Data Processing Strategies to Determine Maximum Oxygen Uptake: A Systematic Scoping Review and Experimental Comparison with Guidelines for Reporting

# Abstract

**Background** Gas exchange data from maximum oxygen uptake (V̇O2max) testing needs post-processing. Different processing strategies may lead to varying V̇O2max values in individuals. The exact processing strategies used in the literature have not yet been systematically investigated. Previous research provided comparisons of methods only on a group level.

**Methods** Out of a random sample, we investigated 242 recently published articles that measured V̇O2max in an appropriate setting. We extracted the reported methods of data processing and their rationale. We compared the most common processing strategies on a data set of 72 standardized exercise tests in trained athletes.

**Results** Most researchers use binned time averages to determine V̇O2max, with a minority using moving time or moving breath averages. Half of the included studies did not report their data processing strategy and almost all articles failed to provide a rationale for the strategy used. The processing strategies found in the literature can lead to median differences in V̇O2max of more than 5% with considerable variation on an individual level.

**Conclusion** Researchers should report their data processing strategy used to determine V̇O2max. We provide a reporting checklist of seven items that can function as a template for reporting.

## Key Points

* Despite calls to use moving averages or digital filter, binned averages remain the most common data processing strategy to determine V̇O2max.
* Different processing strategies lead to varying V̇O2max values, which can hardly be predicted on an individual level.
* Researchers should at minimum report their processing strategy in detail, or preferably share their raw oxygen uptake data and analysis code.

# Introduction

The maximum oxygen uptake (V̇O2max) is one of the most commonly assessed physiological parameters in sports and exercise science [1]. The V̇O2max highly corresponds with endurance performance in heterogeneous groups [2–4] and can be regarded the most relevant physiological predictor of endurance performance [1]. Many training programs aim to target the V̇O2max, and studies evaluate the quality of interventions by measuring changes in V̇O2max.

Researchers predominantly measure V̇O2max during exercise tests to exhaustion. Among the factors that influence the V̇O2max determination are the exercise test protocol [5,6,7], exercise modality [8], motivation [9], and biological variability [10]. The measuring method (e.g. mixing chamber, breath-by-breath) and the metabolic cart model used can influence the measured oxygen uptake — and as such the V̇O2max [11,12]. But even the same oxygen uptake data generated during an exercise test may result in varying outcomes when processed differently [13].

Whether an observed peak in oxygen uptake corresponds to the true maximum has been extensively discussed in the past decades [14–18]. To identify a true maximum, researchers commonly evaluate a set of parameters measured during ramp tests — the so called ‘V̇O2max criteria’ [15]. We will not distinguish between peak and maximum oxygen uptake in this article, as the criteria for V̇O2max (e.g. the primary criterion of a plateau in oxygen uptake or the secondary criterion of the maximum respiratory quotient) do heavily rely on the data processing strategy used [13]. Thus, we speak of V̇O2max as the maximum oxygen uptake measured during an appropriate exercise test regardless of any V̇O2max criteria.

Measured gas exchange data is noisy. Both the biological variability of breathing patterns and the measurement error (as well as irregular breaths such as coughs and swallowing) lead to a highly fluctuating raw oxygen uptake on a breath-by-breath basis. For interpretation the raw data requires some form of processing. Different data processing strategies influence measured parameters of gas exchange [13,19–21]. Midgley et al. [22] found differences between data processing strategies for V̇O2max, but no difference in the reliability of these, inferring that no optimal strategy exists, but that consistency of strategies is key when comparing outcomes. Based on own data in sedentary and moderately trained individuals, Martin-Rincon et al. [23] provided linear-log equations to compare mean V̇O2max values derived from processing strategies with differing averaging interval lengths. There is inconclusive evidence on whether the influence of different processing strategies interacts with different exercise protocols [24,25].

Researchers have proposed a variety of calculation intervals and computational methods to process oxygen uptake data [15,26–29] and calls to standardize processing strategies are frequent [20,24,25,30]. In light of the influence on outcome variables, many articles highlighted the need to report processing strategies in research [13,20,25,30]. Midgley et al. [22] were the first to evaluate reported data processing strategies for breath-by-breath analyses in selected journals. They found that almost all studies reporting the use of binned time averaging, with only 1 in 117 using a moving time and a moving breath average, respectively. One third of the studies did not describe their processing method at all. While providing interesting first insights into reporting and processing practices, the search by Midgley et al. [22] was not systematic and its methods were not described in detail. Robergs et al. [29] wrote that to investigate the current state of data processing strategies, two possible approaches are *“(i) a summary of published research, and (ii) a survey circulated via the Internet to as many exercise physiologists as possible”*. They provided the latter one with a total of 75 respondents, who reported a great variety in data processing strategies. Most researchers reported the use of binned time averages over 30 or 60 seconds. Astonishingly, about half of the respondents admitted that their data processing strategy was rather chosen due to subjective influences as opposed to objective reasoning [29]. Practices in reporting and processing may have changed more than a decade after the publication of these numbers.

Selected data processing strategies have been extensively compared in the literature. Due to the absence of a systematic mapping of current practices, these works lacked the reasoning of which strategies to compare. Many studies compared different averaging intervals, but not averaging types (e.g. moving breath vs. binned time) [13,22]. Martin-Rincon et al. [23] provided formulas for comparing data processing strategies by investigating a data set of sedentary individuals and recreational athletes, using two different metabolic carts. Therefore in this work motivation and measurement devices may have interacted with the influence of processing strategies. Most comparisons only report mean differences between strategies [13,22]. No research has yet compared a variety of systematically derived strategies among a group of trained individuals using a standardized measurement set-up.

Differences in V̇O2max due to data processing can have serious implications in practice. They hinder the assessment of longitudinal data from athletes that participated in diagnostics with differing data processing or the comparability of data from studies within meta-analyses [23]. Data processing strategies directly affect the occurrence of a plateau in oxygen uptake, the primary criterion for V̇O2max [13]. Crucially, in situations where individuals are classified by their V̇O2max — for example when describing the training status of a study population [31,32] or evaluating patients for a heart transplantation [33] — differing processing strategies can lead to misclassifications [20].

This paper aims to review the usage and reporting of different data processing strategies in the scientific literature and investigate their influence on the V̇O2max. The results will help to compare V̇O2max data derived from different processing methods among studies and in individuals. The review allows to assess the implementation of data processing routines and to find malpractices in reporting. The results build a basis for providing recommendations for the reporting of data processing strategies to determine the V̇O2max.

# Methods

This work was preregistered before the start of the project on the [Open Science Framework](https://osf.io/3am4s) [34], following the ‘Inclusive Systematic Review Registration Form’ [35]. Any deviations from the preregistration are indicated in a ‘Transparent Changes’ document [LINK]. Major deviations will also be explicitly stated within the methods section. All data and code of this research project can be found on [GitHub](https://www.github.com/smnnlt/vo2max-processing). We conducted all analyses using R Version 4.2.0 [36] in the R Studio IDE Version 2022.2.2.485 [37].

## Systematic Scoping Review

The aim of the scoping review was to systematically map current practices of data processing for V̇O2max determination in the scientific literature. Since determining V̇O2max is a far too common procedure to perform an exhaustive search, we randomly sampled 500 articles that referred to V̇O2max or similar keywords. Data on processing strategies were extracted from all sampled articles that directly measured V̇O2max using an appropriate testing procedure in humans. The review was performed in accordance to the PRISMA extension for Scoping reviews [38], see LINK for the CHECKLIST.

### Search & Screening

The article search was conducted on 16th March 2022 using [PubMed](https://pubmed.ncbi.nlm.nih.gov/) and [Web of Science](https://www.webofscience.com/wos/woscc/basic-search). The search included articles published from 2017 to the date of the search referring to ‘maximum oxygen uptake’ or equivalent terms in title, abstract or keywords. APPENDIX [LINK] shows the exact search terms used.

The search results from both data bases were merged and checked for the presence of a Digital Object Identifier (DOI). Entries without DOI were excluded to allow for automated removal of duplicates by DOI matching in the next step. This was followed by an automated title scanning to exclude results that were likely no original research articles. All titles that contained one of the following words were excluded: *‘review’, ‘correction’, ‘meta-analysis’, ‘comment’, ‘retraction’, ‘editorial’, ‘erratum’, ‘reply’.*

In accordance with the preregistration we drew a random sample from the search results. The goal of this process was to give an unbiased estimate of the current state of scientific V̇O2max testing. Based on the procedure described in the preregistration, the sample included a total of 500 articles.

The abstracts from the articles included in the random sample were blinded for scanning, by removing any authors identities and journal information. Two of the authors (SN and OJQ) independently scanned the blinded abstracts to filter those that matched one of the exclusion criteria shown in **?@tbl-exclusion**. When the screeners disagreed in their assessment, they resolved the conflict by discussion.

After the abstract screening we retrieved the full-texts for the remaining articles. The full-texts were again independently scanned by two authors (SN and OJQ) to include only those articles that measured V̇O2max using an appropriate testing procedure in humans (see **?@tbl-exclusion** for the detailed full-text exclusion criteria[LINK]). Conflicts were resolved by discussion between the two examiners.

### Data Extraction

We retrieved data from all articles remaining after the abstract and full-text screening. Extraction included the following data:

* metabolic cart used
* measurement type (breath-by-breath, mixing chamber)
* type of outcome for V̇O2max (primary, secondary, other)
* data preprocessing (e.g., filtering)
* data processing software
* interpolation procedure
* data processing/determination of V̇O2max:
  + type (time average, breath average, digital filtering, …)
  + alignment (moving, binned, …)
  + interval (in seconds or breaths, parameters for filtering)
* rationale for the used data processing strategy (e.g. a reference)

The criteria ‘type of outcome’ and ‘rationale’ were added to the extraction list after the abstracts had been scanned, thus they were not stated in the preregistration.

### Data Synthesis

The extracted data is presented in a purely descriptive way. We calculated the relative and absolute frequency for the reporting of the extracted items. Similarly, we counted the use of different strategies for processing data in all articles that reportedly measured breath-by-breath. Total interval duration of averaging procedures were derived from the reported parameters.

## Experimental Comparison

To determine the influence of the most common data processing strategies on the determination of V̇O2max, we compared them on a set of already collected gas exchange data from ramp tests in running.

### Data Source

A total of N = 72 exercise tests were analyzed for this study. Due to a miscalculation, the preregistration had incorrectly stated a number of 76 tests. The data were from previous research on the metabolic profile of endurance runners [39,40]. The tested individuals were experienced distance runners (15 female, 54 male; three of the males participated in both studies). The V̇O2max tests were conducted in March to September 2019 [39] and March to October 2021 [40], respectively, while using identical exercise protocols and test equipment. Participants run on a treadmill (saturn 300/100, h/p/cosmos sports & medical 127 GmbH, Nussdorf-Traunstein, Germany) with 1% inclination for 8 minutes at a velocity of 2.8 m·s-1 as a warm-up. After preparing the gas exchange measures, they started a ramp protocol with an initial speed of 2.8 m·s-1 for two minutes and subsequently increased velocity by 0.15 m·s-1 every 30 seconds. The researchers provided verbal encouragement and terminated the exercise when the participants reached subjective exhaustion.

Gas exchange data were recorded using a ZAN 600 USB device (nSpire Health, Inc., Longmont, CO, United States of America). The device was calibrated with a 3l-syringe pump (nSpire Health, Inc., Longmont, CO, 143 United States of America) and a reference gas (15% O2, 6% CO2) before each measurement. The measured breath-by-breath data is available on [GitHub](https://github.com/smnnlt/vo2max-processing/tree/master/data/ramptests).

### Data Processing

The spiro package version 0.0.4 for R [41] processed the raw gas exchange data. The software includes various algorithms to calculate V̇O2max with user-defined parameters. Moving time-based averages were calculated by first linearly interpolating the breath-by-breath data to full seconds. Subsequently a (center aligned) moving average was calculated over the defined timespan.

For binned time-averages, the breath-by-breath data were initially interpolated to full seconds and then binned into consecutive intervals of constant lengths. The average of each interval was aligned to its center. Incomplete intervals (i.e. the last seconds of measurement) were not considered in the analysis. Note that some authors use a different procedure for determining their bins, starting by the end point of the measurement. However, defining bins beginning by the start of the measurement is a common output option for many gas exchange data analysis software (e.g. Cosmed Omnia). Breath based moving averages were calculated on the raw data.

### Comparison of Methods

To compare different processing methods within and between individuals, we chose to express the V̇O2max normalized to a reference procedure. The reference procedure was the most commonly applied one in current literature as determined by our scoping review. Individual V̇O2max values were expressed in reference to this procedure, where a value of 1 means that the processing method yields exactly the same V̇O2max value as the reference method. We calculated the data for all integer parameter values within the range of the calculation intervals found in the literature during the review. On a group level, we calculated the median and 10%- and 90%-quantiles of each processing strategy.

# Results

## Systematic Scoping Review

Initial search yielded 7529 results of which 4364 remained after automated filtering and removal of duplicates (see flow diagram in [Figure 1](#fig-flow)). Out of the random sample (n = 500), 242 articles were included in the final analysis.

|  |
| --- |
| Figure 1: Flow diagram for the systematic scoping review in accordance with the PRISMA 2020 Statement [42] |

Reporting practices of the methodology of gas exchange measures differed widely across the literature (see [Table 1](#tbl-reporting)). Almost half (44.2%) of the articles did not report any information regarding their data processing strategy. About one in twenty articles (5.8%) provided a rationale for their used strategy. Only a single article [43] reported information regarding all the investigated criteria.

Table 1: Percentage of studies that provided details on the different characteristics of oxygen uptake data processing.

| Metabolic cart | Preprocessing | Software | Processing Strategy | Rationale |
| --- | --- | --- | --- | --- |
| 88.0% | 5.6%\* | 15.0%\* | 55.8% | 5.8% |
| \*only examined within the subgroup of studies using breath-by-breath measurements |  |  |  |  |

Out of the authors that provided information and collected breath-by-breath measurements, most (79.5%) utilized binned averages to determine V̇O2max. Moving time averages, or breath-based averages were uncommon (see [Figure 2](#fig-strategies)). No study used methods of digital filtering to determine V̇O2max.

|  |
| --- |
| Figure 2: Strategies for processing breath-by-breath data in the reviewed literature (n = 88). |

For preprocessing, some authors reported the use of a (linear) interpolation for the breath-by-breath data to seconds (n = 7; 4.3%). A minority of researchers reported the use of data filtering strategies to remove outliers. This included the use of initial data smoothing by a short moving average (3 seconds, n = 1; 5 breaths, n = 3), the manual detection and removal of outliers (n = 2) or an automated removal of outliers (n = 5). For the automated outlier detection authors removed single data points differing from a not further defined local mean by a varying number of standard deviations (2, 3 or 4) or being outside of a 95% confidence interval. When reported, the software used for data processing varied among studies showing a total of more than 15 reported programs (for 30 studies that reported this parameter).

The calculation intervals for time-based averages of mixing chamber and breath-by-breath devices ranged from 5 to 60 seconds (see [Figure 3](#fig-duration)). 30-second intervals were most common to define V̇O2max, while authors also often employed shorter (10-20 s) and longer (60 s) periods.

|  |
| --- |
| Figure 3: Total durations of the calculation interval of V̇O2max in the reviewed studies. |

## Experimental Comparison

The average V̇O2max as determined by a binned 30-second average was 62.2 ± 6.3 ml·min-1·kg-1 (mean ± standard deviation). Applying different data processing strategies for V̇O2max determination leads to different outcome values (see [Figure 4](#fig-comparison)).

|  |
| --- |
| Figure 4: V̇O2max varies by data processing strategy. Values are expressed relative to the V̇O2max from a 30-second binned average — the most common strategy as determined by the review. Solid lines display the median, the shaded areas mark the interval between 10th and 90th percentile. Using moving average leads to systematically higher V̇O2max values compared to binned time averages. Changing the averaging interval (in seconds or breaths) can lead to median changes in V̇O2max as large as 5%. |

Binned time averages systematically generate lower V̇O2max values compared to moving averages. Decreasing the calculation interval to 5 seconds leads to a 3-4% median increase of V̇O2max values. Notably on an individual level these increases may be lower (~2%) or much higher (>10%) than the median value.

Moving time and moving breath averages yield nearly identical values for V̇O2max over all calculation intervals, as most of the trained athletes reached respiratory rates around 60 min-1 in the final minutes of the exercise test (see SUPPL FIG LINK).

# Discussion

We aimed to review current practices of data processing strategies to determine V̇O2max and to compare them on experimental data. Our results show, that current research uses a variety of processing strategies to determine V̇O2max, which directly influences the values obtained. Identical raw breath-by-breath data can result in different V̇O2max values when processed differently. Many articles only incompletely report on their methods, which hinders the reproducibility of V̇O2max measurement.

## Current State of Data Processing

Despite calls to use moving averages or digital filters [26,29], binned time averages remain the most common data processing strategy to determine V̇O2max in the reviewed literature (see [Figure 2](#fig-strategies)). Our numbers are generally in line with the findings of the non-systematic search by Midgley et al. [22] and the survey by Robergs et al. [29]. It is somewhat surprising, that practices have not changed in recent years despite the publication of recommendations, that discourage researchers from using binned averages [29]. Using binned time averages leads to systematically lower V̇O2max values as compared to moving averages (see [Figure 4](#fig-comparison)). The peak in oxygen uptake may be attained between two averaging intervals, resulting in an underestimation of V̇O2max. Binned time averages revoke the most important argument for measuring breath-by-breath: the high temporal resolution of data. Despite these arguments speaking against the use of binned time averages, our review demonstrates that they remain extremely common in the scientific literature.

Breath-based averages seem to be more common (~8%) than reported previously (<1%) [22], but less common than assessed in self-reporting (~17%) [29]. The increasing proportion of breath-based averages may be explained by publications in the recent years advocating for their use [26,29]. The length of the calculation interval for averaging is highly diverse within the literature (see [Figure 3](#fig-duration)). This may reflect contradictory recommendations [15,26]. There is no consensus on using shorter or longer calculation intervals, and an is unclear whether an optimal interval duration even exists. As different interval durations influence V̇O2max by more than 5% on a median level (see [Figure 4](#fig-comparison)), the exact reporting of the data processing strategy remains essential for interpretation.

A major source for choosing binned averages for data processing may be limitations by analysis software. Our review shows that most researchers use the software of the metabolic cart’s manufacturer to analyze the gas exchange data. These software may by default output binned time averages instead of raw breath-by-breath data. Moreover, further processing (e.g., interpolation, moving averages) may require the use of additional software. This may also explain why digital filtering has — despite recommended by Robergs et al. [29] — not been used in a single study reviewed here: standard distributions of common data analysis software (e.g., Microsoft Excel) lack the capability to perform such operations. The common practices of data processing are likely to be both attributed to a lack of awareness and a lack of adequate software solutions.

## Impact of Different Data Processing Strategies

The different data processing strategies found in the literature directly bias the V̇O2max determination (see [Figure 4](#fig-comparison)), and as such influence the classification of individuals, the evaluation of training success, and the assessment of V̇O2max attainment. In accordance with previous findings [13,20,21] longer calculation intervals lead to lower V̇O2max values (see [Figure 4](#fig-comparison)). The analyzed data shows median differences up to 5-7% between processing strategies, which is in accordance with previous research [23]. Some studies reported even higher mean differences of up to 20% [30], but only when including raw breath-by-breath data in the comparison. The evaluation of unprocessed raw data for its maximum is highly erroneous and as such is not performed in research (see [Figure 3](#fig-duration)); so there is no reason to compare it to other strategies. While previous research was often conducted in sedentary or recreationally trained individuals, we now present evidence, that a similar effect of data processing strategies on V̇O2max exists in trained athletes.

Binned time averages lead to systematically lower V̇O2max values compared with moving averages, for the reasons explained above. While this general trend has been acknowledged previously [23], it has not been quantified. Our data suggest a ~1% lower median V̇O2max when using binned averages compared with moving averages of the same calculation interval length. This difference is well within the measurement error of most if not all metabolic carts, but it is systematic and as such may bias the evaluation in scenarios where small changes in V̇O2max are important (e.g. in elite sports).

Moving time and moving breath averages lead to almost identical V̇O2max values on a median level ([Figure 4](#fig-comparison)). This seems natural in that the athletes in this study reached respiratory rates around 60 min-1 (see SUPPL LINK), resulting in equivalent time- and breath-based interval lengths. For an athletic population, V̇O2max values obtained by moving time and moving breath averages can approximately be used interchangeably. Given that less trained individuals display lower respiratory rates during exercise tests to exhaustion [44], this finding will likely not translate to a sedentary population.

The exact impact of data processing strategies on the V̇O2max is highly individual. Most research did present only comparisons of mean values, with results in accordance with those found here [23]. On an individual level, data processing strategies may impact the V̇O2max in different severity. For example, for 10% of the investigated athletes a binned time average of 5 seconds leads to a V̇O2max <3% higher than by a 30-second average, while for another 10% the V̇O2max was >6% higher (see [Figure 4](#fig-comparison)). Current values reported and equation derived compare strategies on a group level [23], which improves comparability of group results for meta-analyses or group classifications. However, on an individual level these equations can only be applied with a high margin of error. Changes in the impact of different data processing strategies on V̇O2max range from 1-2% in some individuals to more than 10% in others. Hence when evaluating V̇O2max data from different tests in a single individual obtained by using different processing methods, there is no way to accurately compare these values even when the processing strategies are reported. While the comparison of V̇O2max from different processing strategies require their reporting for an sufficient analysis on a group level, on the individual level the raw data from each test is required.

## Guidelines for Reporting

To compare and evaluate V̇O2max values from different studies, knowledge of the underlying data processing strategies is crucial. Our review demonstrates that almost half of the studies measuring V̇O2max did not describe their processing strategy. Other aspects of the data processing, such as outlier filtering or the rationale for the chosen procedure were only in rare instances reported (see [Table 1](#tbl-reporting)). [Table 2](#tbl-recommendation) lists seven items that should be reported to provide sufficient information on the data processing strategy used to determine V̇O2max. These items may be reported in form of a checklist, as an in-text enumeration or in sentence form. An example paragraph containing all the relevant information for the original data presented in this paper [39,40] would be:

*“We measured breath-by-breath data during the ramp tests with a ZAN 600 USB device (nSpire Health, Inc., Longmont, CO, United States of America). The raw data was analyzed without any previous filtering by using a low-pass forward-backward Butterworth filter (each filter: 3rd order, 0.04 Hz cut-off) implemented in the spiro package for R version 0.1.0 [41]. This strategy produces similar results as that recommended by Robergs et al. [29], but does not include a time lag.”*

Note that the correct reporting of an exercise test to determine V̇O2max requires more information than those on data processing. Further aspects to be reported include, but are not limited to, the study population, the exercise protocol, the device calibration, and criteria to stop the test. In cases where journals endorse word limits on articles, this reporting — including the reporting on data processing strategies — may be included in supplementary files. The correct and detailed reporting of data processing strategies, as well as other test characteristics, is crucial for interpreting presented V̇O2max values.

The results of this thesis suggest that comprehensive reporting facilitates approximate comparisons of V̇O2max data on a group level derived using different data processing strategies. But on an individual level and for a precise comparison, reporting is not enough, as differences between data processing strategies vary between individuals and are potentially influenced by training status. Sharing of the raw gas exchange data can solve this challenge, as it allows any researcher to recalculate the V̇O2max using their preferred data processing strategy. Most raw gas exchange data files are structured in a simple way, which allows to easily remove any personal information (if this had not been done in the metabolic cart’s system before). In terms of reproducibility of V̇O2max determination, a way even superior to correct reporting and raw data sharing is to additionally share data analysis code. This allows any researcher to independently reproduce the V̇O2max processing conducted within a study, but requires the data analysis to take place in a programming (or at least code generating) environment. Such programs for the purpose of analyzing gas exchange data exist as free open-source software [41,45].

Table 2: Recommendations for reporting data processing strategies to determine the maximum oxygen uptake.

| Reporting Item | Description |
| --- | --- |
| Metabolic Cart | State the exact device model and manufacturer. |
| Measurement Mode | State the measurement mode (e.g., mixing chamber, breath-by-breath,…). |
| Software | State the name and version of the software used for data analysis. |
| Preprocessing | State if and how data underwent any initial modification (e.g., filtering of outliers, interpolation) before analysis. |
| Processing Strategy | State the exact data processing strategy used to determine the VO2max (e.g., binned time averages). |
| Processing Parameters | State the parameters used for the processing strategy (e.g., length of averaging interval). |
| Rationale | State the rationale for using the processing strategy (e.g., reference to recommendations). |

## Limitations

Due to the sheer quantity of the publications investigating V̇O2max, it was not possible to perform an exhaustive review of the literature. The scoping review therefore relies on a random sample which not necessarily captures exact trends of the literature. However, efforts were made — such as random sampling and systematic article exclusions — to ensure the sample to be representative. Notable, almost half of the studies did not report their data processing strategy at all. The data processing strategies used in the literature could only be investigated when studies reported them.

Ambiguities in the reporting of the investigated studies may impact the analysis results. For example some studies using long binned averages (e.g., 60 seconds) may have in fact been using multiple binned averages of shorter duration (e.g., 4x15 second), without describing this correctly. Moreover the exact definitions for building binned averages vary within the literature: While most studies define the binning periods from the beginning of the exercise, some may define them from the endpoint. Additionally some studies reviewed did not define the maximum bin, but a pre-set binned average period as their V̇O2max (for example the last bin, regardless of its value). In situations where the maximum in oxygen uptake is reached considerably before exhaustion (i.e., a long plateau in oxygen uptake exists) this may lead to different results than a traditional binned average processing. We did not separately consider such sub-categories of data processing strategies, as they may not be very common and are often hard to precisely investigate due to ambiguities in their reporting.

This work treated each breath as the single data processing unit of cardiopulmonary exercise testing. However, metabolic carts sample gas fraction and gas flow data at a much higher frequency (e.g., 50 Hz). The data for each breath is subsequently calculated from these raw signals by the means of an algorithm. Different algorithms to generate the breath-by-breath data can lead to different outcomes [46], and as such may also influence the V̇O2max. Hence documenting and reporting of the breath-by-breath algorithm is advisable. Yet many metabolic carts do not describe their default algorithm and limit access to the raw data signal.

The experimental comparison of different data processing strategies was conducted on a standardized data set of exercise tests. This standardization in terms of training status, exercise protocol, and measurement device helps to highlight the impact of different data processing even in a relatively homogeneous data set. However the results may only partly transfer to different settings, such as individuals with less training background.

# References

1. Bassett DRJETH. Limiting factors for maximum oxygen uptake and determinants of endurance performance. Medicine & Science in Sports & Exercise. 2000;32:70. <https://doi.org/10.1097/00005768-200001000-00012>

2. Reaburn P, Dascombe B. Endurance performance in masters athletes. European Review of Aging and Physical Activity. 2008;5:31–42. <https://doi.org/10.1007/s11556-008-0029-2>

3. Costill DL, Thomason H, Roberts E. Fractional utilization of the aerobic capacity during distance running. Medicine & Science in Sports & Exercise. 1973;5:248252. <https://doi.org/10.1249/00005768-197300540-00007>

4. Tanaka K, Takeshima N, Kato T, Niihata S, Ueda K. Critical determinants of endurance performance in middle-aged and elderly endurance runners with heterogeneous training habits. European Journal of Applied Physiology and Occupational Physiology. 1990;59:443–9. <https://doi.org/10.1007/BF02388626>

5. Buchfuhrer MJ, Hansen JE, Robinson TE, Sue DY, Wasserman K, Whipp BJ. Optimizing the exercise protocol for cardiopulmonary assessment. Journal of Applied Physiology. 1983;55:1558–64. <https://doi.org/10.1152/jappl.1983.55.5.1558>

6. Yoon B-K, Kravitz L, Robergs R. V̇O2max, protocol duration, and the v̇O2 plateau. Medicine & Science in Sports & Exercise. 2007;39:11861192. <https://doi.org/10.1249/mss.0b13e318054e304>

7. Midgley AW, Bentley DJ, Luttikholt H, McNaughton LR, Millet GP. Challenging a Dogma of Exercise Physiology. Sports Medicine. 2008;38:441–7. <https://doi.org/10.2165/00007256-200838060-00001>

8. Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, Hamilton-Wessler M, et al. Comparison of the ramp versus standard exercise protocols. Journal of the American College of Cardiology. 1991;17:1334–42. <https://doi.org/10.1016/S0735-1097(10)80144-5>

9. Midgley AW, Marchant DC, Levy AR. A call to action towards an evidence-based approach to using verbal encouragement during maximal exercise testing. Clinical Physiology and Functional Imaging. 2018;38:547–53. <https://doi.org/10.1111/cpf.12454>

10. Katch VL, Sady SS, Freedson P. Biological variability in maximum aerobic power. Medicine & Science in Sports & Exercise. 1982;14:2125. <https://doi.org/10.1249/00005768-198201000-00004>

11. Winkert K, Kirsten J, Kamnig R, Steinacker JM, Treff G. Differences in V̇O2max Measurements Between Breath-by-Breath and Mixing-Chamber Mode in the COSMED K5. International Journal of Sports Physiology and Performance. 2021;16:1335–40. <https://doi.org/10.1123/ijspp.2020-0634>

12. Miles DS, Cox MH, Verde TJ. Four commonly utilized metabolic systems fail to produce similar results during submaximal and maximal exercise. Sports Medicine, Training and Rehabilitation. 1994;5:189–98. <https://doi.org/10.1080/15438629409512016>

13. Astorino TA. Alterations in VO2max and the VO2 plateau with manipulation of sampling interval. Clinical Physiology and Functional Imaging. 2009;29:60–7. <https://doi.org/10.1111/j.1475-097X.2008.00835.x>

14. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. Journal of Applied Physiology. 1955;8:73–80. <https://doi.org/10.1152/jappl.1955.8.1.73>

15. Howley ET, Bassett DR, Welch HG. Criteria for maximal oxygen uptake: Review and commentary. Medicine & Science in Sports & Exercise. 1995;27:12921301. <https://doi.org/10.1249/00005768-199509000-00009>

16. Poole DC, Wilkerson DP, Jones AM. Validity of criteria for establishing maximal O2 uptake during ramp exercise tests. European Journal of Applied Physiology. 2008;102:403–10. <https://doi.org/10.1007/s00421-007-0596-3>

17. Poole DC, Jones AM. Measurement of the maximum oxygen uptake v̇o2max: V̇o2peak is no longer acceptable. Journal of Applied Physiology. 2017;122:997–1002. <https://doi.org/10.1152/japplphysiol.01063.2016>

18. Green S, Askew C. V̇o2peak is an acceptable estimate of cardiorespiratory fitness but not v̇o2max. Journal of Applied Physiology. 2018;125:229–32. <https://doi.org/10.1152/japplphysiol.00850.2017>

19. Matthews JI, Bush BA, Morales FM. Microprocessor Exercise Physiology Systems vs a Nonautomated System: A Comparison of Data Output. CHEST. 1987;92:696–703. <https://doi.org/10.1378/chest.92.4.696>

20. Johnson JS, Carlson JJ, VanderLaan RL, Langholz DE. Effects of sampling interval on peak oxygen consumption in patients evaluated for heart transplantation. CHEST. 1998;113:816–9. <https://doi.org/10.1378/chest.113.3.816>

21. Sell KM, Ghigiarelli JJ, Prendergast JM, Ciani GJ, Martin J, Gonzalez AM. Comparison of V̇o2peak and V̇o2max at Different Sampling Intervals in Collegiate Wrestlers. Journal of Strength and Conditioning Research. 2021;35:2915–7. <https://doi.org/10.1519/JSC.0000000000003887>

22. Midgley AW, McNaughton LR, Carroll S. Effect of the o2 time-averaging interval on the reproducibility of o2max in healthy athletic subjects. Clinical Physiology and Functional Imaging. 2007;27:122–5. <https://doi.org/10.1111/j.1475-097X.2007.00725.x>

23. Martin-Rincon M, González-Henríquez JJ, Losa-Reyna J, Perez-Suarez I, Ponce-González JG, La Calle-Herrero J de, et al. Impact of data averaging strategies on V̇O2max assessment: Mathematical modeling and reliability. Scandinavian Journal of Medicine & Science in Sports. 2019;29:1473–88. <https://doi.org/10.1111/sms.13495>

24. Hill DW, Stephens LP, Blumoff-Ross SA, Poole DC, Smith JC. Effect of sampling strategy on measures of V̇O2peak obtained using commercial breath-by-breath systems. European Journal of Applied Physiology. 2003;89:564–9. <https://doi.org/10.1007/s00421-003-0843-1>

25. Scheadler CM, Garver MJ, Hanson NJ. The gas sampling interval effect on v̇O2peak is independent of exercise protocol. Medicine & Science in Sports & Exercise. 2017;49:19111916. <https://doi.org/10.1249/MSS.0000000000001301>

26. Robergs RA, Burnett A. Methods used to process data from indirect calorimetry and their application to VO2max. Journal of Exercise Physiology Online. 2003;6:44–57. <http://connection.ebscohost.com/c/articles/21794702/methods-used-process-data-from-indirect-calorimetry-their-application-vo2max>

27. ATS/ACCP. ATS/ACCP statement on cardiopulmonary exercise testing. American Journal of Respiratory and Critical Care Medicine. 2003;167:211–77. <https://doi.org/10.1164/rccm.167.2.211>

28. Weir J, Koerner S, Mack B, Masek J, Vanderhoff D, Heiderscheit B. VO2 plateau detection in cycle ergometry. Journal of Exercise Physiology Online. 2004;7:55–62.

29. Robergs RA, Dwyer D, Astorino T. Recommendations for Improved Data Processing from Expired Gas Analysis Indirect Calorimetry. Sports Medicine. 2010;40:95–111. <https://doi.org/10.2165/11319670-000000000-00000>

30. Myers J, Walsh D, Sullivan M, Froelicher V. Effect of sampling on variability and plateau in oxygen uptake. Journal of Applied Physiology. 1990;68:404–10. <https://doi.org/10.1152/jappl.1990.68.1.404>

31. Pauw KD, Roelands B, Cheung SS, Geus B de, Rietjens G, Meeusen R. Guidelines to Classify Subject Groups in Sport-Science Research. International Journal of Sports Physiology and Performance. 2013;8:111–22. <https://doi.org/10.1123/ijspp.8.2.111>

32. Decroix L, Pauw KD, Foster C, Meeusen R. Guidelines to Classify Female Subject Groups in Sport-Science Research. International Journal of Sports Physiology and Performance. 2016;11:204–13. <https://doi.org/10.1123/ijspp.2015-0153>

33. Mancini DM, Eisen H, Kussmaul W, Mull R, Edmunds LH, Wilson JR. Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. Circulation. 1991;83:778–86. <https://doi.org/10.1161/01.cir.83.3.778>

34. Foster ED, Deardorff A. Open Science Framework (OSF). Journal of the Medical Library Association. 2017;105:203–6. <https://doi.org/10.5195/jmla.2017.88>

35. Akker O van den, Peters G-J, Bakker C, Carlsson R, Coles NA, Corker KS, et al. Inclusive systematic review registration form. <https://doi.org/10.31222/osf.io/3nbea>

36. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022. <https://www.r-project.org/>

37. RStudio Team. RStudio: Integrated development environment for r. Boston, MA: RStudio, PBC; 2022. <http://www.rstudio.com/>

38. Tricco AC, Lillie E, Zarin W, O’Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Annals of Internal Medicine. 2018;169:467–73. <https://doi.org/10.7326/M18-0850>

39. Quittmann OJ, Foitschik T, Vafa R, Freitag F, Spearmann N, Nolte S, et al. Augmenting the metabolic profile in endurance running by maximal lactate accumulation rate. under review.

40. Quittmann OJ, Schwarz YM, Nolte S, Fuchs M, Gehlert G, Slowig Y, et al. Relationship between physiological parameters and time trial performance over 1, 2 and 3 km in trained runners. unpublished.

41. Nolte S. Spiro: Manage data from cardiopulmonary exercise testing. 2022. <https://doi.org/10.5281/zenodo.5816170>

42. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. <https://doi.org/10.1136/bmj.n71>

43. Maturana FM, Schellhorn P, Erz G, Burgstahler C, Widmann M, Munz B, et al. Individual cardiovascular responsiveness to work-matched exercise within the moderate- and severe-intensity domains. European Journal of Applied Physiology. 2021;121:2039–59. <https://doi.org/10.1007/s00421-021-04676-7>

44. Blackie SP, Fairbarn MS, McElvaney NG, Wilcox PG, Morrison NJ, Pardy RL. Normal values and ranges for ventilation and breathing pattern at maximal exercise. CHEST. 1991;100:136–42. <https://doi.org/10.1378/chest.100.1.136>

45. Maturana FM. Whippr: Tools for manipulating gas exchange data. 2022. <https://github.com/fmmattioni/whippr>

46. Koschate J, Cettolo V, Hoffmann U, Francescato MP. Breath-by-breath oxygen uptake during running: Effects of different calculation algorithms. Experimental Physiology. 2019;104:1829–40. <https://doi.org/10.1113/EP087916>