Interpreting (and misclassifying) fitness – what do we really mean by 'normal'?

Dr Owen W Tomlinson – University of Exeter Medical School
Professor Craig A Williams FBASES – Sport & Health Science, University of Exeter

---

Owen is an Associate Lecturer at the University of Exeter Medical School, and recipient of the 2022 BASES International Conference Grant.
Craig is a Professor of Paediatric Physiology and Health at the University of Exeter, and Director of the Children's Health & Exercise Research Centre.

"Is that a good result?"

"Is that normal?"

"What should it be?"

How many of us have heard these questions when undertaking testing? Whether it be from athletes and players, patients, clients, study participants, students, or even curious admirers of our testing. No matter the parameter in question, or the setting, it is a valid question whose answer should be simple but is often masked with some uncertainty.

Measuring and Interpreting Fitness

As exercise scientists, we will measure fitness in numerous ways, depending on the participant and aim of the testing session. When utilising fitness testing for people with complex medical conditions (CMC), the use of cardiopulmonary exercise testing, with peak oxygen uptake (VO_{2peak}) as a primary outcome, tends to be favoured in clinical practice. This utilisation is not only due to the establishment of valid and reliable protocols but also because this modality can simultaneously assess multiple systems (and reduces participant burden for the need of multiple tests). Importantly, the significant associations of VO_{2peak} with long-term outcomes such as quality of life, transplant risk and premature mortality provide a robust measurement of prognostic value.

However, one challenge to the uncertainty in answering the question of what is 'normal' begins in the interpretation of the data. Absolute values of VO_{2peak} (presented in $L \cdot min^{-1}$) are biased by body size, whereby larger individuals will present with higher scores by virtue of greater mass (including muscle mass). Scores that are normalised for body mass (presented in $mL \cdot kg^{-1} \cdot min^{-1}$) do not account for lean mass, and therefore individuals with high levels of muscle mass (e.g., body builders, rugby players, rowers) can be unfairly penalised and misclassified with low fitness. Therefore, scaling for fat free mass, using allometric scaling, or using normative data can be utilised to try and minimise these influences.

Normal Reference Values

Therefore, the latter of these options – using normative data – is feasible for many practitioners as this can be applied to all populations (allometry is only applicable to the explicit sample group being studied) and does not require additional physiological measurement (such as lean mass). There are several studies providing commentary and normal reference values (NRV) as to what VO_{2peak} should be for someone of a particular sex, age, and mass, with 29 different sets of NRV published from 2014-2019 alone (Takken et al., 2019).

For those working in a clinical environment, there are a plethora of guidelines and recommendations available for the conduct and interpretation of exercise testing. One such organisation, the Association for Respiratory Technology & Physiology have recently published a statement for use by clinical physiologists in the NHS. Internationally, other professional groups e.g., the American Thoracic Society & American College of Chest Physicians (ATS/ACCP), European Respiratory Society (ERS) and European Cystic Fibrosis Society (ECFS) all have guidelines applicable to a range of respiratory disorders. The latter of these are presented in Table 1, and it is shown that multiple NRV are available depending on testing modality and age group, with markedly different populations in which the NRV are derived.

However, there is discrepancy between these guidelines and how to derive and classify 'normal' fitness. In obtaining data, all studies will have taken different approaches with regards to equipment, protocols, and statistical analyses. Moreover, some will suggest use of a NRV dataset to display data as 'percent of predicted' whereas others will recommend using NRV to derive a percentile to classify abnormal function. In many documents, the seminal work of Jones et al., (1985) has gained traction as a popular NRV, as recommended by the ATS/ACCP,

ERS and ECFS; and being cited over 600 times in the literature. However, this study was published in 1985, conducted on 50 males and 50 females from 15-71 years of age, from a single laboratory in Canada. Therefore, despite its prominence in the literature base, the question of how applicable such a study is for global use, nearly 40 years later, is warranted.

Reviewing the use of NRV

A recent scoping review of 169 studies – conducted by our group – examining use of NRV in studies of cystic fibrosis (a genetic, predominantly respiratory, disease affecting ~100,000 people worldwide) has shown that 36% of studies present data as a 'percent of predicted' but do not state the NRV used. Without such reporting, we cannot be assured that the data is presented relative to the correct age groups and sex or uses the same testing modality. Moreover, only 21% of studies cite the NRV (as per Table 1) that are suggested for use (Tomlinson & Williams, 2022). Such under-reporting of NRV, and discrepancy in their use can lead to incorrect comparisons and interpretation between studies, and potentially negatively affecting clinical practice.

Applying NRV

Moreover, analysis of clinical data within our laboratory has shown that two common NRV proposed for use by the ECFS result in markedly different changes over time. As shown in Figure 1, in a group of 18 children with cystic fibrosis, only ~70% of patients present with the same direction of change when undergoing sequential annual exercise testing (Tomlinson et al., 2020). The results of this analysis convey a worrying picture for long-term interpretation of data and profiling of individual disease progression.

What is the value in using a NRV to present data when one set of values will propose someone increases their fitness by 20%, but another set of values shows it declines by 5% over the course of a year?

<u>Which do we believe?</u>

<u>How does this change treatment?</u>

<u>What do we tell the patient?</u>

The negative impacts of using NRV in a clinical setting have been described comprehensively in a case study from Waterfall *et al.*, (2020), whereby a teenager presenting with breathlessness

and chest pain underwent exercise testing at two different hospitals in England, one year apart. The two tests showed the same absolute values (1.39 $L \cdot min^{-1}$ for both tests) and remarkably similar relative values (23.5 vs. 24.5 $mL \cdot kg^{-1} \cdot min^{-1}$). However, use of NRV indicated large differences in VO_{2peak} (70 v 55%$_{Pred}$) because the two hospital teams used two different NRV for interpretation. Such a simple difference in testing strategy had a very real impact upon patient care, in this instance via a delay of a medical procedure, as it was initially assumed the patient had 'normal fitness'. Life-changing consequences could have been avoided, and can be avoided in the future, if centres align in the use of NRVs.

Summary

How we progress from here is the larger challenge, we must address collaboratively as a community comprising exercise scientists, physiologists, statisticians, educators, clinicians, and policy makers. Studies are needed to develop and validate NRV databases in larger populations; but also, the pooling of existing but newer data to replace poorly designed and implemented datasets, which provide greater confidence in what we define as 'normal'. Comparatively, we have such datasets when defining 'normal' body weight and lung function, so why not fitness?

Whilst we have highlighted some issues surrounding the interpretation of one variable in one population, the same issues associated with choice of procedures and application of NRV will apply across populations and therefore all practitioners should think carefully next time when asked "Is that a good result?"

REFERENCES

Hebestreit, H., Arets, H. G., Aurora, P., Boas, S., Cerny, F., Hulzebos, E. H., Karila, C., Lands, L. C., Lowman, J. D., Swisher, A., Urquhart, D. S., & European Cystic Fibrosis Exercise Working Group. (2015). Statement on exercise testing in cystic fibrosis. *Respiration*, *90*(4), 332–351. https://doi.org/10.1159/000439057

Jones, N. L., Makrides, L., Hitchcock, C., Chypchar, T., & McCartney, N. (1985). Normal standards for an incremental progressive cycle ergometer test. *American Review of Respiratory Disease*, *131*(5), 700–708. https://doi.org/10.1164/arrd.1985.131.5.700

Takken, T., Mylius, C. F., Paap, D., Broeders, W., Hulzebos, H. J., van Brussel, M., & Bongers, B. C. (2019). Reference values for cardiopulmonary exercise testing in healthy subjects—An updated systematic review. *Expert Review of Cardiovascular Therapy*. https://doi.org/10.1080/14779072.2019.1627874

Tomlinson, O. W., Barker, A. R., Trott, J., Oades, P. J., Withers, N. J., & Williams, C. A. (2020). P298 Annual changes in aerobic fitness are biased between normative equations in children and adolescents with cystic fibrosis. *Journal of Cystic Fibrosis*, *19*. https://doi.org/10.1016/s1569-1993(20)30627-5

Tomlinson, O. W., & Williams, C. A. (2022). WS14.05 Normal fitness data in cystic fibrosis—A scoping review. *Journal of Cystic Fibrosis*, *21*, S28. https://doi.org/10.1016/S1569-1993(22)00235-1

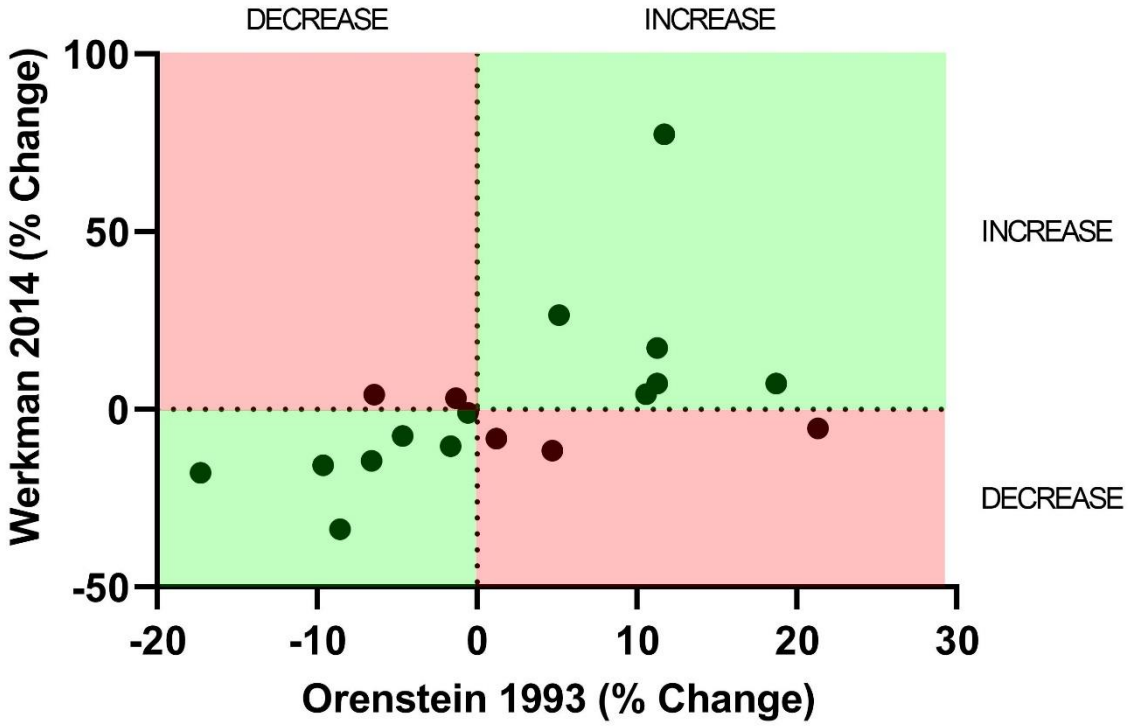
Waterfall, J. L., Burns, P., Shackell, D., Pepke-Zaba, J., Oates, K. E., & Sylvester, K. P. (2020). The risks of applying normative values in paediatric cardiopulmonary exercise testing: A case report. *ERJ Open Research*. https://doi.org/10.1183/23120541.00333-2020

Table 1. Normative reference values suggested for use by the European Cystic Fibrosis Society.

Guideline Document	**Suggested Normative Reference Value**	**Derived Population**
European Cystic Fibrosis Society (Hebestreit et al., 2015)	*Cycle Ergometry* 1. Jones *et al*, 1985; Am Rev Respir Dis; 131: 700–708. 2. Orenstein, 1993; In Rowland TW (ed), Pediatric Laboratory Exercise; pp. 141–163. 3. Werkman *et al*, 2014; Arch Dis Child; 99: 21–25. *Treadmill* 4. Pollock *et al*, 1982; Am Heart J; 103: 363–373. 5. Foster *et al*, 1984; Am Heart J; 107: 1229–1234. 6. Thompson *et al*, 2010; ACSM's Guidelines for Exercise Testing and Prescription.	1. n = 100, male & female, 15-71 years 2. n = not reported, sex unknown, age unknown 3. n = 363, male & female, 14.8 ± 1.7 years 4. n = 49, females, 27 ± 5 years 5. n = 200, males, 43 ± 16 years 6. n = not reported, sex unknown, age unknown

Please see individual guidelines for further explanation for choice of NRV.

Figure 1. Change in VO_{2peak} in 18 teenagers with cystic fibrosis, with data presented using two different sets of normative reference values.



Green boxes indicate that values agree in direction of change (i.e., both increase, or both decrease). Red boxes indicate no agreement in directions of change.