, 0.50, 0.50, 1, ///
   0.20, 0.20, 0.20, 0.10, 0.10, 0.10, 0.15, 0.15, 0.15, 1, ///
   0.20, 0.20, 0.20, 0.10, 0.10,0.10, 0.15, 0.15, 0.15, 0.70, 1, ///
   0.20, 0.20, 0.20, 0.10, 0.10, 0.10, 0.15, 0.15, 0.15, 0.70, 0.70, 1, ///
  -0.10, -0.10, -0.10, -0.02, -0.02, -0.02, -0.05, -0.05, -0.05, 0, 0 , 0, 1)
matrix m = (-0.10,0.11,0.00,-0.20,-0.04,0.22,-0.08,0.06,0.04, 0, 0, 0, 0.5)
matrix sd = (0.90,1.09,1.02,0.90,0.98,1.10,0.91,1.11,0.98,1,1,1,0.5)
drawnorm outcome11 outcome21 outcome31 outcome12 outcome22 outcome32 outcome13 outcome23
outcome33 ind1 ind2 ind3 sex, n(100) corr(c) cstorage(lower) means(m) sds(sd)
gen id = _n
reshape long outcome1 outcome2 outcome3 , i(id) j(factor)
gen id2 = id * 10 + factor
reshape long outcome ind, i(id2) j(time)
replace sex=0 if sex <0.5
replace sex=1 if sex >0.5
mixed outcome c.ind##i.factor ||id: || time:
end
simulate _b _se, seed(54321) reps(1000): test_2
```

Box 2: Example of the STATA code for one of the longitudinal simulations.

Results

Cross-sectional example dataset

(Table 4) shows descriptive information regarding the cross-sectional example dataset regarding subjective psychological well-being in the 2006 measurement of the AGHLS study.

Wellbeing	Males (N=163)	Females (N=180)
Positive mood	3.32 (0.97)	3.04 (1.02)
Vitality	3.30 (0.97)	3.02 (1.03)
General interest	3.55 (0.96)	3.53 (1.03)

Table 4: Mean and standard deviation ( ) of the three subscales of wellbeing for males and females

(Table 5) shows the results of the analyses related to the differences between males and females in three subscales of wellbeing. It should be noted that within multilevel analysis, the dependency of the observations is usually quantified by the intraclass correlation coefficient (ICC) [15]. In the cross-sectional multivariate multilevel analysis, there is one ICC (i.e.,on subject level), which reflects the average correlation between the outcomes within a subject.

	Separate regression analysis	Multivariate multilevel analysis
Positive mood	-0.27 (0.11)	-0.27 (0.11)
Vitality	-0.29 (0.11)	-0.29 (0.11)
General interest	-0.01 (0.11)	-0.01 (0.11)
ICC		0.58

Table 5: Results* of the analyses related to the difference between females and males in three subscales of wellbeing; ICC = intraclass correlation coefficient, *regression coefficients and standard errors ( )

The results show that the regression coefficients indicating the difference between males and females as well as the standard errors of the regression coefficients were exactly the same for the three separate regression analyses and the multivariate multilevel analysis. This, even though, the ICC on subject level was relatively high.

Longitudinal example datasets

(Table 6) shows descriptive information of the longitudinal example datasets, regarding the development of depression and anxiety across three measurements in the NOCDA study (Example 1) and the development of three risk factors for cardiovascular disease (triglycerides, BMI and blood pressure) across three measurements in the AGHLS study (Example 2). In a longitudinal multivariate dataset, there are two ICCs. One reflecting the average correlation between the outcomes within the subject per time-point (i.e., the ICC over outcomes), and one reflecting the average correlation between the measurements over time (i.e., the ICC over time).

Example 1	Measurement 1 (n=399)	Measurement 2 (n=275)	Measurement 3 (n=272)	ICC over time
Depression	17.30 (11.97)	13.43 (11.22)	13.62 (10.86)	0.61
Anxiety	15.31 (10.09)	11.59 (10.10)	11.57 (9.77)	0.62
ICC over outcomes	0.65	0.67	0.65	
Example 2	**Measurement 1 (n=437)**	**Measurement 2 (n=378)**	**Measurement 3 (n=340)**	**ICC over time**
Triglycerides	1.09 (0.70)	1.27 (0.86)	1.17 (0.80)	0.60
BMI	23.29 (2.86)	24.07 (3.13)	24.62 (3.51)	0.84
Mean blood pressure	99.19 (9.02)	100.63 (11.03)	100.39 (9.72)	0.58
ICC over outcomes*	0.31	0.35	0.27	

Table 6: Mean and standard deviation ( ) of the outcomes of the two longitudinal example datasets at the three follow-up measurements; ICC = intraclass correlation coefficient; *ICC based on z-scores across measurements.

From Table 6 it can be seen that in both datasets the number of subjects decreased over time, which is typically seen in longitudinal studies. Regarding the development over time, in the first example dataset there was a sharp decrease from measurement 1 to measurement 2 in both outcomes, while in the second dataset there were only relatively small changes over time. It can also be seen that in the first longitudinal example dataset the ICCs over outcomes within a time-point were slightly higher than the ICCs over time, while in the second longitudinal example dataset, the ICCs over outcomes were much lower than the ICCs over time.

Before performing the various analyses, it should be evaluated which of the two longitudinal multivariate multilevel models should be used in the examples by assessing which model best fits the data. Therefore, for both examples, an intercept only model was analysed. (Table 7) shows the ICCs and the Bayesian Information Criteria (BIC) obtained from the different analyses on the two longitudinal example datasets. The BIC is a model fit indicator which is often used to compare models with each other. A lower BIC value indicates a better model fit [16]. Regarding the first longitudinal multivariate multilevel analysis, the ICC on time level refers to the correlation between the outcomes at a particular time-point, while the ICC on subject level refers to the correlation between the repeated measures of an outcome within a subject. Regarding the second longitudinal multivariate multilevel analysis, the ICC on outcome level refers to the correlation between the repeated measures within an outcome, while the ICC on subject level refers to the correlation between the outcomes within the subject. From the BIC values, it is obvious that for example 1, the first longitudinal multivariate multilevel model fitted best, while for example 2, the second longitudinal multivariate multilevel model fitted best. In the remaining part of the analyses on the example datasets, these best fitting models were used (Table 8).

Example 1		**Example 2**	
Model 1 (Figure 2a)	**Model 2 (Figure 2b)**	**Model 1 (Figure 2a)**	**Model 2 (Figure 2b)**
subject-level: 0.62 time-level: 0.29 BIC: 13569	subject-level: 0.60 outcome-level: 0.16 BIC: 13607	subject-level: 0.38 time-level: 0.00 BIC: 8981	subject-level: 0.48 outcome-level: 0.54 BIC: 8268

Table 7: Intraclass correlation coefficients and model fit of the different multivariate multilevel analysis.

Example 1		**Separate analysis**	**Multivariate analysis**
Depression	Time-point 1 Time-point 2	-3.28 (0.58) -3.34 (0.59)	-3.29 (0.58) -3.42 (0.58)
Anxiety	Time-point 1 Time-point 2	-3.23 (0.50) -3.56 (0.50)	-3.15 (0.58) -3.42 (0.58)
Example 2		**Separate analysis**	**Multivariate analysis**
Triglycerides	Time-point 1 Time-point 2	0.20 (0.045) 0.10 (0.047)	0.19 (0.040) 0.10 (0.042)
BMI	Time-point 1 Time-point 2	0.20 (0.026) 0.38 (0.027)	0.20 (0.040) 0.38 (0.042)
Mean blood pressure	Time-point 1 Time-point 2	0.10 (0.047) 0.07 (0.049)	0.09 (0.040) 0.06 (0.040)

Table 8: Results* of the analyses related to the development over time in different outcomes;*regression coefficients and standard errors ( )

The results of the different analyses investigating the development over time in the two outcomes (depression and anxiety) of the first longitudinal example dataset show that there was a small difference in both the estimated regression coefficients, and the standard errors between the longitudinal multivariate multilevel model and the two separate longitudinal multilevel models. For the second longitudinal example dataset, again small differences in estimated regression coefficients were found. The standard errors obtained from the longitudinal multivariate multilevel analysis on the other hand were lower than the one obtained from the separate analyses for triglycerides and mean blood pressure, but higher than the one obtained from the separate analysis for BMI.

(Table 9) shows the results of the different analyses investigating the difference in outcomes between males and females on average over time, and (Table 10) shows the results of the different analyses investigating the relationship between symptom severity (in the first longitudinal example dataset) and dominance (in the second longitudinal example dataset) and the outcomes on average over time.

Example 1	**Separate analysis**	**Multivariate analysis**
Depression	3.01 (1.04)	2.86 (0.95)
Anxiety	1.93 (0.91)	2.17 (0.95)
Example 2	**Separate analysis**	**Multivariate analysis**
Triglycerides	-0.57 (0.08)	-0.57 (0.08)
BMI	-0.37 (0.09)	-0.36 (0.08)
Mean blood pressure	-0.60 (0.08)	-0.59 (0.08)

Table 9: Results* of the analyses related to the difference between females and males in the different outcomes on average over time; *regression coefficients and standard errors ( ).

The results of the analyses regarding the difference in the outcomes between males and females and the relationship with a time-dependent covariate (symptom severity and dominance) was more or less comparable for both datasets, that is, a small difference in both regression coefficients and standard errors in a non-systematic way. However, in the second example dataset there much bigger difference regarding the analyses with the time-dependent covariate.

Simulations

Cross-sectional simulations

The regression coefficients obtained from the cross-sectional multivariate multilevel analysis and the three separate regression analyses were exactly the same in all simulations on datasets without missing data. However, when the standard deviations differed between the different subscales, the standard errors of the regression coefficients differed remarkably (Tables S1, S2 and S3). From Table S1, it can be seen that different correlations between the subscales had no influence on the differences between the results of the different analyses.

From Tables S2 and S3 it can be seen that the difference between the estimated standard errors increased when the difference between the standard deviations of the different subscales increased and that the estimated standard errors were equal when the standard deviations of the different subscales were equal. Furthermore, it can be seen that the difference between the estimated standard errors was bigger when sample size decreased and that the standard error obtained from the cross-sectional multivariate multilevel analysis is a sort of mean of the standard errors obtained from the separate regression analyses.

Missing data in some of the subscales

(Tables S4 and S5) show the results of the simulation study comparing the cross-sectional multivariate multilevel analysis and the separate regression analyses for the three subscales with different amounts of missing data in some of the subscales ranging from 10% to 40%. Table S4 shows the results of the analyses when missing data was completely at random (MCAR), while table S5 shows the results for the analyses when missing data was at random (MAR).

When missing data was MCAR, the regression coefficients of all analyses were more or less the same. When missing data was MAR, on the other hand, the regression coefficients were different for the two analyses and were slightly in favour for the multivariate multilevel analyses, when compared to the regression coefficients obtained from the analyses on the complete dataset. As expected, the differences between the methods increased when the amount of missing data increased. Besides this, the standard errors were in general a bit higher in the separate regression analyses compared to the multivariate multilevel analyses.

Longitudinal simulations

Before performing the different analyses, comparable to what has been done for the example longitudinal datasets, it had to

be evaluated which of the two longitudinal multivariate multilevel models should be used for the different scenarios. Therefore, intercept only models were analysed for all the scenarios shown in Table 1. Figure 3 shows the BIC values for the analyses of all the different simulated scenarios.

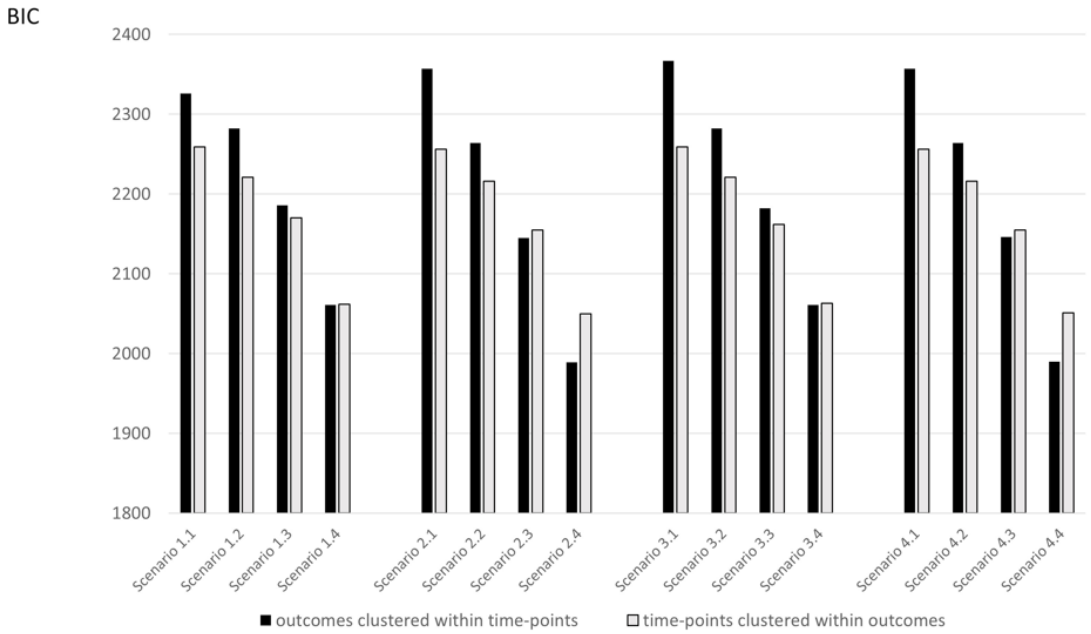

Figure 3: BIC values for intercept only models analysed for the different scenarios shown in Table 1.

From Figure 3 it can be seen that with increasing correlation within an outcome, the longitudinal multivariate multilevel model in which outcomes were clustered within time-points showed a better fit compared to the longitudinal multivariate multilevel model in which repeated measures were clustered within the outcome. For all longitudinal simulations the best fitting model was used for each scenario.

(Tables S6 to S9) show the results of the longitudinal simulation study regarding the development over time. The results show that for the complete dataset and the dataset with MCAR, the regression coefficients of the multivariate longitudinal multilevel analyses and the three separate longitudinal multilevel analyses were identical regardless the scenario used for the simulations. For the MAR datasets, the regression coefficients were slightly different compared to the ones estimated in the complete dataset and also slightly different between the longitudinal multivariate multilevel analyses and the three separate longitudinal multilevel analyses.

(Tables S10 to S13) show the results of the longitudinal simulation study regarding the relationship with a time independent covariate. For the complete dataset, the regression coefficients were equal when estimated with multivariate longitudinal multilevel analyses or with separate longitudinal multilevel analyses. Within the MCAR and MAR datasets, there were small differences in both regression coefficients and standard errors, which were not in favour of one the two methods.

The results of the longitudinal simulation study regarding the relationship with a time dependent covariate (Tables S14 to S17) show a different picture. In this situation, there was a remarkable difference between the regression coefficients obtained from the multivariate longitudinal multilevel analyses and the three separate longitudinal multilevel analyses. This was already the case in the complete dataset. Regarding this, it can be seen that for the multivariate longitudinal multilevel analysis, scenario 1 and 3 led to comparable results as well as scenario 2 and 4.

In other words, the regression coefficients were mostly related to the difference in correlation between the outcomes. For the separate longitudinal multilevel analyses, the results of scenario 1 and 2 and the results of scenarios 3 and 4 led to comparable results, which indicates that the regression coefficients of the separate longitudinal multilevel analyses were related to the difference in correlation over time between the three outcomes. It can also be seen that for all scenarios, changing correlations within a scenario led to different regression coefficients of the multivariate

longitudinal multilevel analyses, while this has no influence on the regression coefficients obtained from the three separate longitudinal multilevel analyses. For the MCAR and MAR datasets, the same differences between the two methods were found and in general, the longitudinal multivariate multilevel analyses seem to perform slightly better than the separate longitudinal multilevel analyses when the regression coefficients were compared to the ones obtained from the complete dataset. Regarding the standard errors, in general, for all scenarios the standard errors obtained from the longitudinal multivariate multilevel analyses were slightly lower than the ones obtained from the separate longitudinal multilevel analyses.

Discussion

The goal of this paper was to investigate multivariate multilevel analysis as a statistical approach to analyse correlated outcomes, such as questionnaire subscales or biological risk factors retrieved from the same subject. The results showed that although multivariate multilevel analysis is theoretically a very elegant solution to analyse correlated outcomes, in many situations it does not lead to different results than the separate analyses of the various correlated outcomes. For cross-sectional studies, only when there is missing data in some of the outcomes, the multivariate multilevel analysis gives more accurate results (i.e., results are closer to the results obtained from the complete dataset). For longitudinal studies, multivariate multilevel analysis is preferred above separate analyses per outcome when a) there are missing data at different time-points and/or b) the relationship with a time-dependent covariate is estimated.

From the cross-sectional analyses performed in the present paper (both on the real life examples and the simulations) two important differences between the multivariate analysis and the commonly applied separate analyses for each outcome were found. First, there is a difference in estimated standard errors between the two methods. When the standard deviations differ between the outcomes, in the cross-sectional multivariate multilevel analysis a pooled standard error is calculated over the different outcomes. Because of that, in the multivariate multilevel analysis, the regression coefficients for the different outcomes all have the same standard error. For the separate analyses for the different outcomes, sometimes a higher standard error is found (for the outcomes with a higher standard deviation) and sometimes a lower standard error (for the outcomes with a lower standard deviation). When the standard deviations are equal for the different outcomes (for instance by using z-scores), and there are no missing data, the standard errors of the cross-sectional multivariate multilevel analysis and the separate analyses for the different outcomes are exactly the same. Second, when there is missing data for some of the outcomes and the missing data is MAR, multivariate multilevel analysis is preferred above separate analyses for the different outcomes. This may not be surprising because it is known that multilevel analysis better deals with missing data than complete case analyses [17], which is basically the type of analysis done when the different outcomes are analysed separately.

The results of the comparison between the longitudinal multivariate multilevel analysis and the separate longitudinal multilevel analyses differed from the findings in the cross-sectional situation. First, in line with a simulation study addressing multivariate longitudinal multilevel analyses of Baldwin et al. [18], it was found that the effect estimates and standard errors did not differ between multivariate longitudinal multilevel analysis and separate longitudinal multilevel analyses under the following conditions: a) the development over time is analysed and/or the relationship with a time-independent covariate is analysed and b) a complete dataset was analysed (i.e., no missing values in the outcomes). However, results differed between the two analytical approaches when these conditions were not met. To illustrate, in both longitudinal real example datasets (with missing values), the regression coefficients obtained from the two methods differed from each other. In the simulation studies, it appeared that these differences in regression coefficients between statistical approaches did not only occur when datasets consisted missing values. In the simulations, it was shown that the regression coefficients differed between the two methods for the models with a time-dependent covariate, irrespective of missing data.

The finding that, even in a complete dataset, the regression coefficients for a time-dependent covariate and the standard errors differed substantially between a multivariate longitudinal multilevel analysis and the separate longitudinal multilevel analyses is probably the most interesting. Due to the fact that multivariate multilevel analysis takes into account the correlation between the outcomes, this method provides more appropriate estimates than separate multilevel analyses. Therefore, multivariate multilevel analysis should be preferred above the separate longitudinal multilevel analyses when time-dependent covariates are analysed in a longitudinal study. The reason why no or only small differences in coefficients were observed in the analyses regarding the development over time and the analyses with a time-independent covariate is because time is not different between subjects and the time-independent covariate is not different between time-points. Because of that, the regression coefficient of the multivariate longitudinal multilevel analysis with complete data was not different from the regression coefficients obtained from the separate longitudinal multilevel analysis. This phenomenon is well known within multilevel analysis [15].

It should be realized that in the present paper, in all examples linear multivariate multilevel analysis was used, assuming a normal distribution of the residuals. It is, however, possible that the residuals are not normally distributed, due to a non-normal

distribution of the outcome variable(s). In that case a different multivariate multilevel analysis can be used. When, for instance, an outcome variable with floor or ceiling effects is analysed, a multivariate to bit multilevel analysis can be used. When the outcome variable is a count outcome, multivariate Poisson multilevel analysis or multivariate negative binomial multilevel analysis can be used [12,15]. In all situations, however, it is expected that the comparison of the results between the different methods will be more or less the same.

Various practical implications arise from this study. First, it should be noted that using longitudinal multivariate multilevel modelling, coefficients depend on the way the clustering of the data is modelled. It is therefore important to first evaluate which clustering (see Figures 2a and b) of the data results in the best model fit, accounting optimally for the dependency of the outcomes, before the actual analyses are conducted. As shown in the simulation study, the choice for a particular model depends on the magnitude of the correlation between the outcomes. When these correlations are relatively low (as in the second example longitudinal dataset), a model is preferred in which the repeated measures are clustered within the outcomes. When the correlations between the outcomes are relatively strong (as in the first example longitudinal dataset) a model is preferred in which the outcomes are clustered within the time-points. Second, it is advised to standardize multiple outcomes (i.e., calculate z-scores per outcome across all measurements) when doing multivariate multilevel analysis, in particular when standard deviations differ between outcomes.

It was shown that when standard deviations differ, a pooled standard deviation is used for the estimation of the standard error of the regression coefficient for each outcome in the multivariate multilevel analysis, which is an unfavorable situation, because it results in imprecise test results. Thirdly, multivariate multilevel analyses may also serve the need for a type of conservative statistical testing, which is commonly, applied in for example the field of psychology. In this field, where analyses of variance (ANOVA) is often used for testing hypotheses, the multivariate test result of a multivariate ANOVA is taken as a first step in the analyses of multiple outcomes (such as questionnaire subscales). In case of non-significance (i.e., no significant group differences, or otherwise, no significant association between the independent variable and the multivariate outcome), no further analyses are performed and it is concluded that the independent variable is not related to any of the outcomes. If the multivariate result is statistically significant, though, the univariate results (from the same analysis) are reported and conclusions are drawn regarding to which outcome the independent variable is related to. While we are convinced that this procedure is inadequate, because it is totally driven by test theory and not by the aim of effect estimation, we feel that we need to highlight the suitability of multivariate multilevel analysis for this analytic strategy for researchers in fields where this approach is still customary. For researchers in epidemiology, where the focus has shifted the past decades from statistical testing to the estimation of effects and confidence intervals, this two-step analytical approach is explicitly not promoted. The right way from the perspective of effect estimation is to report the univariate results, irrespective of the statistical significance of the multivariate result.

Conclusions

In this paper, it was shown that multivariate multilevel analyses, which is commonly used in longitudinal data analyses using one outcome variable measured at different time-points, can also be easily applied when analysing multiple outcomes, such as questionnaire subscales or biological risk factors. While theoretically superior to separate analyses per outcome, multivariate multilevel analysis often produce highly similar results as separate analyses for each outcome. However, results differ increasingly with an increasing amount of missing data and in longitudinal models with time-dependent covariates. Points of attention when conducting the multivariate multilevel analyses are similarity of standard deviations between the outcomes which can be obtained by using z-scores and evaluating the type of clustering for longitudinal models. As the technique is relatively easy to conduct, there is no reason why researchers would not use multivariate multilevel analysis when analysing correlated outcomes in both cross-sectional and longitudinal studies.

Acknowledgements

Not applicable.

Funding

None

Ethics guidelines

Not applicable.

Conflict of interests

The authors declare that they have no conflict of interests.

References

1. Topp CW, Østergaard SD, Søndergaard S, Bech P (2015) The WHO-5 Well-Being Index: A Systematic Review of the Literature. Psychother Psychosom 84: 167-176.
2. Lovibond P, Lovibond S (1995) The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 33: 335-343.

3. Subramanian SV, Kim D, Kawachi I (2005) Covariation in the socioeconomic determinants of self-rated health and happiness: a multivariate multilevel analysis of individuals and communities in the USA. J Epidemiol Community Health 59: 664-669.

4. Kiwanuka HM, Van Damme J, Van Den Noortgate W, Anumendem DN, VanLaar G, et al.(2016) How do student and classroom characteristics affect attitude toward mathematics. A multivariate multilevel analysis. School Effectiveness and School Improvement 28: 1-21.

5. Dessie ZG, Zewatir T, Mwambi H, North D (2020) Multivariate multilevel modelling of quality of life dynamics of HIV infected patients. Health Qual Life Outcomes 18: 80.

6. Hart de Ruyter FJ, Morrema THJ, den Haan J; Netherlands Brain Bank; Twisk JWR, et al. (2023) Correction to: Phosphorylated tau in the retina correlates with tau pathology in the brain in Alzheimer's disease and primary tauopathies. Acta Neuropathol 145: 197-218.

7. Beek AM, Kühl HP, Bondarenko O, Twisk JW, Hofman MB, et al. Delayed contrast-enhanced magnetic resonance imaging for the prediction of regional functional improvement after acute myocardial infarction. J Am Coll Cardiol 42: 895-901.

8. Goldstein H (1987) Multilevel models in educational and social research. London, Griffin; New York, Oxford University Press.

9. Schüle SA, von Kries R, Fromme H, Bolte G (2016) Neighbourhood socioeconomic context, individual socioeconomic position, and overweight in young children: a multilevel study in a large German city. BMC Obes 3: 25.

10. Rice N, Leyland A (1996) Multilevel models: applications to health data. J Health Serv Res Policy 1: 154-164.

11. Merlo J, Östergren PO, Broms K, Bjorck-Linné A, Liedholm H (2001) Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. J Epidemiol Community Health 55: 323-329.

12. Twisk JWR (2023) Applied longitudinal data analysis for medical; science. A practical guide. 3rd edition. Cambridge UK, Cambridge University Press.

13. Wijnstok NJ, Hoekstra T, van Mechelen W, Kemper HC, Twisk JW (2013) Cohort profile: the Amsterdam growth and health longitudinal study. Int J Epidemiol 42: 422-429.

14. du Mortier JAM, Remmerswaal KCP, Batelaan NM, Visser HAD, Twisk JWR, et al. (2021) Predictors of Intensive Treatment in Patients With Obsessive-Compulsive Disorder. Front Psychiatry 12: 1-659401.

15. Twisk JWR. Applied mixed model analysis. A practical guide. Cambridge UK, Cambridge University Press, 2019.

16. Neath AA, Cavanaugh JE (2012) The Bayesian information criterion: background, derivation, and applications. WIREs Computational Statistics 4: 199-203.

17. Twisk J, de Boer M, de Vente W, Heymans M (2013) Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J Clin Epidemiol 66: 1022-1028.

18. Baldwin SA, Imel ZE, Braithwaite SR, Atkins DC (2014) Analyzing multiple outcomes in clinical research using multivariate multilevel models. J Consult Clin Psychol 82: 920-930.