# Validation of the MGC Diagnostics Meridian Metabolic Cart Against the Douglas Bag Method During Maximal-Intensity Exercise

Jack Mislinski and Andrew Subudhi

*Department of Human Physiology and Nutrition, University of Colorado Colorado Springs*

## INTRODUCTION

Maximal oxygen uptake ($VO_2$max) is a defining marker of cardiorespiratory fitness, a strong independent predictor of all-cause mortality,[1,11] and a routine input into clinical risk stratification, exercise prescription, and athletic preparation.[1] In nearly all settings where it is now measured, $VO_2$max is determined not by direct collection of expired gas but by automated breath-by-breath metabolic carts. The accuracy of these systems therefore directly determines the validity of the decisions made downstream. The question this study addresses is straightforward: how well does an automated breath-by-breath cart agree with the Douglas bag method when used to measure gas exchange at maximal exercise intensity?

The criterion method for measuring pulmonary gas exchange during exercise is the Douglas bag method, in which expired air is collected over a defined interval and subsequently analyzed for volume and gas fractions. The Douglas bag method has served as the reference standard for over a century,[12] but it is labor-intensive, time-consuming, requires specialized equipment, and is no longer common practice in most laboratories. The advent of microprocessor-based gas analysis enabled real-time processing of expired gas and ventilation, first through mixing-chamber systems and more recently through breath-by-breath systems that report $VO_2$, $VCO_2$, and minute ventilation (VE) on every breath.[13] Breath-by-breath measurement carries inherent challenges that scale with exercise intensity: at maximal effort, ventilation rises sharply, breathing frequency and depth change, end-tidal gas fractions deviate from steady-state assumptions, and the response time of gas analyzers and the temporal alignment of flow and concentration signals are pushed toward their physical limits. Whether modern automated systems can hold their accuracy under these conditions is not well established.

Validation of automated metabolic carts has been performed primarily at rest and during submaximal, steady-state exercise. Independent simulator-based comparisons have reported $VO_2$

errors ranging from approximately 1% to over 17% across commercially available breath-by-breath systems, with both systematic over- and underestimations and substantial between-system variability.[2] Direct comparisons against the Douglas bag method during human exercise are far less common; the existing literature in this category has been conducted predominantly during cycle ergometry. To our knowledge, no published study has evaluated agreement between an automated breath-by-breath cart and the Douglas bag method during maximal-intensity treadmill running. This gap is consequential because $VO_2$max, by definition a peak and non-steady-state value, is among the most common variables these systems are used to measure, and the conditions under which it is recorded — high ventilation, rapid changes in flow and gas concentration, and approach to physiological limit — are precisely the conditions least represented in the existing validation literature.

The MGC Diagnostics Meridian metabolic cart is a recently introduced breath-by-breath system marketed for cardiopulmonary exercise testing and $VO_2$max measurement, and to our knowledge no independent validation study of the Meridian has been published. The purpose of this study was to evaluate the level of agreement between the MGC Meridian and the Douglas bag method during maximal-intensity treadmill exercise. Healthy adults completed a graded exercise test to volitional exhaustion to establish peak running velocity, followed by a validation trial in which expired gases were collected simultaneously by the Meridian system and Douglas bags during the final stages of a near-maximal incremental protocol. Agreement between the two methods was characterized using Bland–Altman analysis for VE, $VO_2$, $VCO_2$, and respiratory exchange ratio (RER).[3]

## METHODS

### *Study Design*

To assess the validity of an automated metabolic measurement system against the Douglas bag method during maximal-intensity exercise, each participant completed two exercise trials in a single visit. Testing was conducted at the University of Colorado Colorado Springs (~1,840 m). The first trial was a graded exercise test (GXT) to volitional exhaustion, performed solely to establish peak running velocity. The second trial was a validation trial in which the treadmill protocol resumed near peak velocity and expired gases were collected simultaneously by the

automated metabolic cart and by Douglas bags. Peak running velocity from the GXT was used to determine the starting speed for the validation trial: the protocol began two stages below the last fully completed stage and incremented 0.4 mph each minute, allowing collection of Douglas bags at near-maximal and maximal intensities. At least one bag was collected per participant, with additional bags collected when the participant continued to higher stages.

### *Equipment*

#### Automated Metabolic System

Expired gases were analyzed breath-by-breath using an MGC Diagnostics Meridian metabolic cart (MGC Diagnostics, St. Paul, MN). The system employs a pitot tube flow sensor, a galvanic fuel cell $O_2$ analyzer, and an infrared $CO_2$ analyzer. Before each testing session, the system was calibrated according to the manufacturer's standard procedures. Gas calibration was performed using a two-point reference: room air (20.93% $O_2$, 0.04% $CO_2$) and a certified reference gas (12% $O_2$, 5% $CO_2$). Flow calibration was performed using a 3 L calibration syringe interfaced with the system's built-in calibration software. The system allows digital correction for dead space of the mask and any apparatus connected distal to the mouth, which was applied as appropriate during the validation trial.

#### Criterion Methods

Expired gas was collected in 200 L polyvinyl chloride Douglas bags (Harvard Apparatus). Following collection, bags were transported immediately to an adjacent laboratory and analyzed within approximately 20 minutes. Gas fractions of $O_2$ and $CO_2$ were determined using a mass spectrometer (MGA-1100; MA Tech Services, St. Louis, MO). For each bag, the mass spectrometer sampled the bag contents via an automated pump for 30 seconds (30 mL sample volume), and gas fractions were recorded at the 20-second mark once readings had stabilized. Bag volume was determined by evacuating the bag contents into a chain-compensated Tissot wet spirometer (Warren E. Collins, Inc., Braintree, MA) using a 7 L syringe; gas temperature inside the spirometer was recorded for volume correction.

#### Breathing Apparatus for Validation Trial

During the validation trial, participants breathed through a facemask connected to the MGC Diagnostics pitot tube used in the initial trial. The pitot tube was connected in series to a two-way non-rebreathing valve (model 2700; Hans Rudolph, Shawnee, KS) via a rubber coupler.

The Hans Rudolph valve directed expired air into the Douglas bag via 1 m of clean-bore tubing, and the valve assembly was secured to the participant using headgear. The additional dead space introduced by the valve *[EDITOR'S NOTE: confirm value displayed on the MGC cart graphic]* was entered into the MGC Diagnostics software, which applied an automatic correction to the breath-by-breath calculations.

### *Exercise Protocol*

#### Initial Graded Exercise Test

Upon arrival, participants' height and body mass were measured. Participants were then fitted with a facemask connected to the MGC Diagnostics pitot tube and outfitted with a chest-strap heart rate monitor. Warm-up treadmill speed was estimated for each participant by asking their most recent 5 km race time and subtracting approximately 6–8 stages from the corresponding race pace. *[EDITOR'S NOTE: data sheet does not include starting pace range, only validation pace range]* The protocol began with two minutes of standing rest to verify baseline gas exchange readings, followed by five minutes of warm-up. Treadmill speed was then increased by 0.4 mph at the five-minute mark and every minute thereafter until the participant reached volitional exhaustion. Treadmill grade was held constant at 0.5% throughout. Rating of perceived exertion (RPE; Borg 6–20 scale) was recorded each minute. The last fully completed one-minute stage was recorded as the peak stage and the corresponding speed as peak running velocity. If a participant reached volitional exhaustion partway through a stage (e.g., 15 seconds into a given stage), that incomplete stage was not counted, and the preceding fully completed stage was recorded as the peak stage.

#### Validation Trial

Following five minutes of seated rest, participants began the validation trial with two minutes of walking, followed by three minutes of standing while the additional breathing apparatus (Hans Rudolph 2700 valve, connecting tubing, and headgear) was attached. The treadmill protocol then resumed at the speed corresponding to two stages below the participant's peak stage from the initial GXT. Across the nine participants, starting speeds for the validation trial ranged from 7.4 to 12.6 mph. Speed was increased by 0.4 mph each minute, replicating the incremental protocol. Douglas bags were collected during the final 30 seconds of each minute-long stage of the validation trial, with a separate bag used for each collection period. Bag collection was initiated

and terminated manually by an operator who communicated verbally with the test administrator to synchronize bag open and close times with the elapsed time displayed on the MGC Diagnostics system.

### *Data Processing*

Raw, unaveraged breath-by-breath data were exported from the MGC Diagnostics system. For each Douglas bag collection period, the corresponding 30-second epoch of breath-by-breath data was identified by matching elapsed time stamps, and a simple arithmetic mean of all breaths falling within that window was calculated. This 30-second average from the MGC system was then compared to the values derived from the corresponding Douglas bag. For the Douglas bag criterion, VE, $VO_2$, $VCO_2$, and RER were calculated from measured expired gas fractions and ventilation volume using standard indirect calorimetry equations.[16] Volumes were corrected to STPD for $VO_2$ and $VCO_2$ and to BTPS for VE. The MGC Diagnostics software applied STPD and BTPS corrections automatically, while corrections were performed manually for the Douglas bag calculations.

### *Statistical Analysis*

Agreement between the MGC Diagnostics Meridian metabolic cart and the Douglas bag method was assessed using Bland–Altman analysis for four variables: VE (BTPS), $VO_2$ (STPD), $VCO_2$ (STPD), and RER. Mean bias (MGC minus Douglas bag) and 95% limits of agreement (mean bias ± 1.96 SD) were calculated for both absolute and percent differences. Paired-samples t-tests were used to test whether the mean bias differed significantly from zero. Data are reported as mean ± SD unless otherwise stated. Statistical significance was set at $P < 0.05$. All statistical analyses were performed in Microsoft Excel.

## RESULTS

Nine healthy, recreationally active to amateur competitive adult athletes recruited by word of mouth completed the validation protocol and contributed 17 paired bag–cart observations at near-maximal and maximal exercise intensities. Six participants completed two stages during the validation trial, one participant completed three stages, and two participants completed one stage. Demographic and fitness characteristics are summarized in Table 1. Participants were excluded if they did not meet the American College of Sports Medicine pre-participation health screening

guidelines or if they presented any contraindication to maximal exercise testing. All participants provided written informed consent, and the study was approved by the Institutional Review Board at the University of Colorado Colorado Springs.

**Table 1.** *Participant characteristics (n = 9).*

| Characteristic | Mean ± SD (range) |
|---|---|
| n (male / female) | 9 (7 / 2) |
| Age (yr) | 27.3 ± 3.9 (24–37) |
| Height (cm) | 174.3 ± 12.2 (148–188) |
| Body mass (kg) | 71.2 ± 14.3 (44.1–94.0) |
| Peak relative $VO_2$* ($mL \cdot kg^{-1} \cdot min^{-1}$) | 46.6 ± 7.3 (36.7–60.7) |

**Measured during the initial graded exercise test.*

Minute ventilation measured by the MGC system was lower than the Douglas bag criterion by a mean of 2.5 $L \cdot min^{-1}$ (1.9%; $p = 0.010$), a small but statistically significant underestimation. The bias was distributed approximately uniformly across the range of measured ventilation (Fig. 1A; Table 2).

Oxygen uptake was overestimated by the MGC system by a mean of 207 $mL \cdot min^{-1}$ (6.5%; $p < 0.001$), a statistically significant overestimation. The bias was consistent in direction across all 17 paired observations and did not vary systematically with the magnitude of $VO_2$ (Fig. 1B; Table 2).

Carbon dioxide production showed a mean bias of +35 $mL \cdot min^{-1}$ (1.6%) that did not differ statistically from zero ($p = 0.235$). The 95% limits of agreement were considerably wider relative to mean magnitude than those for VE, indicating that paired comparisons could differ in either direction by more than the group mean bias suggested (Fig. 1C; Table 2).

Respiratory exchange ratio was underestimated by the MGC system by a mean of 0.04 RER units (4.6%; $p < 0.001$), a statistically significant underestimation. All 17 paired differences fell below zero, and the 95% limits of agreement were narrow (Fig. 1D; Table 2). This pattern follows arithmetically from the upward bias in $VO_2$ combined with the absence of a significant bias in $VCO_2$: with an inflated denominator and a stable numerator, the ratio falls.

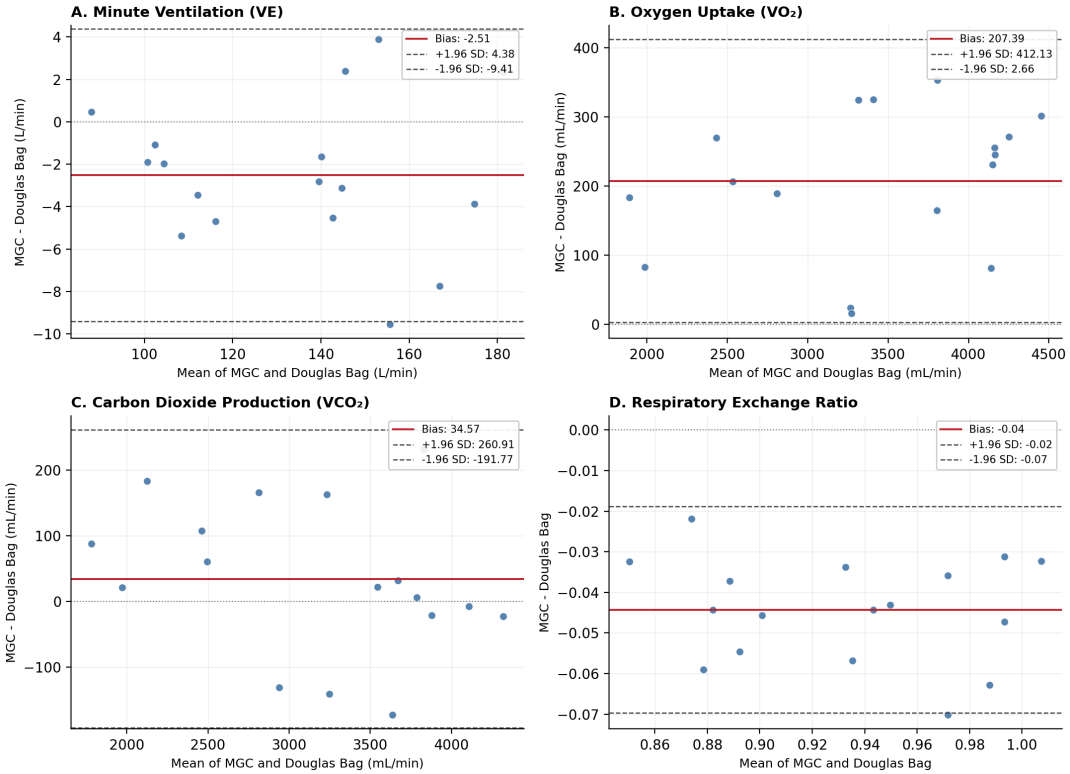


**Figure 1.** *Bland–Altman plots comparing the MGC Diagnostics Meridian metabolic cart and the Douglas bag method for (A) minute ventilation ($\dot{V}E$), (B) oxygen uptake ($\dot{V}O_2$), (C) carbon dioxide production ($\dot{V}CO_2$), and (D) respiratory exchange ratio (RER). Each point represents one paired observation (n = 17). The solid red line indicates the mean bias (MGC − Douglas bag); dashed lines indicate the 95% limits of agreement (mean bias ± 1.96 SD); the dotted reference line marks zero.*

**Table 2.** *Agreement between the MGC Diagnostics Meridian metabolic cart and the Douglas bag method at near-maximal and maximal exercise intensity (n = 17 paired observations from 9 participants). Bias values are reported as both absolute differences (MGC − Douglas bag) and percent differences (100 × [MGC − DB] / DB). The 95% limits of agreement are calculated as mean bias ± 1.96 SD. The p-value column reports the result of a paired t-test of the absolute mean bias against zero.*

| Variable | MGC (mean ± SD) | Douglas Bag (mean ± SD) | Bias, absolute (MGC − DB) | 95% LoA, absolute | Bias, % | 95% LoA, % | p (bias ≠ 0) |
|---|---|---|---|---|---|---|---|
| VE (L/min) | 132.7 ± 28.2 | 135.2 ± 28.4 | −2.5 (3.5) | [−9.4, 4.4] | −1.9% | [−6.5%, 2.8%] | 0.010 |
| $VO_2$ (mL/min) | 3507 ± 837 | 3299 ± 808 | +207 (104) | [3, 412] | +6.5% | [−0.1%, +13.2%] | <0.001 |
| $VCO_2$ (mL/min) | 3185 ± 765 | 3151 ± 799 | +35 (115) | [−192, 261] | +1.6% | [−6.2%, +9.4%] | 0.235 |
| RER | 0.91 ± 0.05 | 0.96 ± 0.05 | −0.04 (0.01) | [−0.07, −0.02] | −4.6% | [−7.2%, −2.1%] | <0.001 |

## DISCUSSION

This study shows that the MGC Diagnostics Meridian metabolic cart does not measure the four primary gas exchange variables with uniform agreement against the Douglas bag method during maximal-intensity treadmill exercise. Oxygen uptake was overestimated by 6.5%, respiratory exchange ratio was underestimated by 4.6%, ventilation was underestimated by 1.9%, and carbon dioxide production showed near-zero mean bias but wide individual-level variability. The two systematic biases — in $VO_2$ and RER — are large enough to alter the practical and clinical interpretation of cart-derived measurements, and they fall at or beyond the upper edge of what recent validation literature has treated as acceptable.

Recent independent validation work in which breath-by-breath CPET systems were compared against metabolic simulators provides a useful frame of reference. Across these studies, $VO_2$ errors have ranged from approximately 1% to over 17% across commercially available systems, with most well-performing systems falling within ±5% and a smaller subset showing errors of 10% or more.[2,4,5,6] The most explicit acceptability threshold in this literature is the 5% criterion described for clinical-cart validation against a simulator.[7] Direct human-exercise comparisons against the Douglas bag method during cycling have shown a similar split, with portable breath-by-breath systems reporting either close agreement or systematic $VO_2$ overestimations of 10–17%.[5,6] Our observed $VO_2$ bias of 6.5% sits in the middle of this distribution: it exceeds the 5% threshold cited in simulator literature and the better-performing portion of the field, but is well below the largest errors reported for portable systems against the Douglas bag during human exercise. By the most commonly cited criterion in the field, the Meridian's $VO_2$ performance at maximal intensity does not pass.

A 6.5% bias in $VO_2$ is consequential because $VO_2$max is recommended by the American Heart Association as a clinical vital sign[1] and is used to place individuals into cardiorespiratory fitness categories that carry independent prognostic information.[1,11] In the present sample, the mean Douglas bag $VO_2$ was 3,299 $mL \cdot min^{-1}$, or approximately 46.3 $mL \cdot kg^{-1} \cdot min^{-1}$, at a mean body mass of 71.2 kg. Applying the observed 6.5% upward bias, the same individual measured by the Meridian would yield a value of approximately 49.4 $mL \cdot kg^{-1} \cdot min^{-1}$ — a shift of about 3 $mL \cdot kg^{-1} \cdot min^{-1}$. In the ACSM percentile bands published in the *Guidelines for Exercise Testing and Prescription*,[8] this magnitude of shift is sufficient to move an average male in the 25–34 yr

range from one fitness category into the next. Because the observed bias varied across observations rather than affecting all participants by the same amount, the practical impact on individual classification is best expressed at the participant level: when peak $VO_2$ from the validation trial was used to assign each of the nine participants to an ACSM cardiorespiratory fitness category, five of nine participants were placed in a different category by the MGC system than by the Douglas bag method, and in every case the cart-derived value placed the participant one category higher.

These findings sit uneasily against the regulatory and tolerance framework under which CPET systems reach the clinic. MGC Diagnostics, like other CPET manufacturers, brings its devices and software to market through the FDA 510(k) substantial-equivalence pathway,[9] which requires demonstration that a new device is comparable to a legally marketed predicate but does not require independent performance validation against a criterion method. The manufacturer does provide a device-specific pre-shipment validation record against a metabolic simulator and will not ship a system that fails an internal 4% accuracy criterion; however, this validation is performed against a simulator and only up to approximately 3.5 $L \cdot min^{-1}$ $VO_2$, leaving the upper range of human $VO_2$max output and the conditions of human-exercise gas exchange uncharacterized at the manufacturer level. Manufacturer documentation for the Meridian Series describes the system as designed to meet or exceed ATS/ERS standards[14] but does not publish a quantitative accuracy specification for $VO_2$, $VCO_2$, or RER under exercise conditions. The honest summary is that the Meridian falls inside the manufacturer's internal simulator criterion at sub-maximal flows, fails the most commonly cited cross-system simulator threshold, and is unvalidated under maximal-intensity human exercise — the condition for which the device is widely used.

The RER finding has a specific consequence worth highlighting. Achievement of RER $\geq$ 1.10 is one of the most widely used secondary criteria for confirming attainment of a true $VO_2$max during graded exercise testing.[10] A systematic 4.6% underestimation means a cart-derived RER of 1.05 corresponds to a Douglas bag value of approximately 1.10. During the validation trial in this study, the maximum cart-derived RER reached 1.02, while the corresponding Douglas bag value was 1.07. Applied as a strict cutoff, the cart-derived RER would have flagged genuinely maximal tests as submaximal in some cases. This pattern argues against using cart-derived RER

as a strict standalone cutoff for confirming maximal effort under these testing conditions, and it supports the broader cautious view that no single secondary criterion should be relied on in isolation.[15]

Several limitations should be considered when interpreting these findings. First, all testing was conducted at moderate altitude (~1,840 m), and the design cannot separate the contribution of altitude from the contribution of maximal intensity to any observed disagreement. Two altitude-specific caveats apply to the RER results. The validation trial was conducted after a recovery period following the initial graded exercise test, during which participants would have hyperventilated and depleted $CO_2$ stores; the RER values observed during the validation trial are therefore likely lower in absolute terms than they would have been had the validation collection occurred during the initial test. Independently, residence and testing at moderate altitude tend to lower resting and exercise $PETCO_2$ relative to sea level, which may further constrain the upper range of achievable RER values. Both effects depress the absolute RER values reported here without biasing the simultaneous cart-versus-bag comparison itself, since both systems sample the same physiologically depressed RER. Second, the sample comprised 17 paired observations from 9 participants, with several participants contributing multiple observations; the paired t-test treats each observation as independent and does not account for within-subject correlation, and a mixed-effects analysis would provide a more rigorous treatment. Third, the participant sample was modest in size and demographically narrow. Fourth, the systems were operated in series rather than in parallel, with the Hans Rudolph valve and connecting tubing positioned distal to the MGC pitot tube; although the added dead space was measured and entered into the MGC software for compensation, it is not clear from manufacturer documentation whether the system corrects for trapped dead-space gas from the prior breath in the same way it accounts for static dead space. Any incomplete compensation of this type could contribute to the observed $VO_2$ bias. Fifth, only a single Meridian unit was tested; system-to-system variability across Meridian units cannot be characterized from these data and may differ from the unit-level agreement reported here.

In summary, this study shows that the MGC Diagnostics Meridian introduces clinically meaningful systematic error in $VO_2$ and RER measurement during maximal exercise, while measuring VE acceptably and producing variable but unbiased $VCO_2$ estimates. Cart-derived

$VO_2$max values obtained during maximal exercise should be interpreted with awareness of a likely upward bias of approximately 6%, the corresponding caveat applied to fitness category placement and exercise prescriptions, and RER should not be relied on as a strict standalone cutoff for confirming maximal effort under these conditions. For laboratories operating at altitude, periodic criterion-method validation should be considered as part of a quality-assurance program. For the regulatory and manufacturer framework that governs these devices, the present findings support the broader recommendation that 510(k) substantial-equivalence clearance be supplemented by independent performance validation under the conditions and intensities at which the devices are actually used in practice.

## REFERENCES


**1.** 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(24):e653–e699. https://doi.org/10.1161/CIR.0000000000000461

**2.** Van Hooren B, Souren T, Bongers BC. Accuracy of respiratory gas variables, substrate, and energy use from 15 CPET systems during simulated and human exercise. Scandinavian Journal of Medicine & Science in Sports. 2024;34(1):e14490. https://doi.org/10.1111/sms.14490

**3.** Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986;327(8476):307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

**4.** Van Hooren B, Souren T, Miqueu F, Bongers BC. Technological and biological reliability, and validity of five different CPET systems during simulated and human exercise. Scandinavian Journal of Medicine & Science in Sports. 2026;36(1):e70184. https://doi.org/10.1111/sms.70184

**5.** Crouter SE, LaMunion SR, Hibbing PR, Kaplan AS, Bassett DR. Accuracy of the COSMED K5 portable calorimeter. PLOS ONE. 2019;14(12):e0226290. https://doi.org/10.1371/journal.pone.0226290

**6.** Macfarlane DJ, Wong P. Validity, reliability and stability of the portable Cortex MetaMax 3B gas analysis system. European Journal of Applied Physiology. 2012;112(7):2539–2547. https://doi.org/10.1007/s00421-011-2230-7

**7.** Falcioni L, Guidetti L, Baldari C, et al. Accuracy and repeatability of the COSMED® Q-NRG max mobile metabolic system. PLOS ONE. 2025;20(3):e0319394. https://doi.org/10.1371/journal.pone.0319394

**8.** American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. 11th ed. Wolters Kluwer; 2021.

**9.** U.S. Food and Drug Administration. Premarket Notification 510(k). Center for Devices and Radiological Health; 2024.

https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/premarket-notification-510k

**10.** Howley ET, Bassett DR, Welch HG. Criteria for maximal oxygen uptake: Review and commentary. Medicine & Science in Sports & Exercise. 1995;27(9):1292–1301. https://doi.org/10.1249/00005768-199509000-00009

**11.** Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA. 2009;301(19):2024–2035. https://doi.org/10.1001/jama.2009.681

**12.** Douglas CG. A method for determining the total respiratory exchange in man. Journal of Physiology. 1911;42:17–18P.

**13.** Macfarlane DJ. Automated metabolic gas analysis systems: a review. Sports Medicine. 2001;31(12):841–861. https://doi.org/10.2165/00007256-200131120-00002

**14.** American Thoracic Society/American College of Chest Physicians. ATS/ACCP statement on cardiopulmonary exercise testing. American Journal of Respiratory and Critical Care Medicine. 2003;167(2):211–277. https://doi.org/10.1164/rccm.167.2.211

**15.** Poole DC, Wilkerson DP, Jones AM. Validity of criteria for establishing maximal $O_2$ uptake during ramp exercise tests. European Journal of Applied Physiology. 2008;102(4):403–410. https://doi.org/10.1007/s00421-007-0596-3

**16.** Weir JB de V. New methods for calculating metabolic rate with special reference to protein metabolism. Journal of Physiology. 1949;109(1–2):1–9. https://doi.org/10.1113/jphysiol.1949.sp004363