Ejection fraction as a statistical index of left ventricular systolic function: The first full allometric scrutiny of its appropriateness and accuracy

Running head: Modeling left ventricular volumes

Lorenzo Lolli¹, Alan M Batterham¹, Greg Atkinson¹

Affiliations:
¹ Health and Social Care Institute, School of Health and Social Care, Teesside University, Middlesbrough, United Kingdom.

Corresponding author:
Lorenzo Lolli
Health and Social Care Institute
School of Health and Social Care, Constantine Building
Teesside University, Middlesbrough, TS1 3BA, United Kingdom
Tel: +44 (0) 1642 342934
Email: L.Lolli@tees.ac.uk

Word count: Manuscript: 4647 words; Summary: 247 words

Items: 1 table; 2 figures; 2 supplemental tables
Summary

Left ventricular ejection fraction (EF) is a ratio that is deemed to accurately normalize stroke volume (SV) to end-diastolic volume (EDV). Ratios are now well-recognised for not normalizing the numerator, in this case SV, consistently for the denominator, EDV. We aimed to provide the very first allometric-based scrutiny of the conventional assumptions that underpin the EF ratio. We allometrically-modeled untransformed SV and EDV measurements from 112 preclinical heart failure patients in the Multi-Ethnic Study of Atherosclerosis (MESA), and 864 chronic heart failure patients in the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) study. An information-theoretic approach was adopted to assess the relative quality of twelve candidate models for normalizing SV to EDV. None of the conventional underlying assumptions for accurate ratio normalization, e.g. an allometric exponent $\approx 1$, were upheld for EF. A two-parameter power function with normal, heteroscedastic error was the best model for scaling SV to EDV in both samples. The allometric exponent (95% confidence interval) was $0.776$ (0.682 to 0.869) in MESA, and $0.860$ (0.857 to 0.864) in TOPCAT. EF was inversely correlated with EDV in MESA ($r = -0.67$, 95%CI: $-0.76$ to $-0.55$) and TOPCAT ($r = -0.41$, 95%CI: $-0.46$ to $-0.35$). Consequently, for fundamental statistical reasons, EF was biased low for people with generally larger EDVs, and vice versa. For the first time, we have demonstrated that EF is an inaccurate statistic for scaling SV to EDV, leading to potential biased inferences for research and individual patients.

Key words: ejection fraction; allometry; heart failure; left ventricle; normalization
Introduction

Left ventricular ejection fraction (EF) is typically calculated as the ratio of stroke volume (SV) to end-diastolic volume (EDV) and expressed as a percentage statistic (Carabello 2002). Ejection fraction represents a criterion measure used to inform clinical decisions in the diagnosis and treatment pathways for heart failure (Dunlay et al., 2017). Heart failure is a multifactorial clinical syndrome resulting from pathological impairments in cardiac function and morphology (Abudiab et al., 2013) and is estimated to affect more than 37.7 million individuals worldwide (Ziaeian & Fonarow 2016). According to recent epidemiological data, hospital admissions due to heart failure are expected to increase by more than 50% by 2035 (Ziaeian & Fonarow 2016). In the United Kingdom, there are approximately 493,000 people living with a definite diagnosis of heart failure (Townsend et al., 2015), which imposes a substantial economic burden on the UK’s National Health Service, accounting for 2.1% of its overall budget (Cook et al., 2014).

Clinically, patients may often progress through a silent and asymptomatic phase of left ventricular systolic dysfunction that characterizes the transition from preclinical to overt heart failure (Goldberg & Jessup 2006). European guidelines for the diagnosis and treatment of acute and chronic heart failure (Ponikowski et al., 2016) differentiate patients with an EF < 40% as “heart failure with reduced ejection fraction” (HFrEF), EF ≥ 50% as “heart failure with preserved ejection fraction” (HFpEF), and a ‘gray zone’ in the range from 40% to 49% (HFmrEF). It has been reported that HF patients with a preserved EF have a 32% lower risk of mortality over a 3-year follow-up period compared with HF patients with a reduced EF (Meta-analysis Global Group in Chronic Heart 2012).

In a previous study, it was highlighted that “in chronic, compensated heart failure with reduced EF, the EF is reduced because the chamber size (denominator of EF equation) is larger, whereas the stroke volume (numerator) is typically similar to that of normal controls” (Borlaug & Redfield.
2011, page 2008). In classical allometry, EF is a ratio size-scaling index, the accuracy of which is reliant on SV varying as a constant proportion of EDV. Like many such ratio statistics, EF is a statistically robust measure of systolic function only if this assumption and other related assumptions are satisfied (Albrecht et al., 1993; Curran-Everett 2013; George et al., 2001; Tanner 1949). For example, recently-published studies have revealed that the percentage flow-mediated dilation index can misrepresent the true size-scaling association between resting and hyperaemic artery diameter, thereby entailing inaccurate inferences regarding human endothelial function (Atkinson & Batterham 2015).

Since EF is the selected statistic for informing the diagnosis and treatment of patients with heart failure (Dunlay et al., 2017; Ponikowski et al., 2016), we hypothesized that the true relationship between the left ventricular systolic and diastolic volumes might not be directly proportional in nature. It is this assumption which underpins the accuracy of the EF ratio statistic and, if false, would lead to biased inferences in research and diagnoses for individual patients. While studies on allometry and scaling in cardiovascular physiology have been traditionally conceived to standardise measures of cardiac structure and function to body size (Dewey et al., 2008), no previous study has comprehensively scrutinised the inherent scaling properties of the EF index itself, i.e. the inherent accuracy of how EF normalises stroke volume for differences in end diastolic volume.

Therefore, using a formal information-theoretic approach, we compared twelve candidate models for scaling SV to EDV in terms of the potential implications for general clinical practice using two samples of data (Studies 1 and 2). In study 1, we analysed data from preclinical heart failure patients enrolled in the Multi-Ethnic Study of Atherosclerosis (MESA), and, in study 2, we analysed data from patients already with chronic HF involved in the Treatment of Preserved Cardiac Function Heart Failure with and Aldosterone Antagonist (TOPCAT) echocardiographic sub-study.
Methods

Study 1 (MESA)

Participants

A detailed study protocol of the MESA has been previously reported (Bild et al., 2002). In brief, the MESA is a prospective, population-based study on the prevalence, incidence, and progression of subclinical cardiovascular disease (Bild et al., 2002). For the present study, participants at the baseline visit were selected based on the established diagnosis for incident heart failure after 8 years of follow-up (Habibi et al., 2014). The adjudication of a hard-cardiovascular event was established by a committee that included a cardiologist, an epidemiologist, and a neurologist. Incident heart failure was classified as definite, probable, or absent. The full criteria for the diagnosis of heart failure in the MESA were also detailed in previous studies (Bluemke et al., 2008; Yeboah et al., 2012). The MESA was approved by the local institutional review boards of each study centre, and participants provided written informed consent. The current study adhered to the ethics and research governance procedures at Teesside University.

Demographic, medical history, metabolic and cardiovascular data for this study were obtained at the MESA baseline examination. Resting blood pressure was determined as the average of the last two measurements in the seated position using a Dinamap model Pro 100 automated oscillometric sphygmomanometer (Critikon, Tampa, Florida). Hypertension was defined as systolic blood pressure > 140 mmHg, diastolic blood pressure > 90 mmHg, or the use of antihypertensive medication. Fasting plasma glucose ≥ 126 mg/dL or the use of anti-diabetic medications defined diabetes mellitus. The glomerular filtration rate (eGFR) was estimated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation (Levey et al., 2009). Smoking history was determined via standardized questionnaires. Body mass index (BMI) was calculated as the ratio of weight to height squared (kg/m²). Lipid profiling from blood samples was performed after a 12-h
fast. Low-density lipoprotein cholesterol was estimated with the Friedewald equation (Friedewald et al., 1972).

Cardiac magnetic resonance imaging

The magnetic resonance imaging (MRI) protocol procedures and reliability of the global left ventricular measurements have been outlined previously (Natori et al., 2006). Briefly, the MRI examination to quantify left ventricular functional and structural parameters consisted of a stack of short- and long-axis echo cine images covering the base-to-apex distance of the left ventricle with a temporal resolution of 50 ms (Natori et al., 2006). The EDV and end-systolic volume (ESV) were calculated using the Simpson’s rule from endocardial and epicardial myocardial borders (Natori et al., 2006). Left ventricular mass was the resultant of the difference between epicardial and endocardial areas times the slice thickness, section gap, and the specific gravity of the myocardium (i.e. 1.05 g/mL) (Natori et al., 2006). Papillary muscle mass was included in the left ventricular cavity volume and excluded from left ventricular mass (Bluemke et al., 2008). EF (%) was conventionally calculated as SV divided by EDV × 100.

Study 2 (TOPCAT)

Study population and definitions

TOPCAT is an intercontinental, double-blind, randomized, placebo-controlled, parallel-group study involving 3445 HF patients recruited at 266 centres in United States, Canada, Russia, Republic of Georgia, Argentina, and Brazil to test the efficacy and safety of an aldosterone antagonist to reduce cardiovascular morbidity and mortality in patients with heart failure and an EF ≥ 45% (Desai et al., 2011). The present study examined participants enrolled in the TOPCAT echocardiographic sub-study, a smaller sample from the TOPCAT trial (Shah et al., 2014). Participants were eligible if they had a technically-valid echocardiographic quantification of the left ventricular volumes derived according to the modified biplane Simpson’s rule, which represents the recommended method by
the American Society of Echocardiography (Lang et al., 2015). The TOPCAT trial was funded by the National Heart, Lung, and Blood Institute (NHLBI), and was approved by the local institutional review boards of each study centre (Pitt et al., 2014). The current study was compliant with the ethics and research governance procedures at Teesside University.

At the baseline visit, each participant underwent record screening, which included self-reported medical history and current medications, a physical examination (e.g. blood pressure, height, weight), and laboratory data collection involving complete blood count, electrolytes, blood urea nitrogen, creatinine, blood glucose, liver function assessment, and urine test for microalbuminuria. (Pitt et al., 2014; Shah et al., 2014). Participants’ chronological age, the inverse of serum creatinine, sex, and ethnicity were obtained to derive eGFR using the four-variable Modification of Diet in Renal Disease algorithm (Levey et al., 1999).

**Echocardiography**

The echocardiographic assessments, procedures, and intra-observer measurement variability in the TOPCAT echocardiographic sub-study have been described in detail previously (Shah et al., 2014). Each study centre submitted echocardiograms in digital or analog format to the core laboratory at the Brigham and Women’s hospital (Desai et al., 2011). Left ventricular endocardial borders were traced manually at the end of the diastolic and systolic phases in the 4- and 2-chamber apical views (Shah et al., 2014). The biplane method of disks (i.e. modified Simpson’s rule) was adopted to assess left ventricular volumes. Left ventricular mass estimation from linear dimension was performed according to the American Society of Echocardiography equation (Lang et al., 2015). Of the 935 echocardiographic measurements that were analyzable quantitatively, left ventricular volumes derived via the modified biplane Simpson’s rule were available in 864 study participants.
Statistical analyses and allometric modeling

The MESA (n = 112) and TOPCAT (n = 864) samples were examined separately. Demographic and clinical characteristics of participants at the baseline examination are presented as mean ± standard deviation (SD) for continuous variables and frequency or percentages for categorical variables.

To examine the scaling relationship between SV and EDV, we performed non-linear regression analyses of untransformed measurements. We fitted three sets of four models, involving two straight lines and two power functions, with multiplicative, log-normal, heteroscedastic error, and additive, normal, homoscedastic or heteroscedastic error, respectively (Packard 2017). Parameter estimates for each model were solved using an iterative protocol based on the Marquardt procedure (Packard 2017). Participants’ chronological age and sex (coded “0” for female, “1” for male) were included as continuous and categorical covariates in the models, respectively. The commonality of \( b \) exponent principle was tested to establish the presence of a common EDV exponent for both sexes (Batterham et al., 1997; Vanderburgh 1998). A substantial sex difference in the allometric exponent, predefined as ± 0.1 in the present study, would reveal a fundamental difference in the relationship between SV and EDV, thereby precluding meaningful comparisons between men and women (Batterham et al., 1997; Vanderburgh 1998).

The Akaike Information Criterion (AIC) was adopted to assess the relative quality of each model in the set of candidates (Burnham et al., 2011). The Akaike difference (\( \Delta \text{AIC} \)) from the estimated best model (i.e. the model with the lowest AIC value; \( \Delta \text{AIC} = 0 \)) was evaluated according to the following scale: 0-2, essentially equivalent; 2-7, plausible alternative; 7-14, weak support; > 14, no empirical support (Burnham et al., 2011). Parameter estimates were interpreted from the best/essentially equivalent models for the examined data. Regression coefficients were reported as point estimates with 95% confidence intervals (CI). Statistical analyses were carried out using SAS® software (PROC MODEL, Version 9.3; SAS Institute, Inc., Cary, NC, 2011), and figures
were produced using IBM Statistical Package for the Social Sciences (SPSS) Statistics v. 23.0 (SPSS, Chicago, IL, USA).

Table 1 about here

**Results**

**Allometric accuracy of the EF ratio in preclinical individuals (MESA)**

Among 5004 study participants with technically-valid measurements of the left ventricle obtained at the baseline visit, 112 participants reported a subsequent diagnosis of heart failure at a median 7.2-year follow-up. Of these participants, 43% were Caucasian (n = 48), 5% Chinese (n = 5), 31% African-American (n = 35), and 21% Hispanic (n = 24). Table 1 shows the summary data of the 112 study participants stratified by sex.

Graphical and statistical criteria indicated that the EF ratio failed to meet underlying assumptions for appropriate scaling. First, there was a large, negative correlation between the ratiometric index and its denominator corresponding to \( r = -0.67 \) (95%CI: \(-0.76\) to \(-0.55\)). This inconsistent normalization for EDV is also shown in Fig. 1a. Second, the linear regression between SV and EDV for the whole sample revealed a positive \( Y \)-intercept value of 44 mL (95%CI: 34 mL to 54 mL). Use of a ratio would only be appropriate if the line describing the bivariate relationship passes through the origin (Figure 1c). Accordingly, the ratio of the coefficient of variations (CV) for EDV to SV was substantially different from the correlation coefficient describing the bivariate relationship between the two variables (1.31 ≠ 0.68). A ratio standard model is valid only if this ratio of CVs is equal to the correlation coefficient between SV and EDV.
The AIC criteria revealed the two-parameter power function with normal, heteroscedastic error, of the form \( Y = a \cdot X^b \) (Figure 2a), to be the best of the twelve models (Supplemental Table 1). The allometric exponent \((b)\) describing the non-linear relationship between SV and EDV was 0.776 (95%CI: 0.682 to 0.869), with no main effects of chronological age and sex as predictor variables in the model. The mean difference in the EDV exponent between men and women was 0.073 (95%CI: -0.090 to 0.235). The EDV measurement spectrum ranged from 47 to 290 mL. Supplemental Table 1 shows the AIC values for each model in the set of candidates. In agreement with the AIC outcomes, the raw residuals from the best model were well-behaved (Figure 2c).

Allometric accuracy of the EF ratio in clinical individuals (TOPCAT)

Among the 864 eligible study participants, 83% were Caucasian \((n = 714)\), 13% were Black \((n = 114)\), less than 1% Asian \((n = 4)\), and 3% \((n = 30)\) were defined as a minor mixed-ethnic group.

Demographic and cardiovascular functional parameters of the 864 study participants are illustrated in Table 1.

The moderate, inverse association between the ratiometric EF and EDV corresponding to \( r = -0.41 \) (95%CI: -0.46 to -0.35) demonstrated that the conventional ratiometric EF index does not consistently control for the effects of EDV (Figure 1b). Likewise, the positive Y-intercept value of 13 mL (95%CI: 11 mL to 14 mL) observed in the bivariate relationship between SV and EDV indicated the failure of the ratiometric EF to meet another underlying assumption of ratio scaling models (Figure 1d). In fact, the substantial difference between CVx/CVy and the observed correlation coefficient between SV and EDV \((1.12 \neq 0.88)\) provided additional evidence about the inappropriateness of the EF ratio also for this data set. The two-parameter power function with normal, heteroscedastic error, of the form \( Y = a \cdot X^b \) (Figure 2b), emerged as the best model in the pool of twelve candidates (Supplemental Table 2). The allometric exponent \((b)\) describing the non-linear relationship between SV and EDV was 0.860 (95%CI: 0.857 to 0.864), with a substantial
main effect of chronological age in the model. The mean difference in the EDV exponent between men and women was 0.087 (95%CI: 0.080 to 0.093). The EDV measurements ranged from 27 to 233 mL for this TOPCAT sub-sample. The AIC values for each model in the set of candidates are shown in the Supplemental Table 2. The raw residuals from the best model plotted against the predicted values were found to be well-behaved (Figure 2d).

**How the EF ratio can misdiagnose individuals**

In both MESA and TOPCAT, application of allometric scaling methods revealed a substantial discrepancy between ratio and adjusted estimates of EF for some individuals on a between-subject basis. For example, in MESA, a 64-year-old, Caucasian man with no history of hard cardiovascular event, hypertension, a fasting glucose level of 87 mg/dl, and eGFR of 69.4 mL/min/1.73m², and a blood pressure of 124/73 mmHg, presented an SV of 91 mL within the age-specific range. On the other hand, left ventricular EDV (251 mL), ESV (160 mL), and mass (254 g) were markedly outside the physiological parameters. Although the calculated EF ratio was 36%, use of the more appropriate size-scaling model revealed an adjusted-EF of 41%. The allometric normalization of SV for differences in EDV in MESA thus revealed an absolute underestimation of the relative systolic function for this individual corresponding to 5%. In TOPCAT, ratio and allometric scaling approaches were found to provide substantially different estimates of EF in a 77-year-old, Caucasian woman with history of angina, hypertension, atrial fibrillation, a fasting glucose level of 91 mg/dl, an eGFR of 70.2 mL/min/1.73m², and on β-blockers therapy. The observed left ventricular EDV, ESV, SV, and mass were 48 mL, 23 mL, 25 mL, and 256 g, respectively. Notwithstanding the relatively small SV observed in this patient, the EF ratio of 52% indicated a preserved systolic function. Conversely, the more appropriate adjusted-EF estimate of 47% revealed
a substantial 5% overestimation of the true EF. Accordingly, the most appropriate size-scaling
model provided a more sensible estimate of EF, ultimately in line with the abnormal global
longitudinal strain of \(-13\%\) observed in this patient.

Discussion

For the first time, we report here that SV does not vary in direct proportion to EDV. The use of the
EF ratio must, as a fundamental assumption for accuracy, be used only when the association
between numerator and denominator is directly proportional in nature. This incompatibility of the
EF ratio has far-reaching implications, including the potential for biasing clinical and physiological
insights into the human left ventricular systolic function. Specifically, estimates of relative SV are
biased low for larger EDV measures, and vice versa. We contend that, although the EF ratio index
is simple to calculate, it can contribute to misdiagnoses in heart failure (Figure 2 a, b). Of the 23
patients who were found to have a reduced EF in the TOPCAT sample, 5 of these patients (22%)
were misclassified. In fact, the mean difference of 3.6% (95%CI: 2.6% to 4.5%) between the
ratiometric and allometrically-adjusted EF estimates indicated that these patients had a mid-range
EF. We also highlight the fact that, in the TOPCAT study, HFpEF patients were specifically
recruited (Desai et al., 2011; Pitt et al., 2014; Solomon et al., 2016). As a consequence, only
approximately 3% of the patients in the TOPCAT sample had a reduced EF. This proportion would
be substantially larger in a random sample of HF patients, as would the range of measured EDVs.

For example, in the PREVEND study, at a median follow-up of 11.5 years, the reported proportion
of patients with HFrEF was 66% (Brouwers et al., 2013). Therefore, a “reduced” misclassification
proportion of 22% could have wider ramifications in a random sample of HF patients.

In both the MESA and TOPCAT samples, the AIC criteria indicated that the two-parameter power
function with normal, heteroscedastic error was the superior model for describing left ventricular
systolic function rather than the EF ratio model of straight line with zero intercept (Supplemental
Table 1 and 2). The EDV scaling exponents observed in both MESA and TOPCAT samples described unambiguously the negative allometric relationship ($b < 1$) between the volume of blood pumped from the ventricle during each cardiac cycle and atrial filling at the end of the diastolic phase both in preclinical and overt heart failure patients (Packard 2017). A simple ratio would have empirical and physiological support only if these allometric exponents were found to be equivalent to 1. Furthermore, as a potential solution to the scaling problems with EF ratio, the present study provides a novel approach to derive EF measures adjusted properly for EDV differences working in the raw arithmetic space and using the residuals from the best model (Albrecht et al., 1993).

Our study findings also appeared to shed light on the reported sex differences in EF both in healthy (Chung et al., 2006; Yeon et al., 2015) and diseased (Davies et al., 2001; Martinez-Selles et al., 2012) populations, whereby women typically show a higher EF compared with men. Yeon and colleagues, who examined a large sub-population of the Framingham Heart Study Offspring Cohort (n=1794) using cardiac MRI, reported a mean (±SD) EF of 68% ± 5 in women and 66% ± 5 in men (Yeon et al., 2015). Nevertheless, the observed EDV was found to be substantially smaller in women than in men (Yeon et al., 2015). Indeed, the mean sex-based differences we observed in absolute EDV both in MESA and TOPCAT samples were in line with the current evidence (Gori et al., 2014; Salton et al., 2002; Yeon et al., 2015). The 95%CI for the mean EDV difference between men and women was 23 mL to 59 mL in MESA, and 22 mL to 30 mL in TOPCAT. On the other hand, application of allometric scaling methods revealed trivial sex-based differences in EF (Table 1). Specifically, in MESA, the observed mean difference of 5.7% (95%CI: 1.0% to 10.5%) indicated that women had a substantially greater EF ratio than men. Conversely, there was a trivial difference in EF of 1.6% (95%CI: −2.5% to 5.8%) between the sexes based on allometrically-adjusted individual EF estimates. Likewise, trivial differences in the adjusted-EF were also observed in the larger TOPCAT population. While EF ratio estimates indicated a substantial


difference of 2.6% (95%CI: 1.6% to 3.7%) between the sexes, the observed mean difference in the
adjusted-EF of 0.5% (95%CI: —0.5% to 1.5%) was again found to be trivial.

Not only did the procedures used for normalizing left ventricular SV relative to EDV unveil the
unappreciated potential of the EF ratio% to provide biased individual estimates, but they also permit
an accurate determination of properly normalized EF measures (Albrecht et al., 1993; Laird 1983)
for new clinical patients showing hallmarks akin to the reference population. Conceptually, the sum
of a new heart failure patient’s individual residual (Albrecht et al., 1993), by definition the
difference between the observed and predicted EF, and the reference MESA sample mean EF of
63.7% can provide the clinician with a size-adjusted measure of EF for the new person examined in
the clinic. The prediction equation resulting from the best model parameter estimates in MESA
(Supplemental Table 1), with the EF ratio as the dependent variable, was \( \text{EF} = 1.74298 \times \)
\( \text{EDV}^{-0.22326} \times \exp(\text{chronological age} \times 0.001831) \times \exp(\text{sex} \times -0.03121) \)
and yields a predicted estimate of EF. To illustrate further the importance of the proposed approach, we also re-examined
here the clinical case of a patient with a definite diagnosis of heart failure, known chronological
age, sex, and left ventricular functional parameters measured between 2010 and 2012 as part of the
fifth examination of MESA (Liu et al., 2013). Demographic characteristics and parameters of
cardiac function were obtained for a 76-year-old, African-American woman with no history of
myocardial infarction or coronary heart disease, hypertension, treated diabetes, a blood pressure of
149/84 mmHg, an eGFR of 30.2 mL/min/1.73m\(^2\), and on \( \beta \)-blockers therapy. Left ventricular EDV
(114 mL), ESV (51 mL), SV (63 mL), and mass (140 g) measures were within the physiological
parameters (Natori et al., 2006). While the EF ratio of 55% was substantially above the threshold
defining HFpEF, the more appropriate adjusted-EF was a lower 49% and revealed a substantial
overestimation of the true relative systolic function corresponding to 6%. A similar trend was
observed in the case a follow-up assessment of a 64-year-old Caucasian man with a definite
diagnosis of heart failure in MESA (Liu et al., 2013). The patient presented a history of myocardial
infarction, coronary heart disease, hypertension, sinus bradycardia, treated diabetes, a blood pressure of 143/72 mmHg, an eGFR of 77.9 mL/min/1.73m², and was on β-blockers therapy. Left ventricular EDV (235 mL), ESV (142 mL), and mass (233 g) measures were substantially elevated, whereas the observed SV (93 mL) was within the physiological parameters (Natori et al., 2006). While the EF ratio of 39% allegedly suggested an HFrEF diagnosis (Ponikowski et al., 2016), the more appropriate allometrically adjusted-EF was a higher 47% and revealed a substantial underestimation of the true relative systolic function corresponding to 8%. From a clinical standpoint, the approach described herein is deemed superior to the traditional formulation of power-function ratios ($Y/X^b$), which typically display distributional patterns dependent on the size of the scaling variable (Albrecht et al., 1993).

In a failing heart, it is well-established that changes in EDV are likely to affect EF to a much greater extent than potential differences in SV, which typically tend to be of a smaller magnitude (Cohn et al., 2000). With use of the traditional EF ratio index, substantial and uncontrolled variations in EDV have the unappreciated potential of generating artefactual variability in the estimated amount of fractional volume that is ejected during each cardiac cycle, regardless of the observed SV (Konstam 2003). A landmark study on the pathophysiological characterization of heart failure revealed trivial differences in SV between patients with chronic heart failure and healthy controls (Kitzman et al., 2002). In contrast, the mean EF was substantially higher in people with HFpEF compared with the observed values in both HFrEF patients and, paradoxically, healthy participants (Kitzman et al., 2002). Similarly, the mean EF was found to be larger in patients with left ventricular hypertrophy than healthy individuals despite significantly smaller left ventricular chamber dimensions (Aurigemma et al., 1995). Additionally, a recent study has demonstrated the unappreciated impact of geometric confounders, primarily increased wall thickness and reduced EDV, hindering a reliable interpretation of EF (Stokke et al., 2017). Despite significant and proportional reductions in SV and, more importantly, EDV which could result in a preserved EF, global longitudinal and
circumferential strain can yet be substantially impaired (Stokke et al., 2017). This line of evidence, alongside our study findings (Figure 1), appears to underline further the potential inadequacy of a EF ratio for stratifying cardiovascular patients since, for example, the development of an increased relative wall thickness could allow a preserved EF irrespective of a depressed myocardial shortening (Aurigemma et al., 1995). Since the physiological range of SV is finite, any substantial increase in EDV would result in a consequent inflation of the ESV and concomitant reduction of EF or vice versa (Li 1996). In relative terms, lack of adequate control for pathophysiological changes in cardiac morphology influencing left ventricular cavity volume in diastole can bias the EF ratio and, ultimately, lead to misclassifying a patient’s clinical profile (Konstam 2003). Furthermore, the seldom appreciated drawbacks of adopting a ratiometric scaling approach may also provide an index of relative systolic function spuriously labile to any variation in preload and afterload (Carabello 2002; Kalogeropoulos & Butler 2017). When SV is appropriately scaled to EDV using allometric methods, the confounding effects of EDV differences can be therefore removed and allow clinically meaningful inter-individual and group comparisons.

Limitations

Notwithstanding the fact that we examined the scaling relationship between SV and EDV among both preclinical and chronic heart failure patients, missing observations of cardiac structure and function of patients with acute decompensated heart failure limit a general application of the observed outcomes for taxonomic classifications in the ensuing stages of this pathological disorder. Additionally, the adoption of different imaging techniques for the assessment of left ventricular volumes in MESA and TOPCAT could be another limitation of the present study, even though the point estimates of the EDV exponents were not found to be substantially different between the samples (Figure 2 a, b). Finally, the distribution of EF frequencies, and implicitly the relatively small left ventricular volumes, might have influenced the precision of the point estimate for the EDV allometric exponent due to the substantially greater proportion of participants with a EF ratio
These results appear to warrant further research investigating the scaling properties of the EF index using allometric methods in large samples of acute and chronic heart failure patients being heterogeneous in left ventricular size, and the related implications from clinical and epidemiological perspectives.

Conclusions

Ratio scaling appears to limit the validity of EF as the traditional measure of the human systolic function unless it is adequately normalized for differences in EDV. The residual size correlation of a EF ratio might preclude a clinically meaningful assessment of cardiac function, ultimately yielding substantially biased estimates of EF for some individuals. A comprehensive integration of absolute measures of the heart function (i.e., left ventricular ESV), clinical parameters, and relevant biomarkers might embody a more pragmatic approach for the optimal pre-emptive screening, decision-making, and therapeutics than the limited scrutiny of a ratiometric EF index failing to serve its purpose in an unbiased manner. Further research will be required to examine scaling properties of the EF% index within large, heterogeneous populations of healthy and diseased individuals for determining the construct validity of the index as a clinical biomarker for risk stratification and therapeutic decisions.

Disclosures

The authors have no conflict of interest to disclose regarding this publication.

Sources of funding

None


Borlaug BA and Redfield MM. Diastolic and systolic heart failure are distinct phenotypes within the heart failure spectrum. *Circulation* (2011); 123: 2006-2013.


Carabello BA. Evolution of the study of left ventricular function - Everything old is new again. *Circulation* (2002); 105: 2701-2703.
Chung AK, Das SR, Leonard D, et al. Women have higher left ventricular ejection fractions than men independent of differences in left ventricular volume - The Dallas Heart Study. *Circulation* (2006); **113**: 1597-1604.


Goldberg LR and Jessup M. Stage B heart failure: management of asymptomatic left ventricular systolic dysfunction. *Circulation* (2006); **113**: 2851-2860.


Meta-analysis Global Group in Chronic Heart F. The survival of patients with heart failure with preserved or reduced left ventricular ejection fraction: an individual patient data meta-analysis. *Eur Heart J* (2012); 33: 1750-1757.


Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* (2016); 37: 2129-2200.


Table Legends

Table 1. Summary data for the study participants in MESA and TOPCAT stratified by sex.

Figure Legends

Figure 1. Scatterplots showing the inverse relationship between the ratiometric EF and EDV in MESA (a), \( r = -0.67 \) (95%CI: \(-0.76\) to \(-0.55\)) and TOPCAT (b), \( r = -0.41 \) (95%CI: \(-0.46\) to \(-0.35\)), and linear bivariate relationship between SV and EDV in MESA (c), \( r = 0.68 \) (95%CI: \(0.57\) to \(0.77\)), Y-intercept = 44 mL (95%CI: 34 mL to 54 mL) and TOPCAT (d), \( r = 0.88 \) (95%CI: \(0.86\) to \(0.89\)), Y-intercept = 13 mL (95%CI: 11 mL to 14 mL).

Figure 2. Scatterplots showing the allometric relationship SV and EDV from the multivariable model, \( SV = 1.75 \cdot EDV^{0.78} \) in MESA (a), and \( SV = 1.22 \cdot EDV^{0.86} \) in TOPCAT (b), and raw residuals against the predicted values from the 2-parameter power function with normal, heteroscedastic error in MESA (c) and TOPCAT (d) samples.

Supplemental Material Legends

Supplemental Table 1. Statistical models fitted to untransformed data for left ventricular stroke volume and end-diastolic volume in the MESA.

Supplemental Table 2. Statistical models fitted to untransformed data for left ventricular stroke volume and end-diastolic volume in the TOPCAT echocardiographic sub-study.
<table>
<thead>
<tr>
<th>Variable</th>
<th>MESA</th>
<th>TOPCAT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men (n = 75)</td>
<td>Men (n = 433)</td>
</tr>
<tr>
<td></td>
<td>Women (n = 37)</td>
<td>Women (n = 431)</td>
</tr>
<tr>
<td>Age, y</td>
<td>68.3 ± 8.0</td>
<td>69.5 ± 9.5</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>85.6 ± 15.6</td>
<td>95.8 ± 21.8</td>
</tr>
<tr>
<td>Height, cm</td>
<td>173.4 ± 7.6</td>
<td>174.2 ± 8.2</td>
</tr>
<tr>
<td>Waist circumference, cm</td>
<td>102.2 ± 11.8</td>
<td>107.3 ± 16.2</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>28.4 ± 4.4</td>
<td>31.4 ± 6.2</td>
</tr>
<tr>
<td>Diabetes mellitus, n (%)</td>
<td>18 (24)</td>
<td>176 (41)</td>
</tr>
<tr>
<td>eGFR, mL/min/1.73m²</td>
<td>72.2 ± 20.6</td>
<td>69.1 ± 20.6</td>
</tr>
<tr>
<td>Systolic blood pressure, mmHg</td>
<td>135 ± 20</td>
<td>127 ± 14</td>
</tr>
<tr>
<td>Diastolic blood pressure, mmHg</td>
<td>74 ± 11</td>
<td>73 ± 11</td>
</tr>
<tr>
<td>Heart rate, beats/min</td>
<td>64 ± 11</td>
<td>68 ± 12</td>
</tr>
<tr>
<td>LV mass, g</td>
<td>206 ± 54</td>
<td>244 ± 68</td>
</tr>
<tr>
<td>LV end-diastolic volume, mL</td>
<td>163 ± 49</td>
<td>112 ± 34</td>
</tr>
<tr>
<td>LV end-systolic volume, mL</td>
<td>66 ± 39</td>
<td>48 ± 22</td>
</tr>
<tr>
<td>LV stroke volume, mL</td>
<td>97 ± 23</td>
<td>64 ± 17</td>
</tr>
<tr>
<td>Adjusted LV stroke volume, mL</td>
<td>90 ± 15</td>
<td>58 ± 8</td>
</tr>
<tr>
<td>Ratiometric EF, (%)</td>
<td>61.8 ± 11.5</td>
<td>58.3 ± 8.4</td>
</tr>
<tr>
<td>Normalized EF, (%)</td>
<td>63.1 ± 10.3</td>
<td>59.4 ± 8.1</td>
</tr>
<tr>
<td>Normalized EF, mL/mL (%)</td>
<td>186.4 ± 30.4</td>
<td>111.8 ± 15.2</td>
</tr>
</tbody>
</table>

Data are presented as mean ± SD for continuous variables, and frequency or percentages for categorical variables. #: 2-parameter power function with normal, heteroscedastic error; †: power function ratio; ‡: indicates missing observations; EF: left ventricular ejection fraction; BMI: body-mass index; eGFR: estimated glomerular filtration rate; LV: left ventricular. The normalized parameters of systolic function (footnote †) were derived directly from the model residuals working in the raw arithmetic data space, with the ratiometric EF or LV stroke volume as the dependent variable and LV end-diastolic volume, chronological age, and sex as predictors. Each participant’s residual was added to the predicted mean ratio at the mean LV end-diastolic volume in the whole sample, to obtain an adjusted EF or LV stroke volume free from the influence of LV end-diastolic volume (Albrecht et al., 1993; Laird 1983). The normalized index (footnote †) was directly derived from the ratio of LV stroke volume to end-diastolic volume raised to the power of 0.78 and 0.86 in the MESA and TOPCAT samples, respectively.
**Supplemental Table 1.** Statistical models fitted to untransformed data for left ventricular stroke volume and end-diastolic volume in the MESA

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight line, no intercept, with normal, homoscedastic error</td>
<td>959.9</td>
<td>50.4</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Straight line, intercept, with normal, homoscedastic error</td>
<td>956.6</td>
<td>47.2</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Three-parameter power function with normal, homoscedastic error</td>
<td>949.4</td>
<td>40.0</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Failed to converge. Re-arranged, convergence criterion changed to 0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-parameter power function with normal, homoscedastic error</td>
<td>948.3</td>
<td>38.8</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Straight line, no intercept, with lognormal heteroscedastic error</td>
<td>945.0</td>
<td>35.6</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Straight line, intercept, with lognormal heteroscedastic error</td>
<td>942.1</td>
<td>32.6</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Two-parameter power function with lognormal, heteroscedastic error</td>
<td>930.9</td>
<td>21.4</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Three-parameter power function with lognormal, heteroscedastic error</td>
<td>922.8</td>
<td>13.3</td>
<td>weak support</td>
</tr>
<tr>
<td>Failed to converge. Re-arranged, convergence criterion changed to 0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Straight line, intercept, with normal, heteroscedastic error</td>
<td>918.8</td>
<td>9.4</td>
<td>weak support</td>
</tr>
<tr>
<td>Straight line, no intercept, with normal, heteroscedastic error</td>
<td>917.6</td>
<td>8.2</td>
<td>weak support</td>
</tr>
<tr>
<td>Three-parameter power function with normal, heteroscedastic error</td>
<td>915.1</td>
<td>5.7</td>
<td>plausible alternative</td>
</tr>
<tr>
<td>Failed to converge. Re-arranged, convergence criterion changed to 0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-parameter power function with normal, heteroscedastic error</td>
<td>909.4</td>
<td>0</td>
<td>best</td>
</tr>
</tbody>
</table>

AIC = Akaike’s information criterion; ΔAIC = Akaike difference
### Supplemental Table 2. Statistical models fitted to untransformed data for left ventricular stroke volume and end-diastolic volume in the TOPCAT echocardiographic sub-study

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-parameter power function with normal, homoscedastic error</td>
<td>6236.8</td>
<td>422.8</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Failed to converge. Convergence criterion changed to 0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Straight line, no intercept, with normal, homoscedastic error</td>
<td>6183.8</td>
<td>369.7</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Straight line, intercept, with normal, homoscedastic error</td>
<td>6129.7</td>
<td>315.6</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Two-parameter power function with normal, homoscedastic error</td>
<td>6093.7</td>
<td>279.6</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Straight line, no intercept, with lognormal heteroscedastic error</td>
<td>6024.0</td>
<td>210.0</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Three-parameter power function with normal, heteroscedastic error</td>
<td>6008.2</td>
<td>194.2</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Failed to converge. Convergence criterion changed to 0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Straight line, intercept, with lognormal heteroscedastic error</td>
<td>5989.8</td>
<td>175.8</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Two-parameter power function with lognormal, heteroscedastic error</td>
<td>5958.3</td>
<td>144.2</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Three-parameter power function with lognormal, heteroscedastic error</td>
<td>5917.5</td>
<td>103.4</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Straight line, no intercept, with normal, heteroscedastic error</td>
<td>5871.5</td>
<td>57.4</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Straight line, intercept, with normal, heteroscedastic error</td>
<td>5842.5</td>
<td>28.5</td>
<td>no empirical support</td>
</tr>
<tr>
<td>Two-parameter power function with normal, heteroscedastic error</td>
<td>5814.1</td>
<td>0</td>
<td>best</td>
</tr>
</tbody>
</table>

AIC = Akaike’s information criterion; ΔAIC = Akaike difference