REVIEW

Our obsession with normal values

Frank A Flachskampf MD PhD[1] and Petros Nihoyannopoulos MD PhD[2,3]

[1]Department of Medical Sciences, Clinical Physiology and Cardiology, Akademiska, Uppsala University, Uppsala, Sweden
[2]Imperial College London, NHLI, Hammersmith Hospital, London, UK
[3]Cardiology, Athens University, Athens, Greece

Correspondence should be addressed to F A Flachskampf: frank.flachskampf@medsci.uu.se

Abstract

Normal values provide the background for interpretation of quantitative imaging data and thus are essential information for daily routine. Nevertheless, the ways how normal values are obtained, presented and interpreted, often do not receive the attention they deserve. We review the concepts of normalcy, the implications of typical normal ranges including the types of distribution of normal data, the possibilities to index for confounding biological factors like body surface area and the limitations of the very concept of normal values, demonstrating that there are no easy statistical solutions for difficult clinical problems.

Key Words
- normal values
- multimodality
- guidelines
- statistics

Introduction

In echocardiography, as indeed in all cardiac imaging, normal values form the foundation for interpreting quantitative parameters. This applies to morphologic continuous parameters such as linear, area or volume measurements (including ventricular mass, which is calculated from myocardial volume), as well as Doppler or speckle-tracking based blood or tissue velocities and derived complex parameters such as strain or strain rate. Most laboratories have normal value tables hanging at a well visible spot, and compilations of such values are among the most often looked up or downloaded documents in cardiac imaging (1, 2, 3, 4).

While it seems obvious that we can only identify pathologic values if we know the normal range of a parameter, it is easy to become overconfident in the utility and wisdom of normal values. This applies in particular to imagers at the beginning of their learning curves or cardiologists whose main focus is not on imaging. In the following, we review the rationale and practical application of normal values in cardiac imaging with respect to underlying assumptions and clinical relevance. We focus here on the use of normal values in echocardiography, but the same arguments can be made for all imaging modalities.

Underlying concepts

Normal individuals

Normal values, by definition, must be gleaned from a population of 'normal', healthy individuals. Since these, again by definition, tend not to undergo cardiac imaging, early normal value compilations were derived from studies of individuals who appeared to be sufficiently unsuspicious of cardiac disease – often laboratory personnel, researchers, students and their relatives. Sometimes, these were individuals sent to the laboratory for imaging but found not to have detectable cardiac disease, such as persons with 'functional' murmurs, persons undergoing check-ups without pathologic findings and others. Note that this, strictly speaking, involves a logical circle, since normal values are derived from persons deemed to have normal findings, which implies that some notion of normalcy already exists. Further, such informal criteria result in a

www.echorespract.com
https://doi.org/10.1530/ERP-17-0082
© 2018 The authors
Published by Bioscientifica Ltd

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

clear selection of persons: for example, often volunteers are recruited from hospital or academic personnel who typically do not represent the age range and other features of the general population. Such 'convenience samples' of individuals cannot be really representative of the general (healthy) population, although the latter is the collective we ultimately would wish to have as our standard of comparison. Nevertheless, many local 'normal value' tables just reflect such samples, often of only few individuals with very limited age range and unbalanced sex distribution. Selection bias in the recruitment of participants in a normal sample leads to loss of generalizability or 'external validity': what is true for the biased sample may not be true for the general (or a particular) population.

To address this problem, meta-analysis studies have been performed utilizing large compilations of normal data from diverse populations, ethnicities and age ranges, for example the EchoNoRMAL study (5, 6) analyzing data from over 22,404 'normal' persons. However, although based on person-level data, these data were acquired all over the world, with corresponding variability between the original readers' habits and biases. These concerns could only be addressed in a future large-scale, prospective core-study using a core-laboratory for uniform analysis.

While selection bias can be circumvented by more careful though cumbersome recruitment procedures, some difficult questions concerning the normalcy of a normal collective remain, most importantly: who is really healthy? Cardiovascular health is obviously not universally given in the general population. For example, about half of the general population is expected to develop arterial hypertension during their lifetime. Further, the average body mass index of the general population deviates considerably from what is believed to be 'healthy' (i.e., predictive of the longest life expectancy). Similar concerns apply for blood lipids and blood glucose. Even if we define explicitly the 'healthy controls' as persons in whom manifest cardiovascular disease and risk factors (such as hypertension or obesity) are excluded, thorny questions remain. A particularly difficult conundrum is age dependency. It seems straightforward that there is a decline in exercise capacity with age, but this decline can be blunted or halted by physical training. For example, diastolic left ventricular function, measured invasively or non-invasively, exhibits an age-dependent decline, with a well-described age dependency of echo parameters of diastolic function. On the other hand, that decline is much less in aged persons who continue in a highly trained state (7). For a non-cardiac example of a nearly universal, age-dependent pathology, osteoporosis comes to mind. What is normal, then?

Which populations should be separately tabulated? Which parameters should be indexed and how?

Human beings, including healthy individuals, differ by height, weight, sex, ethnicity and many other features, for example, body fat/lean mass. Ignoring these fundamental properties affecting most if not all measurable variables may obscure or falsely create pathological measurement values. Thus, to the chagrin of cardiac imagers, normal values tables went from a simple, all-encompassing master table to a multitude of sub-tables allowing for these underlying differences, often excused by citing ubiquitous digital memory and computing capacity. Publication activity has thrived on addressing normal values for specific populations, often defined by geographic region. Some of the differences found for example between the sexes or different ethnicities are quite small, but how much is negligible and by what standard?

To account for categorical variables such as sex or ethnicity we use separate normal tables for the same parameters, e.g. left ventricular volume or mass. Age, although a continuous parameter, is also typically categorized into age ranges, for which separate normal values are established. Furthermore, body size is a major determinant of cardiovascular dimensions. This factor is accounted for by dividing the parameter by an index of body size, an operation called 'normalizing', indexing or scaling. For example, left ventricular volumes are usually presented as indices (end-systolic and end-diastolic left ventricular volume index or stroke volume index) by dividing by body surface area. The rationale is that the division by body surface area will remove the dependence of the left ventricular volume on individual body size and therefore make it easier to detect pathology. However, surprisingly little consensus exists which morphological parameters need such 'normalization' and which index of body size should be used for this. At present, linear measures, for example, the left ventricular diameters, are not routinely indexed, while volumes, for example, chamber volumes, are. Area measurements, such as stenotic valvular areas, are sometimes indexed, especially if the patient in question is uncommonly small or large. The reason for this inhomogeneity is quite simple – given that there is an influence of body size on cardiac dimensions, this influence will be stronger on volumes, which can be thought of as linear measurements to the third power, than on linear measurements, and areas (linear measurements to the second power) fall in between. An important factor that has prevented universal use of indexing is that most literature linking measurements (for example, aortic diameters) to prognosis is based on un-indexed measurements.

www.echorespract.com
https://doi.org/10.1530/ERP-17-0082
© 2018 The authors
Published by Bioscientifica Ltd

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

For the thorny issue of how to normalize measurements to overall body size measures such as surface, height, weight and their powers (allometric scaling), we refer to the excellent recent overview by Oxborough *et al*. (6) and others (8). In general, it would seem best to normalize in a way respecting geometric similarity, i.e., normalizing linear measurements to linear body size (height), area measurements to for example, surface body area and volume or mass measurements to body mass or volume. Unfortunately, both body mass and body surface area are substantially influenced by obesity (the latter since body surface area is calculated from nomograms using body height and weight), which in itself should only have minor influence of cardiovascular dimensions. Again, the ideal solution, normalization for fat-free mass, is cumbersome since this parameter is not readily available.

How should normal values be presented?

The simplest way of defining the limits of normalcy is to, quite literally, state the upper and lower value limits of a parameter in a sample of the normal population: the normal range. Since this is obviously very sensitive to the concrete normal sample chosen, other standard formulations are often used. Most often, it is assumed that the parameter in question among the population will follow a Gaussian ('bell-shaped') distribution, indeed a 'normal distribution' (Fig. 1). In an ideal normal distribution, the interval of 1 standard deviation (SD) below and above the mean value will comprise 68%, of ±2 SD's 95%, and of ±3 SD's 99.7% of all values. Habitually, the interval of ±2 SD's from the mean is used as the definition of normalcy for a normally distributed continuous variable. While appealing in its simplicity and elegance, we need to keep in mind the following limitations:

1. Not all measurements in a normal population follow a Gaussian (normal) distribution, and parameters can follow a multimodal or skewed distribution. This is for example the case for coronary calcifications as measured by the Agatston score (9). A more flexible, but less convenient way of establishing a normal range is referencing quantile cut-offs, for example, the range of a parameter excluding the 2.5% or 5% lowest or highest values ('percentiles') of a normal sample and then identifying the range of the remaining data points. This type of analysis is entirely independent of the actual distribution of the data and does not require that it be normally distributed.
2. By the definition of the 95% interval, 5% of a normal population would fall into the abnormal (2.5% too low or 2.5% too high) ranges. This is a relatively large number especially when considering large populations, such as the number of healthy individuals seen at a busy echo laboratory. For a given parameter, say, left ventricular end-diastolic diameter, we can expect that 5 in 100 truly 'normal' individuals would be falsely labeled as abnormal if one follows this definition. Further, since we evaluate more than one parameter in each echo exam, the consequences are even more dire. If we evaluate 20 quantitative parameters per study (and if they are normally distributed in the healthy population), and use 95% intervals as normal ranges for these parameters, we would 'in the long run' obtain on average one falsely pathologic measurement in each healthy patient we examine, even if our measurements are perfect, just by the definition of the 95% interval. Therefore, in practice, clinical uncertainty persists at the borders of such normalcy ranges. Of course, one could use instead ±1 SD, thus excluding in the long run 32% of true normals, or ±3 SD's, which is nearly identical with the normal range, but thus does not provide a 'security margin' against pathological values.

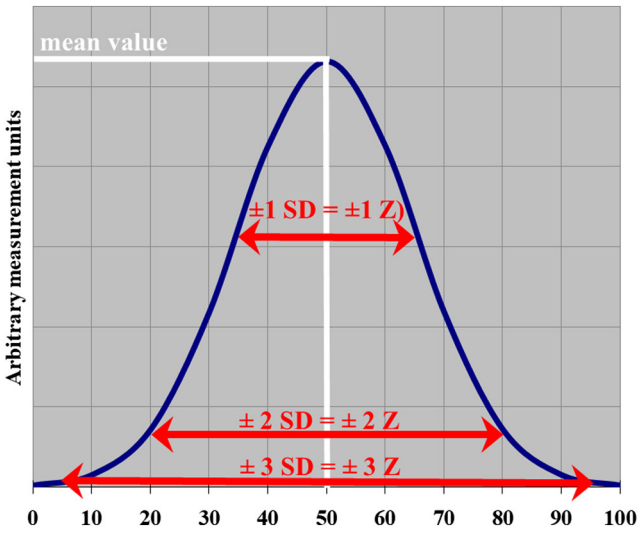


Figure 1
Example graph of a normally distributed variable. The *x*-axis displays the cumulative probability in percent of the variable values, which are shown on the *y*-axis. All values (cumulative probability of 100%) lie under the bell-shaped curve. Fifty percent of values are lower and 50% are higher than the mean value (mean and median of a normal distribution are the same). The SD describes the breadth of the curve. The intervals corresponding to ±1, ±2, and ±3 SD's (SD, or *Z*-scores; red horizontal double arrows), encompass 68, 95, and 99.7 of all values, respectively.

Handling mean and SD can be simplified by using *Z*-scores, which have gained popularity in particular in pediatric cardiology, where rapidly and substantially changing

www.echorespract.com
https://doi.org/10.1530/ERP-17-0082
© 2018 The authors
Published by Bioscientifica Ltd

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

body sizes in infants and adolescents must be dealt with (10). *Z*-scores are based on mean value and SD of a given parameter, which can be indexed to a variable such as body surface area, height or powers of such variables (11). The formula for the *Z*-score of a particular measurement x is $Z\text{-score} = (x - x_{mean})/SD$, where x_{mean} is the mean value and SD is the standard deviation of that parameter in a population of normals of that age and/or body size. A *Z*-score of 1 is the interval from x_{mean} to $x_{mean} + 1$ SD, etc. 95% of normal values thus fall within *Z*-scores of ±2. Besides the assumption of a normal distribution of the values, there is also the underlying assumption of 'homoscedasticity' (or absence of heteroscedasticity), which means that variance itself should be independent of the absolute values of the variable under consideration. This is often not the case in reality: for example, in a study of normal values for Doppler variables in children, variance increased with age (12). Another important limitation of *Z*-scores is that calculation mostly requires access to a computer.

Thus, the presentation of normal values should contain a precise definition of the measurement protocol (e.g. left atrial area by biplane Simpson's rule). If data are normally distributed they should be described by mean and SD, after stratifying or indexing for factors with proven independent influence, such as age, sex, ethnicity, measures of body size, etc. The limit here is how complex and extensive the data are allowed to be presented. Variability can instead be described in terms of *Z*-scores. If data are not normally distributed, the most straightforward manner is to present them as medians with quantiles, for example, quartiles or percentiles.

The challenge of different modalities and techniques

Imaging parameters can be obtained by different imaging modalities. This regularly leads to systematic differences in values for the same parameter, for example, for cardiac chamber volumes including left ventricular volumes. For a number of reasons, cardiac chamber volumes measured by magnetic resonance are substantially larger than by echocardiography, and cardiac computed tomography volumes are again larger than magnetic resonance volumes (13). These differences are systematic, i.e., not due to errors but inherent to methodology, and ultimately irreconcilable.

Furthermore, within one modality there often exist several ways or protocols to measure certain parameters. In echocardiography, for example, left atrial size can be measured in at least 5 ways:

- as a linear antero-posterior diameter in the parasternal long-axis view;
- as an area, usually in the apical four-chamber view;
- as a volume using an area-length method (monoplane or biplane);
- as a volume using Simpson's rule (monoplane or biplane);
- as a volume using 3D echo;

each of these methods giving (slightly) different results.

Thus, clearly, claiming that one modality is superior to another is rather absurd as each is using different techniques for measuring the same thing, whether by using ultrasound reflections (echo), proton spins (magnetic resonance imaging) or positron-electron collisions (positron emission tomography); however, none of them is closer to the 'real heart' than the other.

Finally, technical issues such as hardware and software from different manufacturers (or even different product generations from the same manufacturer) may introduce non-negligible measurement variability. An important example of this is the well-appreciated manufacturer dependency of echocardiographic strain calculations.

While some of the cited differences have to be acknowledged as systematic, unavoidable biases necessitating separate sets of normal values, it is nevertheless important to avoid comparing apples and pears by unifying measurement conventions and protocols. The lack of such unified conventions has been particularly evident in the basic and simple measurement of diameters of the ascending aorta. Practically all conventions have been recommended at some point by some professional society, from inner-edge to inner-edge to outer-edge to outer-edge and from systolic to diastolic measurement (14, 15). Predictably, different measurement protocols will lead to artificial measurement discrepancies, further leading to confusion and potentially adverse management decisions. Occasionally, cut-offs used and validated in studies may influence the definitions of normalcy and abnormality. Clearly, more cooperative effort for standardization across modalities is necessary here for the benefit of patients and imagers alike.

Conclusion

Normal values are indispensable, yet, problematic. They appear to provide certainty where often in reality we cannot be sure. While the literature presents an increasingly complex body of tabulated data, one should be aware of the fundamental limitations of the notion and

www.echorespract.com
https://doi.org/10.1530/ERP-17-0082
© 2018 The authors
Published by Bioscientifica Ltd

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

notation of normal values. Difficult clinical judgments (e.g., interpretation of an aortic diameter in a very small or very large person) are not easily solved by normal value compilations. We should use normal values with caution and understand them not as normative, but as an orientation. Furthermore, more collaborative efforts are needed to harmonize and unify measurement protocols and conventions across imaging modalities but accepting the inherent differences for each modality.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding
This work did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References

1 Dalen H, Thorstensen A, Aase SA, Ingul CB, Torp H, Vatten LJ & Stoylen A. Segmental and global longitudinal strain and strain rate based on echocardiography of 1266 healthy individuals: the HUNT study in Norway. *European Journal of Echocardiography* 2010 **11** 176–183. (https://doi.org/10.1093/ejechocard/jep194)

2 Kou S, Caballero L, Dulgheru R, Voilliot D, De Sousa C, Kacharava G, Athanassopoulos GD, Barone D, Baroni M, Cardim N, *et al*. Echocardiographic reference ranges for normal cardiac chamber size: results from the NORRE study. *European Heart Journal: Cardiovascular Imaging* 2014 **15** 680–690. (https://doi.org/10.1093/ehjci/jet284)

3 Caballero L, Kou S, Dulgheru R, Gonjilashvili N, Athanassopoulos GD, Barone D, Baroni M, Cardim N, Gomez de Diego JJ, Oliva MJ, *et al*. Echocardiographic reference ranges for normal cardiac Doppler data: results from the NORRE Study. *European Heart Journal: Cardiovascular Imaging* 2015 **16** 1031–1041. (https://doi.org/10.1093/ehjci/jev083)

4 Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, *et al*. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *European Heart Journal: Cardiovascular Imaging* 2015 **16** 233–270. (https://doi.org/10.1093/ehjci/jev014)

5 Echocardiographic Normal Ranges Meta-Analysis of the Left heart (EchoNoRMAL) Collaboration. A meta-analysis of echocardiographic measurements of the left heart for the development of normative reference ranges in a large international cohort: the EchoNoRMAL study. *European Heart Journal: Cardiovascular Imaging* 2014 **15** 341–348. (https://doi.org/10.1093/ehjci/jet240)

6 Echocardiographic Normal Ranges Meta-Analysis of the Left Heart Collaboration. Ethnic-specific normative reference values for echocardiographic LA and LV size, LV mass, and systolic function: the EchoNoRMAL study. *JACC: Cardiovascular Imaging* 2015 **8** 656–665. (https://doi.org/10.1016/j.jcmg.2015.02.014)

7 Prasad A, Popovic ZB, Arbab-Zadeh A, Fu Q, Palmer D, Dijk E, Greenberg NL, Garcia MJ, Thomas JD & Levine BD. The effects of aging and physical activity on Doppler measures of diastolic function. *American Journal of Cardiology* 2007 **99** 1629–1636. (https://doi.org/10.1016/j.amjcard.2007.01.050)

8 Batterham A, George K, Whyte G, Sharma S & McKenna W. Scaling cardiac structural data by body dimensions: a review of theory, practice and problems'. *International Journal of Sports Medicine* 1999 **8** 495–502. (https://doi.org/10.1055/s-1999-8844)

9 Pletcher MJ, Sibley CT, Pignone M, Vittinghoff E & Greenland P. Interpretation of the coronary artery calcium score in combination with conventional cardiovascular risk factors: the Multi-Ethnic Study of Atherosclerosis (MESA). *Circulation* 2013 **128** 1076–1084. (https://doi.org/10.1161/CIRCULATIONAHA.113.002598)

10 Lopez L, Colan SD, Frommelt PC, Ensing GJ, Kendall K, Younoszai AK, Lai WW & Geva T. Recommendations for quantification methods during the performance of a pediatric echocardiogram: a report from the Pediatric Measurements Writing Group of the American Society of Echocardiography Pediatric and Congenital Heart Disease Council. *Journal of the American Society of Echocardiography* 2010 **23** 465–495. (https://doi.org/10.1016/j.echo.2010.03.019)

11 Oxborough D, Ghani S, Harkness A, Lloyd G, Moody W, Ring L, Sandoval J, Senior R, Sheikh N, Stout M, *et al*. Impact of methodology and the use of allometric scaling on the echocardiographic assessment of the aortic root and arch: a study by the Research and Audit Sub-Committee of the British Society of Echocardiography. *Echo Research and Practice* 2014 **1** 1–9. (https://doi.org/10.1530/ERP-14-0004)

12 Colan SD. The why and how of Z scores. *Journal of the American Society of Echocardiography* 2013 **26** 38–40. (https://doi.org/10.1016/j.echo.2012.11.005)

13 Dallaire F, Slorach C, Hui W, Sarkola T, Friedberg MK, Bradley TJ, Jaeggi E, Dragulescu A, Har RL, Cherney DZ & Mertens L. Reference values for pulse wave Doppler and tissue Doppler imaging in pediatric echocardiography. *Circulation: Cardiovascular Imaging*. 2015 **8** e002167. (https://doi.org/10.1161/CIRCIMAGING.114.002167)

14 Sugeng L, Mor-Avi V, Weinert L, Niel J, Ebner C, Steringer-Mascherbauer R, Schmidt F, Galuschky C, Schummers G, Lang RM, *et al*. Quantitative assessment of left ventricular size and function: side-by-side comparison of real-time three-dimensional echocardiography and computed tomography with magnetic resonance reference. *Circulation* 2006 **114** 654–661. (https://doi.org/10.1161/CIRCULATIONAHA.106.626143)

15 Asch FM, Yuridetsky E, Prakash SK, Roman MJ, Weinsaft JW, Weissman G, Weigold WG, Morris SA, Ravekes WJ, Holmes KW, *et al*. The need for standardized methods for measuring the aorta: multimodality core lab experience from the National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovasculac Conditions (GenTAC). *JACC: Cardiovascular Imaging* 2016 **9** 219–226. (https://doi.org/10.1016/j.jcmg.2015.06.023)

16 Flachskampf FA. How exactly do you measure that aorta? Lessons from multimodality imaging (editorial). *JACC: Cardiovascular Imaging* 2016 **9** 227–229. (https://doi.org/10.1016/j.jcmg.2015.07.017)

Received in final form 23 February 2018
Accepted 21 March 2018
Accepted Preprint published online 21 March 2018

© 2018 The authors
Published by Bioscientifica Ltd

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.