



Using multilevel linear growth models to examine participant performance on different cardiorespiratory fitness assessments

Peter D Hart

Health Promotion Research, Havre, Montana, United States

Abstract

Background: The ability of a cardiorespiratory fitness (CRF) assessment to differentiate individuals is a valuable characteristic. Little research exists, however, regarding the extent to which individual variation in CRF differs in the same magnitude across different CRF tests.

Purpose: The purpose of this study was to use multilevel linear growth models to examine the inter-individual variation of within-individual variation of CRF field test scores.

Methods: Data for this research came from N=131 college students attending a rural public university. Four (4) CRF field assessments were administered to each participant in random order with each yielding estimated maximal oxygen consumption (VO₂max, ml/kg/min). Random intercept and random slope multilevel growth models were evaluated with CRF tests (level 1) nested within participants (level 2). CRF tests were coded 0 = treadmill, 1 = step, 2 = non-exercise, and 3 = beep and sex was coded 1 = male and 0 = female.

Results: Results from the unconditional means model justified the multilevel analysis of the CRF data (ICC = .19, $p < .001$). A fixed slope for CRF test ($b = -4.95$, $p < .001$) was significantly related to CRF scores but did not show significant random variation. A final random intercept model was selected with significant level 2 predictors of age ($b = -0.33$, $p < .001$) and sex ($b = 12.59$, $p < .001$), significant level 1 predictor of CRF test ($b = -3.41$, $p < .001$), and significant cross-level interaction predictor of sex-by-test ($b = -2.32$, $p < .001$). The final model explained 42% and 70% of level 1 and level 2 variance, respectively.

Conclusion: Results from this study show that CRF field assessments suffer from considerable variation. Sex-specific inter-individual variation showed consistent within-person variation, with males displaying a slightly steeper linear trajectory across CRF tests.

Keywords: multilevel analysis, growth curve, cardiorespiratory fitness, measurement

Introduction

Multilevel linear growth curve modeling is a modern statistical analysis of repeated measure designs and can examine inter-individual variation of within-individual variation^[1]. This is generally accomplished by modeling the hierarchical (correlated) nesting of repeated measurements within individuals. In an equivalence reliability study, participants are assessed repeatedly across different tests, therefore, different tests are nested within participants. Multilevel linear growth curve models allow for the examination of four (4) main questions^[2]. First, is there significant variation around the overall mean CRF score in the population? Second, do CRF scores systematically change across repeated trials (tests)? Third, do CRF scores vary in their growth trajectories (systematic change) across repeated trials. Fourth, can variation in participant CRF scores or growth trajectories be explained by level 2 (participant-level) and/or level 1 (test-level) predictors, respectively. The purpose of this study is to address these four questions and evaluate the equivalence reliability of CRF field tests administered to college students.

Methods

Study Variables

Participants for this study were college students enrolled at a rural public university. Four different fitness tests were administered for the Cardiorespiratory (CR) trait. These

included (1) Multi-stage fitness (Beep) test^[3], (2) Queens College step (Step) test^[4], (3) Ebbeling VO₂max Tread Mill (TM) test^[5], and (4) George Non-Exercise (NE) test^[6]. CRF tests were administered in random order across a period of one week. Each test allowed for the estimation of maximal oxygen consumption (VO₂max) in the common measured units of ml/kg/min. Test coding was assigned based on descriptive statistics as follows: TM = 0, Step = 1, NE = 2, and Beep = 3. Participant sex was used as a model factor and coded as male = 1 and female = 0. Finally, participant age was used in this study as a covariate.

Statistical Analysis

Descriptive statistics were computed for study variables both overall and sex-specifically. Multilevel linear growth modeling was used to address four main questions using four different models. Model 1 is the unconditional means model and used to determine if there was significant variation around the overall CRF mean. From this model, the intraclass correlation coefficient (ICC) was computed as measures of intercept variance and independence violation. The ICC was computed as: $ICC = \tau_{00} / (\tau_{00} + \sigma^2)$, where τ_{00} is intercept variance and σ^2 is residual variance. Model 2 is the linear growth model with fixed slope and used to determine if systematically change is present across repeated trials. Model 3 is model 2 but also estimates slope variance to determine if CRF scores vary in their growth trajectories

across repeated trials. Model 4 is model 3 with random slope removed and added person-level (level 2) time-invariant covariates to determine if more intercept variance can be explained. From this model, the proportion reduction in variance (PRV) was computed at each level. Specifically, PRV.L1 is proportion reduction in model 1 level 1 residual variance by the final model and PRV.L2 is proportion reduction in model 2 level 2 intercept variance by the final model. The PRV is important because residual variance represents the variability of an individual around his or her mean and intercept variance represents differences between individuals. SAS PROC Mixed was used for the multilevel modeling using maximum likelihood estimation, unstructured covariances, and Satterthwaite degrees of freedom for coefficient t tests [7].

Results

Table 1 displays simple statistics on the CRF variables and age by sex. Significant (ps < .001) sex differences were seen across all CRF outcomes. Figure 1 displays trends in linear CRF growth across the four CRF tests for each participant. The curves indicate large variation among the CRF intercepts. Variation in linear slopes across participants is less apparent from the graph. Table 2 contains the estimates from the two-level linear growth models examining CRF growth across different CRF tests. Model 1 is the unconditional means model with overall fixed grand mean of 45.6 ml/kg/min. Model 1 quantifies the amount of variation both between and within participants and confirms the need for multilevel modeling with a substantial between person variance (ICC = .19, p < .001). Model 2 is the linear growth model with fixed CRF test slope. Model 2 indicates a significant fixed linear trajectory (b = -4.95, p < .001), which also adjusts the multilevel model by reducing residual variance and increasing intercept variance. Model 3 is the linear growth model with fixed slope additionally estimating random slope variance. Model 3 is specifically examining between-person variance in within-person changes across different CRF tests. The random slopes model is, however, not supported, as seen by the non-significant slope variance, non-significant slope-intercept covariance, and increased AIC and BIC values relative to model 2.

Model 4 is therefore the conditional linear growth model with fixed CRF test slope and person-level (level 2) time-invariant covariates explaining intercept variance. Model 4 indicates that age (b = -0.33, p < .001), sex (b = 12.59, p < .001), CRF test (b = -3.41, p < .001), and the cross-level interaction of sex-by-test (b = -2.32, p < .001) significantly predict CRF scores. Furthermore, the level 2 predictors explained 70% of the level 2 (intercept) variance. Specifically, both lower age and being male increased predicted mean CRF scores, with males seeing an even steeper linear decline across tests. Figure 2 displays the sex-by-test interaction depicted by the final multilevel model. Figure 2 makes it easy to envision the difference in

predicted mean CRF scores by sex as well as the difference in fixed slopes between sex.

Discussion

The purpose of this study was to employ multilevel linear growth modeling to address four (4) questions regarding the equivalence reliability of CRF field tests. The first question, regarding significant variation around the overall mean CRF score, was addressed from the first multilevel model. Results clearly supported the need for multilevel modeling with a substantial between person variance. Specifically, almost 20% of the variance in CRF scores exists between participants. The second question, regarding the extent to which CRF scores systematically change across repeated trials of different tests, clearly showed a systematic bias. This finding was supported by the significant fixed linear trajectory of approximately -5 ml/kg/min stepwise, beginning with TM and ending with Beep. The third question, regarding the extent to which an individual’s CRF scores vary in their growth trajectories (systematic change) across repeated trials, clearly showed a lack of participant slope variation. This finding supports the fixed linear trajectory finding from the previous question by showing that the bias is consistent across all participants and hence of systematic nature. The fourth question, regarding the extent that variation in participant CRF scores could be explained by level 2 (participant-level) predictors, clearly showed that age, sex, and a sex-by-test interaction significantly and substantially reduced intercept variance. The largest effect responsible for this reduction in variance is sex differences in CRF scores. These findings are not uncommon, since males are known to have greater CRF than females and CRF declines with age [8]. The collective examination of these four (4) questions cast doubt on the equivalence reliability of CRF field tests administered to college students. Although these findings support the consistent ranking of participants across different CRF tests (i.e., random intercepts but not random slopes), the systematic bias across tests questions the equal difficulty requirement in equivalent tests [9].

Results from this multilevel linear growth curve analysis should be considered along with the limitations of the study. One limitation is the population from which the sample was drawn. This study was conducted on college students attending a relatively small and rural public university. Given this fact, the findings from this study may not necessarily generalize to larger universities. A second limitation concerning this study was the lack of level 1 predictors that could possibly explain within individual differences. For example, the Beep test is a test more commonly administered to school-aged children and requires high motivation in order to get valid CRF scores. College students may lack the same motivation inherent in school-aged children, which may explain its systematically lower scores.

Table 1: Descriptive statistics on study variables by sex.

Variable	Overall (n = 131)		Male (n = 87)		Female (n = 44)		t p
	Mean	SD	Mean	SD	Mean	SD	
Age (yr)	21.8	5.06	21.2	3.40	23.1	7.21	.106
TM (ml/kg/min)	50.8	10.04	54.7	8.54	43.3	8.40	<.001
Step (ml/kg/min)	49.0	10.73	53.8	9.71	39.3	4.02	<.001
NE (ml/kg/min)	47.7	7.36	50.0	6.55	43.2	6.86	<.001
Beep (ml/kg/min)	34.8	8.35	36.9	8.05	30.6	7.37	<.001

Note. t test is for sex differences.

Table 2: Estimates from two-level linear growth models examining CRF growth across different CRF tests (N = 131).

	Model 1		Model 2		Model 3		Model 4	
Fixed effects								
Intercept	45.569	(0.609)	52.994	(0.760)	52.994	(0.776)	51.872	(2.285)
Test			-4.951	(0.304)	-4.951	(0.280)	-3.411	(0.516)
Age							-0.332	(0.087)
Sex							12.588	(1.331)
Sex*Test							-2.319	(0.633)
Random effects								
Residual	101.360	(7.231)	60.514	(4.317)	65.066	(5.685)	58.515	(4.174)
Intercept	23.161	(6.260)	33.373	(6.089)	33.421	(10.538)	10.020	(3.219)
Test. Slope					-2.731	(1.705)		
Covariance					1.653	(3.398)		
Model Statistics								
ICC	.19							
PRV.L1							.42	
PRV.L2							.70	
AIC	3998.3		3797.6		3797.6		3701.7	
BIC	4006.9		3809.1		3814.9		3721.8	

Note. Entries show parameter estimates with standard errors in parentheses. Estimates in bold are significant (p<.05). Values based on SAS PROC Mixed. Estimation Method = ML; Covariance structure = unstructured. Satterthwaite degrees of freedom used in coefficient t tests. Test coding is TM = 0, Step = 1, NE = 2, and Beep = 3. ICC is intraclass correlation coefficient. PRV.L1 is proportion reduction in level 1 residual variance by final model. PRV.L2 is proportion reduction in level 2 intercept variance by final model. Model 1 is the unconditional means model. Model 2 is the linear growth model with fixed slope. Model 3 is model 2 also estimating slope variance. Model 4 is model 3 with random slope removed and added person-level (level 2) time-invariant covariates explaining intercept variance. Residual variance represents the variability of an individual around his or her mean. Intercept variance represents differences between individuals. Sex is coded 1=male and 0=female.

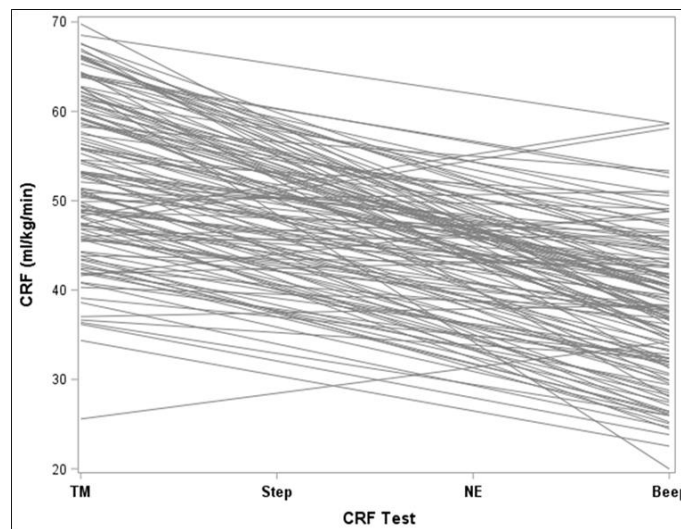


Fig 1: Plot of CRF scores regressed on CRF test by participant (N=131).

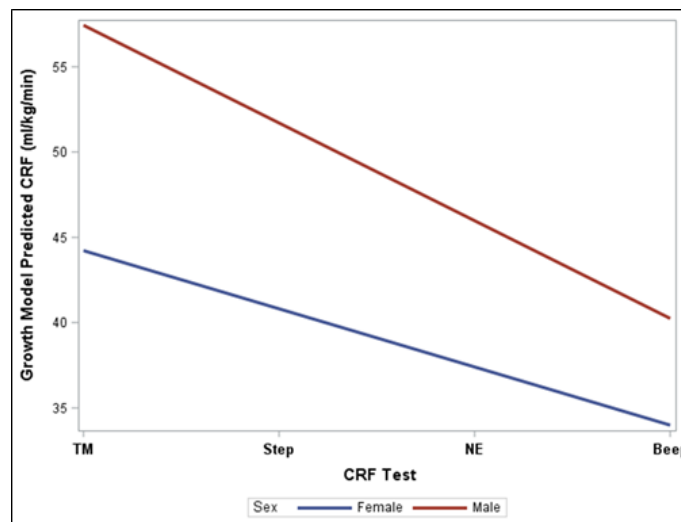


Fig 2: Plot of model predicted CRF scores regressed on CRF test by sex (N=131).

Conclusions

Results from this study support the consistent ranking of participants across different CRF tests (i.e., random intercepts but not random slopes). However, the systematic bias across CRF tests questions the equal difficulty requirement in equivalent tests. The sex-specific findings additionally showed males displaying a slightly steeper linear trajectory across CRF tests. These findings suggest further investigation into CRF field tests that are appropriate for college student populations.

References

1. Curran PJ, Obeidat K, Losardo D. Twelve frequently asked questions about growth curve modeling. *Journal of cognition and development*. 2010; 11(2):121-36.
2. Bell BA, Ene M, Smiley W, Schoeneberger JA. A multilevel model primer using SAS PROC MIXED. In *SAS global forum*. 2013; 433:1-19.
3. Ramsbottom R, Brewer J, Williams C. A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine*. 1988; 22(4):141-4.
4. McArdle WD, Katch FI, Katch VL. *Exercise physiology: nutrition, energy, and human performance*. Lippincott Williams & Wilkins, 2010.
5. Ebbeling CB, Ward A, Puleo EM, Widrick JE, Rippe JM. Development of a single-stage submaximal treadmill walking test. *Medicine and science in sports and exercise*. 1991; 23(8):966-73.
6. George JD, Stone WJ, Burkett LN. Non-exercise VO₂max estimation for physically active college students. *Medicine and science in sports and exercise*. 1997; 29(3):415-23.
7. Wang J, Xie H, Fisher JF. *Multilevel models: applications using SAS®*. Walter de Gruyter; 2011, 23.
8. Raven P, Wasserman D, Squires W, Murray T. *Exercise Physiology*. Nelson Education, 2012.
9. Price LR. *Psychometric methods: Theory into practice*. Guilford Publications, 2016, 12.