Comparing individual and population differences in minute ventilation/carbon dioxide production slopes using centile growth curves and log-linear allometry

Alan M. Nevill[1], Jonathan Myers[2,3], Leonard A. Kaminsky [4], Ross Arena[5] and Tony D. Myers[6]

[1]Faculty of Education, Health and Wellbeing, University of Wolverhampton, Walsall, UK. [2]Division of Cardiology, VA Palo Alto Healthcare System, Palo Alto, CA, USA. [3]Stanford University, Stanford, CA, USA. [4]Fisher Institute of Health and Well-Being and Clinical Exercise Physiology Laboratory, Ball State University, Muncie, IN, USA. [5]Dept of Physical Therapy, College of Applied Sciences, University of Illinois at Chicago, Chicago, IL, USA. [6]Sport and Health, Newman University, Birmingham, UK.

Corresponding author: Alan M. Nevill (a.m.nevill@wlv.ac.uk)

Abstract
Identifying vulnerable groups and/or individuals' cardiorespiratory fitness (CRF) is an important challenge for clinicians/researchers alike. To quantify CRF accurately, the assessment of several variables is now standard practice including maximal oxygen uptake (V'_{CO_2}) and ventilatory efficiency, the latter assessed using the minute ventilation/carbon dioxide production (V'_E/V'_{CO_2}) slope. Recently, reference values (centiles) for V'_E/V'_{CO_2} slopes for males and females aged 20 to 80 have been published, using cardiopulmonary exercise testing (CPX) data (treadmill protocol) from the Fitness Registry and the Importance of Exercise National Database (FRIEND Registry).

In the current observational study we provide centile curves for the FRIEND Registry V'_E/V'_{CO_2} slopes, fitted using the generalised additive model for location, scale and shape (GAMLSS), to provide individuals with a more precise estimate of where their V'_E/V'_{CO_2} slopes fall within the population. We also confirm that by adopting allometric models (incorporating a log transformation), the resulting ANCOVAs provided more normal and homoscedastic residuals, with superior goodness-of-fit using the Akaike information criterion (AIC)=14671 (compared with traditional ANCOVA's AIC=15008) that confirms allometric models are vastly superior to traditional ANCOVA models.

In conclusion, providing sex-by-age centile curves rather than referring to reference tables for ventilatory efficiency (V'_E/V'_{CO_2} slopes) will provide more accurate estimates of where an individual's particular V'_E/V'_{CO_2} slope falls within the population. Also, by adopting allometric models researchers are more likely to identify real and valid inferences when analysing population/group differences in V'_E/V'_{CO_2} slopes.

Introduction
Numerous studies and subsequent position statements have confirmed the importance of cardiopulmonary exercise testing (CPX) as the gold-standard approach for the assessment of cardiorespiratory fitness [1–3]. The term cardiorespiratory fitness (CRF) refers to the ability of the circulatory and respiratory systems to supply oxygen to skeletal muscles during sustained physical activity. The clinical and research focus of CPX is usually focused on the measurement/quantification of peak or maximal oxygen uptake (V'_{O_2}). Accordingly, considerable effort has been made to establish reference values for peak/maximal V'_{O_2} across the lifespan in both males and females [4–6].

Although peak/maximal V'_{O_2} is the gold-standard measurement for exercise capacity specifically, it does not capture all of the characteristics associated with CRF, or all of the CPX responses that are useful in stratifying risk [3]. To quantify CPX responses more comprehensively, the assessment of other parallel

variables is necessary. One such response is ventilatory efficiency, most commonly assessed as the minute ventilation/carbon dioxide production (V'_E/V'_{CO_2}) slope, a measure that dramatically improves diagnostic and prognostic resolution as well as more accurately quantifies the degree of therapeutic efficacy for surgical, pharmacological and lifestyle interventions [1, 3].

In their recent study, Arena *et al.* [7] provide important reference values as discrete point estimates for V'_E/V'_{CO_2} slopes across the lifespan in both males and females (reporting mean values at 10 per cent intervals per decile). Although a helpful guide, these discrete point estimates require a certain degree of "linear interpolation" to estimate precisely where an individual's V'_E/V'_{CO_2} slope falls within the table (*e.g.*, somewhere between the 80th and 90th centile for a 68-year-old). An alternative **continuous** "growth" curve-fitting methodology for centile reference values (fitted using the Lamda-mu-sigma (LMS) proposed by Cole [8]) will provide a more accurate estimate.

Comparing population differences in V'_E/V'_{CO_2} slopes is usually performed using ANOVA or ANCOVA. These analyses assume these data are normally distributed and homoscedastic. As far as the ratio variables such as peak/maximal V'_{O_2} (mL·kg^{-1}·min^{-1}) and V'_E/V'_{CO_2} slopes are concerned, these assumptions are unlikely to be satisfied [6] with evidence of the heteroscedastic spread of the V'_E/V'_{CO_2} slope data with age in the centile plots shown in figure 1a and b. Ratio data such as peak/maximal V'_{O_2} (mL·kg^{-1}·min^{-1}) and the V'_E/V'_{CO_2} slope will invariably benefit from the log transformation [9]. Hence, the purpose of the current study is as follows: 1) to provide "individual" comparisons using the centile curves for V'_E/V'_{CO_2} slope values rather than point estimates (to help overcome the linear interpolation issues described above), curves estimated using the generalised additive model for location, scale and shape (GAMLSS), fitted using the GAMLSS package [10] in R [11]; and 2) to compare population differences in the V'_E/V'_{CO_2} slope using a more valid analysis (ANOVA/ANCOVA) with the help of an allometric model (incorporating

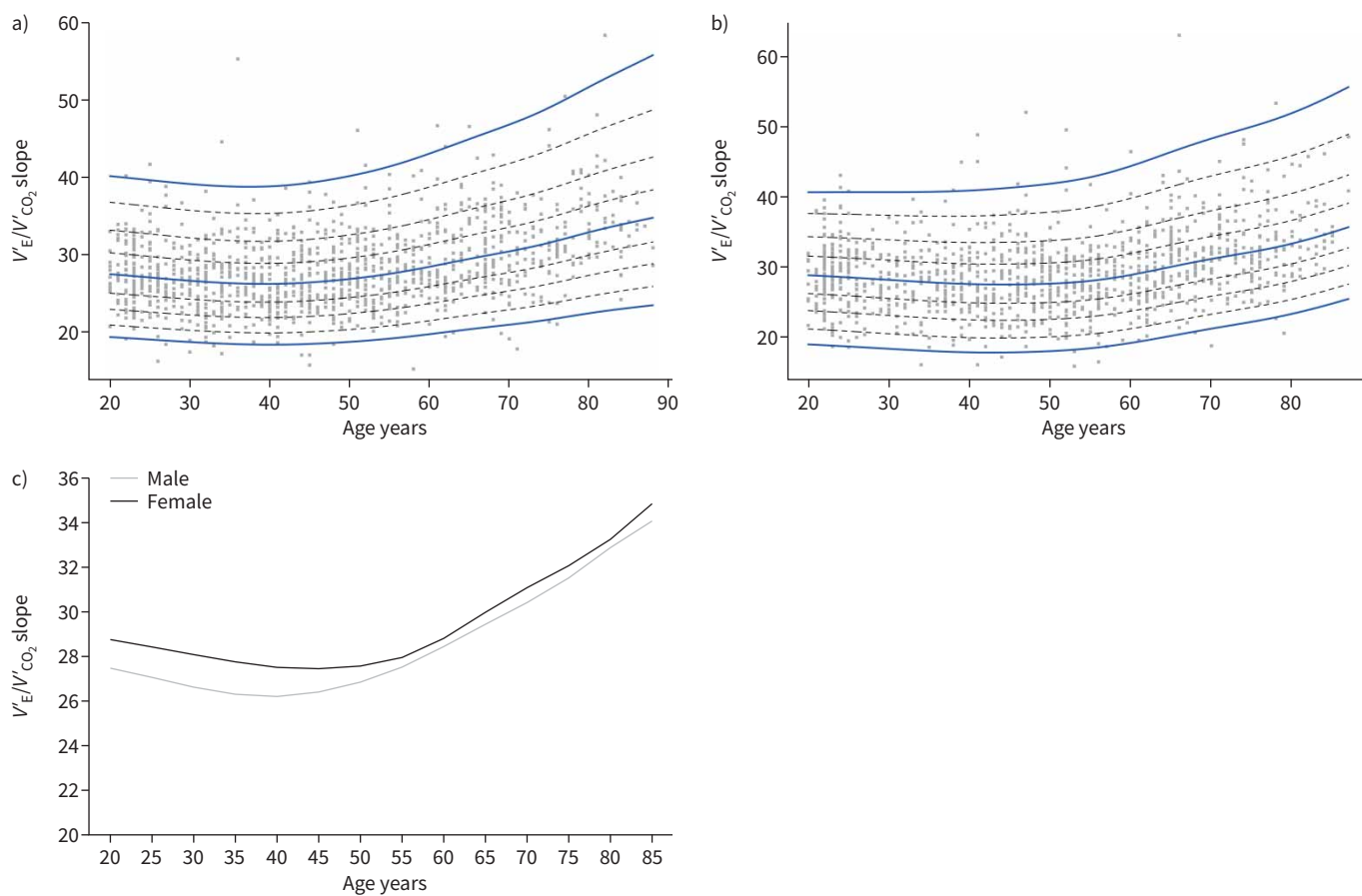

FIGURE 1 a) The centile curves for minute ventilation/carbon dioxide production (V'_E/V'_{CO_2}) slope by age for males. b) The centile curves for V'_E/V'_{CO_2} slope by age for females. c) The 50th centile curves for V'_E/V'_{CO_2} slope for males and females by age.

TABLE 1 Participating CPX laboratories
1) Ball State University, Muncie, Indiana
2) Brigham and Women's Hospital, Boston, Massachusetts
3) Brooke Army Medical Center, Houston, Texas
4) Goshen Heart and Vascular Center, Goshen, Indiana
5) Massachusetts General Hospital, Boston, Massachusetts
6) Saint Francis University, Loretto, Pennsylvania
7) San Francisco State University, San Francisco, California
8) University of Kansas Medical Center, Kansas City, Kansas
9) Veteran's Hospital – Palo Alto, Palo Alto, California
10) Yale University, New Haven, Connecticut

a log transformation) that are likely to provide more normally and homoscedastic residuals, thus making any inference concerning population differences more trustworthy.

Methods

The FRIEND registry

The FRIEND Registry was established with the primary charge of establishing reference CRF values across the adult lifespan [4, 12]. The CPX laboratories are currently contributing data to the FRIEND Registry function consistent with recommendations provided in recently published guidelines including the use of valid and reliable calibration and testing procedures as well as employing experienced personnel qualified to conduct exercise tests [13, 14]. Participating CPX laboratories were responsible for obtaining local institutional review board approval, or exemption from review, for inclusion in the FRIEND Registry, providing documentation that they were authorised to submit de-identified, coded data to the core CPX laboratory housed at the University of Illinois at Chicago (UIC). Institutional review board approval for the core CPX laboratory was also obtained at UIC.

Cohort characteristics, inclusion/exclusion criteria and study data points

The current analysis included 2511 tests from the 10 CPX laboratories listed in table 1. All CPX tests undertaken by the laboratories adopted a treadmill protocol. The mean±SD time to completion of the exercise tests was 11.23±3.06 min.

Inclusion criteria for the current analysis included CPX data on males and females: 1) aged >20 years; and 2) with a maximal CPX performed on a treadmill following established guidelines [1, 2]. Note that only treadmill protocols were performed in all 10 laboratories. Any subject identified as having a pre-existing medical condition was excluded. Thus, the cohort included apparently healthy subjects (i.e. subjects

TABLE 2 Descriptive data for all subjects by sex and age group in years											
Sex	Age group	n	Age years	SD	Weight kg	SD	Height cm	SD	BMI kg·m^{-2}	SD
Female	20.00	252	23.91	2.283	64.97	2.983	146.57	34.319	24.40	5.364
	30.00	156	34.83	2.970	65.25	2.375	171.91	48.196	28.37	7.735
	40.00	206	44.79	2.871	64.44	2.614	163.41	43.749	27.66	7.128
	50.00	227	53.78	2.728	64.38	2.845	165.01	40.204	28.02	6.666
	60.00	209	64.17	2.787	64.46	2.537	160.40	33.313	27.19	5.650
	70.00	140	74.19	2.810	63.69	2.543	154.72	27.747	26.85	4.491
	80.00	43	82.37	2.000	62.30	2.306	139.92	25.157	25.37	4.249
	Total	1233	48.85	17.848	64.48	2.740	159.02	38.996	26.90	6.344
Male	20.00	242	24.34	2.532	70.52	3.030	180.40	32.833	25.51	4.482
	30.00	221	34.39	2.939	70.68	2.924	204.40	51.681	28.72	6.738
	40.00	252	44.50	2.941	70.72	2.870	208.93	37.885	29.35	4.927
	50.00	220	54.04	2.769	70.32	2.848	203.79	39.039	29.00	5.458
	60.00	181	64.60	2.941	70.00	2.876	194.83	36.489	27.95	4.825
	70.00	111	73.99	2.862	69.44	3.118	188.40	31.797	27.52	4.380
	80.00	51	82.78	2.248	69.33	3.042	170.77	22.598	24.99	2.720
	Total	1278	47.51	17.336	70.34	2.961	196.56	40.423	27.92	5.375

BMI: body mass index.

without a pre-existing medical condition) who achieved maximal exertion by established peak respiratory exchange ratio (RER) criteria [1]. Descriptive data for all subjects by sex and age groups are provided in table 2.

V'_{O_2}, peak RER, peak heart rate (HR) and the V'_E/V'_{CO_2} slope were the CPX variables reported in the current study. Each participating CPX laboratory input VE and VCO$_2$ data from the initiation of exercise to peak and calculated the V'_E/V'_{CO_2} slope *via* least squares linear regression (y=mx+b, m=slope).

Statistical methods (to assess individual differences using centile curves)

The centile curves for the V'_E/V'_{CO_2} slope by age were produced using the lamda-mu-sigma (LMS) method. This method assumes that using a power transformation, data can be appropriately normalised, by stretching one tail of the distribution and shrinking the other and so removing skewness [8]. The power transformation applied in the present models used the Box–Cox t (BCT; μ, σ, ν, τ). This distribution is defined by Y^v having a shifted truncated t distribution with τ degrees of freedom, μ is the median of the distribution, $\sigma \cdot (\tau/(\tau-2))^{0.5}$ is approximation of the coefficient of variation, ν controls the skewness and τ the kurtosis of the distribution [15]. The generalised additive model for location, scale and shape (GAMLSS) allows each of the parameters of the distribution to be modelled as linear and/or nonlinear parametric, and/or smooth non-parametric functions of explanatory variables. A Fisher scoring algorithm was used to fit the model by maximising a penalised likelihood [10]. The analysis and centile curves were fitted and plotted using the GAMLSS package [16] in R [11]. Post-estimation diagnostics for these models included not only standard QQ-plots but also de-trended normal QQ-plots (worm plots) to check the age-conditional normality of the transformed data [17].

Statistical methods (to assess population differences)

The additive, linear model for V'_E/V'_{CO_2} slope proposed by [7] is given by

$$V'_E/V'_{CO_2} \text{ slope} = a + b_1 \cdot age + b_2 \cdot BMI + \varepsilon, \tag{1}$$

where ε is an additive error term that is assumed to be both normally distributed and homoscedastic (remains constant throughout the range of observations). The intercept "a" was allowed to vary with sex.

An alternative multiplicative model for the V'_E/V'_{CO_2} slope with allometric body size components originally proposed for V'_{O_2max} by NEVILL and HOLDER [18], and subsequently adapted by NEVILL and COOKE [19] to include an age^2 term, is given by

$$V'_E/V'_{CO_2} \text{ slope} = M^{k_1} \cdot H^{k_2} \cdot \exp(a + b_1 \cdot age + b_2 \cdot age^2) \cdot \varepsilon, \tag{2}$$

where M=mass, H=height and "ε" is a multiplicative, error ratio that assumes the error will be in proportion to V'_E/V'_{CO_2} slope.

The model (Eq. 2) can be linearised with a log transformation (using Ln=log$_e$). A linear regression or ANCOVA analysis on Ln(V'_E/V'_{CO_2} slope) can then be used to estimate the unknown parameters in the log-transformed model, *i.e.* the transformed model (Eq. 3) is now additive that conforms with the assumptions associated with ordinary least squares and ANOVA:

$$Ln(V'_E/V'_{CO_2} \text{ slope}) = k_1 \cdot Ln(M) + k_2 \cdot Ln(H) + a + b_1 \cdot age + b_2 \cdot age^2 + Ln(\varepsilon), \tag{3}$$

where the residual errors Ln(ε) are assumed to be normally distributed and homoscedastic. For researchers unfamiliar with analysing these allometric models (Eq. 3), the dependent variable is defined as the log-transformed V'_E/V'_{CO_2} slopes, Ln(V'_E/V'_{CO_2} slope), and Ln(M), Ln(H), age and age^2 are incorporated as covariates. Population differences in V'_E/V'_{CO_2} slope between groups, for example sex, can be assessed by allowing the intercept "a" to vary for each fixed factor in the ANOVA/ANCOVA analyses.

Normality was assessed using the Shapiro–Wilk and Kolmogorov–Smirnov tests. The Shapiro–Wilk statistic measures how well the data follow a normal distribution by correlating the association between the measured data and the calculated normal scores. If the correlation coefficient is near 1, the population is likely to be normal. Larger values for the Kolmogorov–Smirnov statistic (KS) indicate that the data do not follow the normal distribution.

Model comparison (goodness-of-fit) between the linear and allometric models was assessed using the Akaike information criterion (AIC). Traditionally, R^2 is frequently used to measure goodness-of-fit. However, higher R^2 values do not always indicate a better fit. Higher R^2 can indicate overfitting, and adding noise variables will also inflate R^2. So R^2, while useful, is not necessarily the best method of comparing competing models. An alternative method of model comparison is to use the AIC. AIC can be conceptualised as a "distance" or error between the data and a model, lower values meaning a better model. Unlike R^2, which rewards added noise variables with a higher (better) value, adding noise variables will increase AIC which penalises its values (lower being better). A difference between two AIC values <2 is really irrelevant; differences >2 ⩽6, evidence that the lower model is better/superior is somewhat positive; >6 ⩽10, the evidence that the model with the lower AIC value is better/superior is strong; >10, the evidence that the model with the lower AIC is better/superior is very strong. Given differences between AIC values can be challenging to comprehend, beyond appreciating that one model is better than another, evidence ratios (see NEVILL and COOKE [19]) will also be calculated to present the magnitude of the difference in terms of how many times better the best model is in terms of its predictive accuracy [20–22].

Results

Individual comparisons

The centile curves for the V'_E/V'_{CO_2} slope by age are given for males and females separately in figure 1a and b, and for comparative purposes, the 50th percentile values for the male and female subjects by age in figure 1c.

These curves enable the reader to estimate an individual's V'_E/V'_{CO_2} slope value within a population of healthy individuals (N=2511) for comparative purposes. Point-estimate centile tables by age for male and female individuals are also given in table 3.

Population comparisons
Linear, additive models

Comparing population differences in V'_E/V'_{CO_2} is usually performed using ANOVA or ANCOVA. To assess differences in sex using Eq. 1, we obtain

$$V'_E/V'_{CO_2} \text{ slope} = 26.4 + 0.09(\text{age}) - 0.08(\text{BMI in kg·m}^{-2}) + 0.83(\text{sex}),$$

where the variable sex is incorporated as a (0, 1) indicator variable, given by male=0 and female=1. However, the total R value for this equation was 0.33 with an adjusted R-square of 0.10 and a standard error of the estimate (SEE) of 4.8. Parameter estimates and 95% confidence intervals are given in table 4.

When the residuals were plotted against the "fits" or predicted (V'_E/V'_{CO_2} slope) values, the plot (figure 2a) illustrates a lack-of-fit (the "fits" quadratic term was B=0.26, t=8.14; p<0.001) and heteroscedasticity (confirmed when the absolute residuals were correlated against the predicted values (n=2511), r=0.089, p<0.001).

The Shapiro–Wilk statistic was r=0.967 and the Kolmogorov–Smirnov statistic KS=0.05, indicating that the residuals were not normally distributed (p<0.001).

Multiplicative allometric model

Fitting the multiplicative allometric model to the FRIEND's data using Eq. 3, we obtained the fitted parameters given in table 5.

In order to compare the quality of fit (AIC) values using the same parameters as linear model Eq. 1, but adopting the allometric model structure, we obtained AIC=14 671.04. Using the allometric model with age (but not age^2), height, weight and gender as predictors, AIC=14 672.33. Note that the AIC for the linear model Eq. 1 was 15 007.99 (table 4), a difference in AIC=336.95. The evidence ratio for this difference in AIC values indicates the allometric model is 3.921438×10^{71} times better in terms of model parsimony than the additive linear model. Parsimony is important because it helps discriminate signal from the noise, facilitating generalisation and increasing predictive ability for new data.

When the residuals were plotted against the "fits" or predicted Ln(V'_E/V'_{CO_2} slope) values, the plot (figure 2b) illustrates acceptable fit (the "fits" quadratic term was B=−0.9, t=1.72; p=0.086) and homoscedasticity (confirmed when the absolute residuals were correlated against the predicted values (N=2511), r=0.000, p=0.998).

TABLE 3 Centile table for minute ventilation/carbon dioxide production (V'_E/V'_{CO_2}) slope for healthy 20- to 85-year-olds (centiles 1–99)

Age years	Centiles								
	C1	C3	C10	C25	C50	C75	C90	C97	C99
Males									
20	19.34	20.88	22.92	25.02	27.48	30.24	33.18	36.79	40.18
25	19.01	20.54	22.56	24.63	27.05	29.77	32.68	36.26	39.65
30	18.69	20.21	22.20	24.24	26.62	29.31	32.18	35.75	39.15
35	18.45	19.96	21.94	23.96	26.31	28.97	31.83	35.40	38.83
40	18.35	19.87	21.85	23.86	26.21	28.86	31.73	35.34	38.84
45	18.44	19.99	22.01	24.04	26.40	29.08	31.99	35.68	39.31
50	18.71	20.31	22.37	24.44	26.85	29.58	32.56	36.38	40.18
55	19.13	20.79	22.92	25.05	27.52	30.32	33.40	37.39	41.42
60	19.71	21.45	23.67	25.88	28.44	31.34	34.55	38.76	43.08
65	20.34	22.18	24.51	26.81	29.46	32.47	35.83	40.30	44.97
70	20.93	22.87	25.30	27.69	30.43	33.55	37.06	41.81	46.84
75	21.60	23.66	26.22	28.70	31.55	34.79	38.48	43.54	49.04
80	22.41	24.61	27.32	29.93	32.91	36.31	40.21	45.66	51.72
85	23.10	25.45	28.31	31.03	34.12	37.67	41.77	47.63	54.31
Females									
20	18.90	21.10	23.73	26.15	28.77	31.50	34.27	37.57	40.60
25	18.61	20.79	23.40	25.83	28.46	31.23	34.07	37.46	40.62
30	18.28	20.43	23.03	25.46	28.11	30.92	33.81	37.32	40.62
35	17.97	20.09	22.67	25.10	27.76	30.61	33.57	37.20	40.66
40	17.77	19.86	22.42	24.84	27.51	30.40	33.44	37.21	40.85
45	17.76	19.82	22.35	24.77	27.45	30.37	33.49	37.41	41.24
50	17.94	19.97	22.48	24.88	27.57	30.52	33.71	37.78	41.82
55	18.36	20.37	22.87	25.26	27.97	30.96	34.23	38.47	42.76
60	19.11	21.14	23.65	26.07	28.82	31.89	35.28	39.74	44.34
65	20.12	22.17	24.71	27.17	29.97	33.13	36.66	41.38	46.35
70	21.15	23.21	25.77	28.25	31.09	34.32	37.97	42.93	48.26
75	22.12	24.18	26.75	29.23	32.09	35.36	39.10	44.28	49.96
80	23.26	25.34	27.91	30.41	33.28	36.60	40.44	45.84	51.90
85	24.73	26.84	29.44	31.97	34.90	38.29	42.25	47.93	54.45

The Shapiro–Wilk statistic was r=0.992 (compared with r=0.967 using linear Eq. 1) and the Kolmogorov–Smirnov statistic KS=0.031 (compared with KS=0.05 using Eq. 1), indicating that the residuals were still not normally distributed (p<0.001) but behave considerably better (more normal) than the residuals obtained from the linear additive model (Eq. 1).

To illustrate the possible inferential pitfalls that might occur when investigating population differences using linear models, we re-examined the V'_E/V'_{CO_2} slopes from the FRIEND dataset using allometric

TABLE 4 Parameter estimates and 95% confidence intervals obtained from fitting Eq. 1

Parameter estimates						
Dependent variable: V'_E/V'_{CO_2} slope						
Parameter	B	SE	t	Sig.	95% confidence interval	
					Lower bound	Upper bound
Intercept	26.38	0.53	50.12	0.00	25.35	27.42
Age	0.09	0.01	15.89	0.00	0.08	0.10
BMI	−0.08	0.02	−4.89	0.00	−0.11	−0.05
Sex (male)	0.83	0.19	4.32	0.00	0.45	1.21

R squared=0.105 (adjusted R squared=0.104); Akaike information criterion (AIC)=15 009. Reference or baseline group=females. BMI: body mass index.
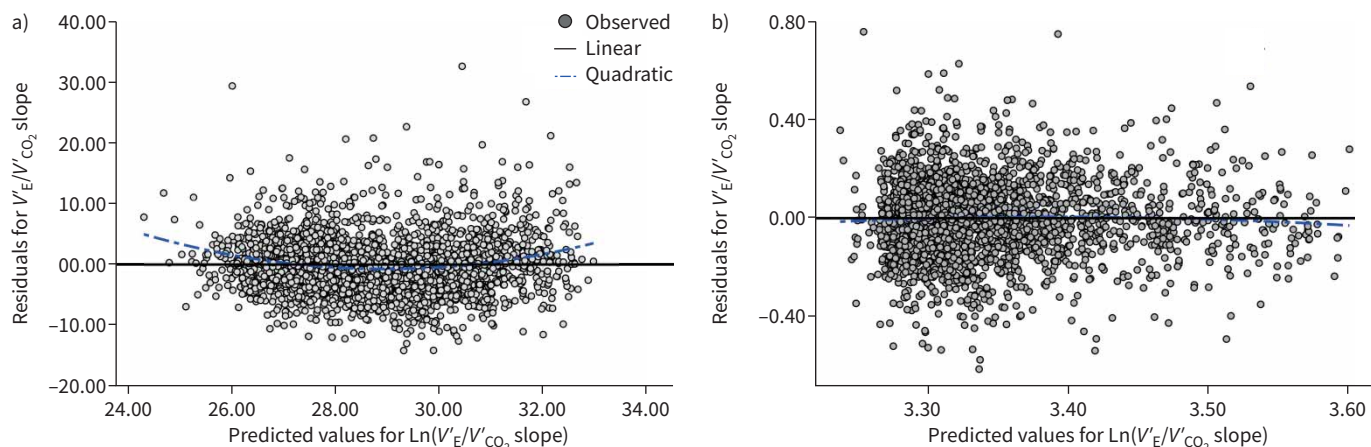
FIGURE 2 a) The residuals *versus* the "fits" or predicted (minute ventilation/carbon dioxide production (V'_E/V'_{CO_2} slope) values using Eq. 1. b) The residuals *versus* the "fits" or predicted (V'_E/V'_{CO_2} slope) values using Eq. 3.

models but incorporating a sex-by-age interaction term as well as a sex main effect and allowing subjects to be nested within sites. The re-analysis revealed a significant age-by-sex interaction ($p<0.001$) that can be clearly seen in figure 1c, a finding that appears to have been overlooked by ARENA *et al.* [7] using the linear additive model (Eq. 1).

Discussion

The first aim/purpose of the current study was to provide centile "growth" curves (by age) for ventilatory efficiency as recorded by the V'_E/V'_{CO_2} slope values. ARENA *et al.* [7] provided discrete point-estimate centile reference values for the V'_E/V'_{CO_2} slope by age (using deciles) and sex, when all treadmill-testing exercise data were used to calculate the slopes. In the current study we report **continuous** "growth" curves for the centile reference values (fitted using the lamda-mu-sigma (LMS) method proposed by COLE [8]) that provides the reader with the opportunity to more accurately **interpolate** an individual's reference values between the point estimates (deciles) reported by ARENA *et al.* [7]. In addition to this inherent error (or lack of precision) associated from using point estimates provided by reference tables, the reference tables would only allow a CPX administrator to say how close to "normal" the patient's V'_E/V'_{CO_2} slope value was. In comparison, the centile curves provide a valuable added context associated with the additional level of precision. The interpretation of centiles is straightforward. For example, in the case of an individual's V'_E/V'_{CO_2} slope and age, if their estimate is on the 25th centile, it means that for every 100 individuals of that age, 25 would have a lower V'_E/V'_{CO_2} slope and 75 a higher V'_E/V'_{CO_2} slope. These results may prove useful in enhancing the interpretation of CPX results. We recognise that similar centile curves have been published elsewhere recently by WAGNER *et al.* [23] but using a different CPX protocol (bicycle ergometry) and with a reduced sample size [24].

TABLE 5 Parameter estimates and 95% confidence intervals obtained from fitting Eq. 3						
Parameter estimates						
Dependent variable: Ln(V'_E/V'_{CO_2} slope)						
Parameter	**B**	**SE**	**t**	**Sig.**	**95% confidence interval**	
					Lower bound	**Upper bound**
Intercept	3.448	0.324	10.639	0.000	2.813	4.084
Age	−0.009	0.001	−8.770	0.000	−0.011	−0.007
Age^2	0.00012	1.064E-5	11.604	0.000	0.000	0.000
LnMass	−0.052	0.017	−3.147	0.002	−0.085	−0.020
LnHeight	0.067	0.080	0.832	0.405	−0.091	0.224
Sex (male)	0.024	0.009	2.594	0.010	0.006	0.042

R squared=0.151 (adjusted R squared=0.149); Akaike information criterion (AIC)=14 679.3. Reference or baseline group=females.

Comparing the centile curves reported by WAGNER *et al.* [23] using bicycle ergometry with those reported in the current study using a treadmill protocol, we detected large differences. The 50th centiles for the male V'_E/V'_{CO_2} slopes reported by WAGNER *et al.* [23] ranged from 34.2 to 37.2 (for 30- to 60-year-old males, respectively). In contrast, 50th centiles for the male V'_E/V'_{CO_2} slopes reported in the current study (treadmill protocol) ranged from 26.31 to 29.46 for 30- to 60-year-old males, respectively. A similar contrast is observed with female V'_E/V'_{CO_2} slopes. The 50th centiles for the female V'_E/V'_{CO_2} slope scores reported by WAGNER *et al.* [23] ranged from 34.0 to 36.5 (30 - to 60 -year-old females, respectively). The 50th centiles for the female V'_E/V'_{CO_2} slope scores reported in the current study (treadmill) ranged from 27.76 to 29.97 for 30- to 60-year-old females, respectively. These differences suggest that the protocol used to obtain the V'_E/V'_{CO_2} slopes is crucial and outlines the necessity to refer to the appropriate centile curves depending on the CPX protocol used.

The second aim/purpose of the current study was to recommend how researchers should compare population differences in the V'_E/V'_{CO_2} slopes. The majority of researchers exploring population differences in V'_E/V'_{CO_2} slopes have used traditional ANOVA, ANCOVA or regression methods (such as Eq. 1) to explore differences between categorical variables such as age, sex, smoking habits, BMI categories and exercise habits [5, 7, 25, 26]. However, the range of ratio variables (such as V'_{O_2max} ($mL \cdot kg^{-1} \cdot min^{-1}$) and the V'_E/V'_{CO_2} slope) that are also associated with body size are invariably bounded by zero to the left (cannot be negative) but unbounded to the right. This leads its frequency distribution to be positively skewed and not normally distributed.

The solution to overcome such problems would appear to adopt an alternative multiplicative model with allometric body size components (Eq. 2). The model can be linearised with a log transformation (using $Ln = log_e$) and linear regression or ANCOVA analysis on $Ln(V'_E/V'_{CO_2}$ slope) can then be used to estimate the unknown parameters (see Eq. 3). The log transformation will naturally overcome both the positive skewness and heteroscedasticity, the latter seen clearly in the centile curves figure 1a and b, where the more extreme curves (*e.g.* 90% and 10% centile) diverge with increasing age in both the male and female figures.

The allometric model (Eq. 2) performs better than the linear model (Eq. 1) based on ALL model comparison criteria. 1) The explained variance was greater using the allometric model ($R^2=0.151$) compared with the linear models ($R^2=0.105$). 2) Both the Shapiro–Wilk statistic and Kolmogorov–Smirnov tests indicate that the saved residuals from the linear and allometric models were not normally distributed. However, the residuals from the allometric model were closer to a normal distribution than the residuals saved from the linear model, with the Q-Q plot correlations for the linear and allometric models being 0.967 and 0.992, respectively. 3) The residuals *versus* the predicted values (fits) plots provided evidence of curvature (lack-of-fit) as seen with the linear model in figure 2a. No such evidence was apparent with the allometric model in figure 2b. 4) The goodness-of-fit assessed using the AIC was much lower (AIC=14 679.3) for the allometric model compared with the linear additive model (AIC=15 008.0). The 328.7 difference between models points to a major improvement, some 71 orders of magnitude greater in terms of parsimony and therefore predictive accuracy.

Clearly, these results suggest that allometric models (incorporating a log transformation) are likely to provide a better fit to the V'_E/V'_{CO_2} slope values with more normally and homoscedastic residuals, thus making any inference concerning population differences more valid and trustworthy. This was reinforced when the allometric model identified the sex-by-age interaction (see figure 1c) that appears to have been missed using linear additive models (ARENA *et al.* 2020 [7]). Some authors [25] have also reported an age-by-sex interaction (in their table 4), but as yet no mechanism has been provided to explain why ventilatory efficiency (V'_E/V'_{CO_2} slope) is higher in females than males in younger age groups but converges (little or no difference) in the older age groups. This can be considered a limitation of the study.

In summary, providing centile curves (by age and sex) rather than having to refer to mean reference tables for ventilatory efficiency (V'_E/V'_{CO_2} slopes) should provide individuals with a more accurate estimate of where their particular V'_E/V'_{CO_2} slope falls within the general population. Also, by adopting the allometric model (Eq. 2), fitted using the log-transformed model/equation (Eq. 3), researchers are more likely to identify real and/or valid inferences, leading to more trustworthy interpretations of population differences (and interactions) when analysing V'_E/V'_{CO_2} slopes.

References

1   Balady GJ, Arena R, Sietsema K, *et al.* Clinician's guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. *Circulation* 2010; 122: 191–225.

2   Guazzi M, Adams V, Conraads V, *et al.* EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Circulation* 2012; 126: 2261–2274.

3   Ross R, Blair SN, Arena R, *et al.* Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. *Circulation* 2016; 134: e653–e699.

4   Kaminsky LA, Arena R, Myers J. Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing: data from the Fitness Registry and the Importance of Exercise National Database. *Mayo Clin Proc* 2015; 90: 1515–1523.

5   Loe H, Steinshamn S, Wisløff U. Cardio-respiratory reference data in 4631 healthy men and women 20–90 years: the HUNT 3 fitness study. *PLoS ONE* 2014; 9: e113884.

6   Ingle L, Rigby A, Brodie D, *et al.* Normative reference values for estimated cardiorespiratory fitness in apparently healthy British men and women. *PLoS ONE* 2020; 15: e0240099.

7   Arena R, Myers J, Harper M, *et al.* The VE/VCO2 slope during maximal treadmill cardiopulmonary exercise testing: reference standards from FRIEND (Fitness Registry and the Importance of Exercise): An International Database. J Cardiopulm Rehabil Prev 2021; 41: 194-198.

8   Cole TJ. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr* 1990; 44: 45–60.

9   Nevill AM. Why the analysis of performance variables recorded on the ratio scale will invariably benefit from a log transformation. *J Sports Sci* 1997; 15: 457–458.

10 Rigby RA, Stasinopoulos DM. Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. *Stat Model* 2006; 6: 209–229.

11 R Core Team. R: A language and environment for statistical computing, 2020. https://www.R-project.org.

12 Kaminsky LA, Arena R, Beckie TM, *et al.* The importance of cardiorespiratory fitness in the United States: the need for a national registry: a policy statement from the American Heart Association. *Circulation* 2013; 127: 652–662.

13 Myers J, Arena R, Franklin B, *et al.* Recommendations for clinical exercise laboratories: a scientific statement from the American Heart Association. *Circulation* 2009; 119: 3144–3161.

14 Myers J, Forman DE, Balady GJ, *et al.* Supervision of exercise testing by nonphysicians: a scientific statement from the American Heart Association. *Circulation* 2014; 130: 1014–1027.

15 Stasinopoulos MD, Rigby RA, Bastiani FD. GAMLSS: a distributional regression approach. *Stat Model* 2018; 18: 248–273.

16 Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *J R Stat Soc Ser C Appl Stat* 2005; 54: 507–554.

17 Buuren S van Fredriks, M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Stat Med* 2001; 20: 1259–1277.

18 Nevill AM, Holder RL. Modelling maximum oxygen uptake—a case-study in nonlinear regression model formulation and comparison. *J R Stat Soc Ser C Appl Stat* 1994; 43: 653–666.

19 Nevill AM, Cooke CB. The dangers of estimating V̇O2max using linear, nonexercise prediction models. *Med Sci Sports Exerc* 2017; 49: 1036–1042.

20 Anderson D, Burnham K. Model selection and multi-model inference. *Second Springer-Verl* 2004; 63: 10.

21 Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control* 1974; 19: 716–723.

22 Spiegelhalter D, Best NG, Carlin BP, Linde A van der. Bayesian measures of model complexity and fit. *Qual Control Appl Stat* 2003; 48: 431–432.

23 Wagner J, Knaier R, Infanger D, *et al.* Novel CPET reference values in healthy adults: associations with physical activity. *Med Sci Sports Exerc* 2021; 53: 26–37.

24 Cole TJ. Sample size and sample composition for constructing growth reference centiles. *Stat Methods Med Res* 2021; 30: 488–507.

25 Koch B, Schäper C, Ittermann T, *et al.* Reference values for cardiopulmonary exercise testing in healthy volunteers: the SHIP study. *Eur Respir J* 2009; 33: 389–397.

26 Salvioni E, Corrà U, Piepoli M, *et al.* Gender and age normalization and ventilation efficiency during exercise in heart failure with reduced ejection fraction. *ESC Heart Fail* 2020; 7: 368–377.