

## **SUPPLEMENTARY DATA**

### **The QUALIREHAB study group:**

Sophie Guillaumont (Montpellier University Hospital, France); Omar Oubari (Anamnese, France); Jerome Bourreau (Anamnese, France); Johanna Calderon (University of Montpellier, France); Stephane Moniotte (Brussels Cliniques universitaires Saint-Luc, Belgium); Jelena Hubrechts (Brussels Cliniques universitaires Saint-Luc, Belgium); Esteban Peiró-Molina (IIS La Fe Valencia, Spain); Erika Rezola (IIS Biogipuzkoa Donostia/San Sebastian, Spain); Begoña Manso (Hospital Universitario Virgen del Rocío Sevilla, Spain); Gonzalo Oñoro (Gerencia de Atencion Primaria Salud Madrid, Spain); Francisco-José Ferrer-Sargues (Universidad Cardenal Herrera CEU); Alice Bordessoule (Geneva University Hospital, Switzerland); Arthur Gavotto (Montpellier University Hospital, France); Stephan Matecki (Montpellier University Hospital, France); Oscar Werner (Nantes University Hospital); Marion Audie (Bordeaux University Hospital, France); Cecile Jore (Bordeaux University Hospital, France); Mathieu Andrianoely (University of Bordeaux, France); Caroline Neyraud (Bordeaux University Hospital, France); Anne-Cecile Huby (Bordeaux University Hospital, France); Marine Tortigue (Lyon University Hospital, France); Arielle Desir (Washington DC Children's National Hospital, USA); Jared Hershenson (Washington DC Children's National Hospital, USA); Justin Burton (Washington DC Children's National Hospital, USA); Pascal Amedro (Washington DC Children's National Hospital, USA).

### **TECHNICAL DETAILS**

All 3 participating Spanish CPET laboratories used the same technical equipment, including: *a*) pediatric face masks (Hans Rudolph, Shawnee, KS, USA); *b*) 12-lead EKG (CardioSoft, GE Healthcare, Little Chalfont, UK; CardioPart 12 Blue, Amedtec Medizintechnik, Aue, GE); *c*) breath-by-breath analysis software (Windows 10, Cortex Biophysik GmbH, Leipzig, Germany; or Windows 10, Bluecherry, Geratherm Respiratory GmbH, Bad Kissingen, Germany); *d*) Cortex Metalyzer 3B gas analyzer (Cortex Biophysik GmbH, Leipzig, Germany); *e*) pulse oximeters

(Masimo, CA, USA; or ear pulse oximeter, Geratherm, Germany); *f*) manual sphygmomanometers for blood pressure measurement.

The pediatric CPET equipment used in the reference development cohort was comparable, except for the calibrated gas analyzers, which included the Quark CPET (Cosmed Srl, Pavona di Albano, Italy) and the Oxycon Pro (Jaeger, Erich Jaeger GmbH, Hoechberg, Germany).

**Table S1.** Center-specific clinical characteristics and performance of VO<sub>2</sub>max reference models in the Spanish validation cohort

	Valencia (n = 87)	San Sebastián (n = 64)	Seville (n = 24)
<i>Clinical characteristics</i>			
Age, y	12.9 ± 1.8	11.9 ± 1.8	11.1 ± 2.4
Girls	26 (29.9)	28 (43.8)	18 (75.0)
BMI, kg/m <sup>2</sup>	19.9 ± 3.0	19.0 ± 2.5	18.5 ± 3.2
BMI ≥85th percentile	22 (25.3)	15 (23.4)	5 (20.8)
Measured VO <sub>2</sub> max, mL/min	1913 ± 492	1763 ± 446	1356 ± 231
<i>Z-score model</i>			
VO <sub>2</sub> max Z-score	-0.46 ± 1.04	-0.25 ± 0.84	-0.98 ± 1.04
Impaired VO <sub>2</sub> max <sup>a</sup>	16 (18.4)	3 (4.7)	5 (20.8)
VO <sub>2</sub> max difference <sup>b</sup> , mL/min	-101.9 ± 285.3	-39.0 ± 212.4	-224.9 ± 260.6
R <sup>2</sup> (observed vs predicted)	0.66	0.80	0.56
<i>Height-based linear equation</i>			
VO <sub>2</sub> max % predicted	88.6 ± 14.9	93.5 ± 13.1	87.3 ± 15.7
Impaired VO <sub>2</sub> max <sup>c</sup>	24 (27.6)	10 (15.6)	5 (20.8)
VO <sub>2</sub> max difference <sup>b</sup> , mL/min	-280.8 ± 358.0	-140.8 ± 248.0	-269.1 ± 378.1
R <sup>2</sup> (observed vs predicted)	0.61	0.74	0.49
<i>Weight-based linear equation</i>			
VO <sub>2</sub> max % predicted	89.7 ± 14.8	96.3 ± 12.1	89.6 ± 19.2
Impaired VO <sub>2</sub> max <sup>c</sup>	24 (27.6)	6 (9.4)	5 (20.8)
VO <sub>2</sub> max difference <sup>b</sup> , mL/min	-266.7 ± 384.7	-80.0 ± 225.9	-262.5 ± 485.5
R <sup>2</sup> (observed vs predicted)	0.62	0.77	0.34

The data are expressed as mean  $\pm$  standard deviation or No. (%). Values are reported descriptively to explore center-related variability in clinical characteristics and model performance.<sup>a</sup> Impaired VO<sub>2</sub>max defined as VO<sub>2</sub>max Z-score < -1.64 (5th percentile of the reference population).

<sup>b</sup> Difference calculated as observed minus predicted VO<sub>2</sub>max.

<sup>c</sup> Impaired VO<sub>2</sub>max defined as VO<sub>2</sub>max < 80% of predicted using linear reference equations.

**Table S2.** Agreement in classification of impaired aerobic fitness between VO<sub>2</sub>max reference models

Reference model comparison	Normal/normal (No.)	Normal/impaired (No.)	Impaired/normal (No.)	Impaired/impaired (No.)	Percent agreement (%)	Cohen's $\kappa$ (95% CI)	<i>P</i>
Z-score vs height-based linear equation	135	16	1	23	89.1	0.67 (0.51-0.79)	< .001
Z-score vs weight-based linear equation	132	19	8	16	84.6	0.45 (0.27-0.61)	< .001
Height-based vs weight-based linear equation	122	14	18	21	80.6	0.45 (0.29-0.60)	< .001

Agreement in classification of impaired aerobic fitness between vo<sub>2</sub>max reference models in the Spanish cohort (n = 175). Cells indicate the number of participants classified as having normal or impaired aerobic fitness according to each pair of reference models, with the first term referring to the first-listed model and the second term to the comparator model. Percent agreement and Cohen's kappa ( $\kappa$ ) with 95% confidence intervals (95%CI) are reported to quantify concordance beyond chance. *P* values correspond to tests of agreement significance.

**Figure S1. Bland-Altman analysis of VO<sub>2</sub>max prediction among reference models in the Spanish cohort**

Bland-Altman plots showing the agreement between observed and predicted VO<sub>2</sub>max values for (A) the Z-score model, (B) the height-based linear equation, and (C) the weight-based linear equation in the Spanish cohort. The x-axis represents the mean of observed and predicted VO<sub>2</sub>max (mL/min), and the y-axis represents the difference between observed and predicted values (observed – predicted). Solid horizontal lines indicate the mean bias, and dashed lines represent the limits of agreement ( $\pm 1.96$  standard deviation).

