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

Bachelor thesis

from

Simon Nolte

German Sport University Cologne

Cologne 2022

Thesis supervisor:
Dr. Oliver Jan Quittmann
Institute of Movement and Neurosciences

Affirmation in lieu of an oath

Herewith I affirm in lieu of an oath that I have authored this Master thesis independently and did not use any other sources and tools than indicated. All citations, either direct quotations or passages which were reproduced verbatim or nearby-verbatim from publications, are indicated and the respective references are named. The same is true for tables and figures. I did not submit this piece of work in the same or similar way or in extracts in another assignment.

____

Personally signed

Zusammenfassung (German abstract)

Table of contents

List of Figures

List of Tables

1 Introduction

1.1 Background

1.1.1 Maximum Oxygen Uptake: Definition and Relevance

Performance in endurance sports is limited by the human physiology. Athletes need to supply energy to the contracting muscles for locomotion, a process that happens mainly via the oxidative phosphorylation. A higher maximal activity of the oxidative phosphorylation allows to supply more energy, and thus to move faster (Holloszy & Coyle, 1984). On a whole-body level the highest activity of oxidative phosphorylation can be approximated by measuring the maximum oxygen uptake ($\dot{V}O_{2max}$).

$\dot{V}O_{2max}$ is defined as the highest rate a body can consume oxygen. Determined by measuring gas exchange data, it is one of the most common assessed exercise parameters in sports science and medicine. $\dot{V}O_{2max}$ highly corresponds with endurance performance in heterogeneous groups (Costill et al., 1973; Reaburn & Dascombe, 2008; Tanaka et al., 1990) and can be regarded the most relevant physiological parameter for predicting endurance performance (Bassett, 2000). Many training programs aim to target the $\dot{V}O_{2max}$, and studies evaluate the quality of interventions by measuring changes in $\dot{V}O_{2max.}$

1.1.2 Measuring the Maximum Oxygen Uptake

Researchers measure $\dot{V}O_{2max}$ during exercise tests to exhaustion. In a laboratory setting such tests commonly take place on treadmills or cycling ergometer. Fatigue prior to reaching the $\dot{V}O_{2max}$ results in underestimating its true value. Therefore most exercise protocols consist of a continuous or stepwise, fast increase in load. Buchfuhrer et al. (1983) and Yoon et al. (2007) suggested an optimal duration of 8-12 or 8-10 minutes for testing $\dot{V}O_{2max}$. While this narrow limits have been challenged, considerably shorter or longer protocols are advised against (Midgley et al., 2008). Despite protocol type and protocol duration, factors such as motivation (Midgley et al., 2018), exercise modality (Myers et al., 1991) and biological variability (Katch et al., 1982) bias the $\dot{V}O_{2max}$ determination.

Early research in $\dot{V}O_{2max}$ used bags to collect expired air for later analysis— the so called Douglas bag method (Shephard, 2017). Today's modern metabolic carts allow for simultaneous collection and analysis of gases, using either mixing chambers (which often sample data at fixed time intervals, e.g., 15 seconds) or breath-by-breath methods (Macfarlane, 2001). While mixing chambers are regarded as more reliable (Beijst et al., 2013), measuring breath-by-breath generates data with a higher temporal resolution (Roecker et al., 2005). The measuring method and the metabolic cart model can influence the measured oxygen uptake — and as such the $\dot{V}O_{2max}$ (Miles et al., 1994; Winkert et al., 2021). But even the same oxygen uptake data generated during an exercise test may result in varying outcomes when processed differently.

1.1.3 Peak vs. Maximum Oxygen Uptake

Whether an observed peak in oxygen uptake corresponds to the true maximum has been extensively discussed in the past decades (Green & Askew, 2018; Howley et al., 1995; Poole et al., 2008; Poole & Jones, 2017; Taylor et al., 1955). To identify a 'true' maximum, researchers commonly evaluate a set of parameters measured during the ramp test — the so called '$\dot{V}O_{2max}$ criteria' (Howley et al., 1995). If these criteria are (partly) not fulfilled, researchers are advised to speak of a peak oxygen uptake instead of $\dot{V}O_{2max}$ (Howley et al., 1995). In this thesis I will not distinguish between peak and maximum oxygen uptake, as the criteria for $\dot{V}O_{2max}$ (e.g. the primary criterion of a plateau in oxygen uptake or the secondary criterion of the maximum respiratory quotient) do by themselves heavily rely on the data processing strategy used (Astorino, 2009). I thus speak of $\dot{V}O_{2max}$ as the maximum oxygen uptake measured during an appropriate exercise test regardless of any $\dot{V}O_{2max}$ criteria.

1.2 Previous Research

Measured breath-by-breath data is noisy (see grey points in Figure 1). 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. For interpretation the raw data requires some form of processing.

Different data processing strategies influence measured parameters of gas exchange. As soon as when the first automated systems for gas exchange measurement were available, Matthews et al. (1987) reported that different processing strategies lead to different $\dot{V}O_{2max}$ values. Since then many researchers investigated the influence of different data averaging intervals on $\dot{V}O_{2max}$, unsurprisingly showing that the shorter the calculation interval, the higher the $\dot{V}O_{2max}$ obtained (Astorino, 2009; Johnson et al., 1998; Sell et al., 2021). Midgley et al. (2007) found differences between data processing strategies for $\dot{V}O_{2max}$, but no difference in the reliability of these, highlighting that no optimal strategy exists in terms of reliability, but that consistency of strategies is key when comparing outcomes. Based on own data in sedentary and moderately trained individuals, Martin-Rincon et al. (2019) provided linear-log equations to compare mean $\dot{V}O_{2max}$ 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 (Hill et al., 2003; Scheadler et al., 2017).

Differences in $\dot{V}O_{2max}$ due to data processing can cause serious implications in practice. They hinder the comparability of data from studies within meta-analysis (Martin-Rincon et al., 2019) or the assessment of longitudinal data from athletes that participated in diagnostics with differing data processing. Data processing strategies directly affect the occurrence of a plateau in oxygen uptake, the primary criterion of $\dot{V}O_{2max}$ (Astorino, 2009). Crucially, in situation where individuals are classified by their $\dot{V}O_{2max}$ — for example when describing the training status of a study population (Decroix et al., 2016; Pauw et al., 2013) or evaluating patients for a heart transplantation (Mancini et al., 1991) — differing processing strategies can lead to misclassifications (Johnson et al., 1998).

Calls to standardize data processing strategies of gas exchange data are frequent (Hill et al., 2003; Johnson et al., 1998; Myers et al., 1990; Scheadler et al., 2017). Howley et al. (1995) argued to use longer calculation intervals (60 seconds), as shorter intervals may introduce bias towards extreme data values and thus could systematically overestimate true $\dot{V}O_{2max}$. Based on synthetic and experimental data, Robergs & Burnett (2003a) opposed these conclusion, stating that shorter calculation intervals are less erroneous. They further argued to use breath-based moving averages instead of time averages. In terms of data processing, the current ATS/ACCP guidelines for cardiopulmonary exercise testing remain vague, stating

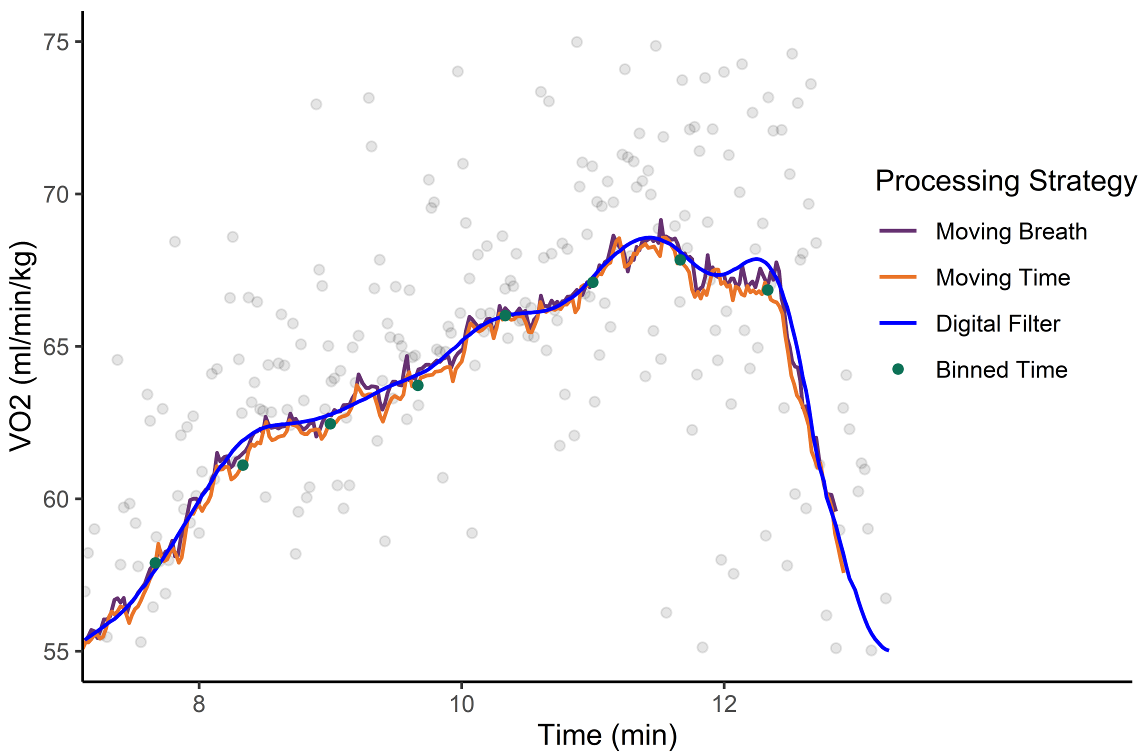

Figure 1: Raw breath by breath data processed by different strategies during the final minutes of a ramp test to exhaustion. Grey points display oxygen uptake from the single breaths. The moving averages were calculated over 30 breaths and seconds, respectively. The binned average was calculated over 30-second intervals with the mean values aligned to the center of each interval. For digital filtering a forward-backward (zero phase) low-pass Butterworth filter with the parameters 0.04 Hz (cut-off frequency) and 3 (order) for each filter was applied. For more details on the used data processing strategies, see the methods section.

that *“30 to 60-second intervals for averaging data are recommended, although 20-seconds intervals may be acceptable*” (ATS/ACCP, 2003). In a classic opinion piece, Robergs et al. (2010) argued for using digital filtering (i.e. a low-pass Butterworth filter) for processing gas exchange data, a strategy initially introduced for gas exchange measures by Weir et al. (2004). However, Robergs et al. (2010) acknowledged the lack of accessible software implementing such procedures.

In absence of one commonly accepted method, data processing varies among the literature. In light of the influence on outcome variables, many articles highlighted the need to report processing strategies in research articles (Astorino, 2009; Johnson et al., 1998; Myers et al., 1990; Scheadler et al., 2017). Midgley et al. (2007) were the first to evaluate reported data processing strategies for breath-by-breath analyses in selected journals. They found that almost all studies reporting their methods choose 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. (2007) was not systematic and its methods were not described in detail. Robergs et al. (2010) 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, which 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. Practices in reporting and processing may have changed more than a decade after the publication of such numbers.

1.3 Aim

In this thesis I will review the usage and reporting of different data processing strategies in the scientific literature and investigate their influence on the $\dot{V}O_{2max}$.

To date, the practices of data processing to determine $\dot{V}O_{2max}$ in the actual scientific work remain largely unknown. Previous research on this topic has only performed unsystematic searches (Midgley et al., 2007) or non-representative surveys (Robergs et al., 2010). Moreover, recent recommendation (Robergs et al., 2010) as well as advances in measurement devices and analysis software may have changed processing practices in the past 15 years.

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 extensively compared different averaging intervals, but not averaging types (e.g. moving breath vs. binned time) (Astorino, 2009; Midgley et al., 2007). Martin-Rincon et al. (2019) based their comparisons on a data set of sedentary or recreational trained athletes using two different metabolic carts. Therefore motivation and measurement devices may have interacted with the influence of processing strategies. Differences in $\dot{V}O_{2max}$ due to data processing strategies may have serious implication for individuals (Johnson et al., 1998), yet almost all comparisons only report mean differences between strategies. No research has yet compared a variety of systematically derived strategies among a group of well-trained individuals using a standardized measurement set-up.

This thesis will first map current practices of data processing for $\dot{V}O_{2max}$ determination by the means of a systematic scoping review of current scientific literature. Secondly, I will compare different data processing strategies in relation to the most common applied on a set of 72 standardized treadmill tests in well-trained individuals. The results will help to compare $\dot{V}O_{2max}$ data derived from different processing methods among studies and in individuals. The review allows to assess the implementation of current data processing recommendations and to find malpractices in reporting. The results build a basis for providing new recommendations for data processing and its reporting in determining the $\dot{V}O_{2max.}$

2 Methods

The work presented in this thesis was preregistered before the start of the project on the Open Science Framework (Foster & Deardorff, 2017), following the 'Inclusive Systematic Review Registration Form' (Akker et al., 2020). Any deviations from the preregistration are indicated in the 'Transparent Changes' document (Appendix A.1). Major deviations will also be explicitly stated within the methods section. All data and code of this research project can be found on GitHub.

I conducted all analyses using R Version 4.1.2 (R Core Team, 2022) within R Studio Version 2022.2.2.485 (RStudio Team, 2022). The thesis was entirely written in Quarto Version 0.9.380 (Allaire et al., 2022) . The attached packages and default settings are documented in Appendix A.4.

2.1 Systematic Scoping Review

The aim of the scoping review was to systematically map current practices of data processing for $\dot{V}O_{2max}$ determination in the scientific literature. Since determining $\dot{V}O_{2max}$ is a far too common procedure to perform an exhaustive search, I randomly sampled 500 articles that referred to $\dot{V}O_{2max}$ or similar keywords. Data on processing strategies were extracted from all sampled articles that directly measured $\dot{V}O_{2max}$ using an appropriate testing procedure in humans.

The review was performed in accordance to the PRISMA extension for Scoping reviews (Tricco et al., 2018). Appendix A.2 contains the corresponding reporting checklist.

2.1.1 Search & Screening

The article search was conducted on 16th March 2022 using PubMed and Web of Science. The search included articles published from 2017 to 2022 referring to 'maximum oxygen uptake' or equivalent terms in title, abstract or keywords. Table 1 shows the exact search terms used.

Table 1: Search strings for the systematic scoping review.

Source	Search String
PubMed	`((((("maximum oxygen uptake") OR ("maximal oxygen uptake")) OR ("VO2max")) OR ("maximum oxygen consumption")) OR ("maximal oxygen consumption")) AND (("2017/01/01"[Date - Publication] : "3000"[Date - Publication]))`
Web of Science	`(((((ALL=("maximum oxygen uptake")) OR ALL=("maximal oxygen uptake")) OR ALL=("VO2max")) OR ALL=("maximum oxygen consumption")) OR ALL=("maximal oxygen consumption")) AND PY=(2017-2022)`

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. After removing the duplicates I conducted an automated title scanning to exclude results that were likely no original research articles. Table 2 displays the exclusion terms for the automated title screening.

Table 2: Exclusion terms for the automated title screening after removal of duplicates. If the title matched at least one of the given terms, the article was excluded.

Search terms for exclusion by title
review, correction, meta-analysis, comment, retraction, editorial, erratum, reply

In accordance with the preregistration I 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 $\dot{V}O_{2max}$ 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. This meant removing any further information not relevant for the screening—such as authors or journal—leaving only the title, abstract and an ID of the article (see Appendix A.3 for an example). Two researchers independently scanned the abstracts to filter those that matched one of the exclusion criteria shown in Table 3. When the screeners disagreed in their assessments, they resolved the conflict by discussion.

Table 3: Exclusion criteria for the screening process. During the abstract screening only the criteria marked with an asterix were evaluated; during full text screening all criteria were assessed.

	Criterium	Details
A*	not in English	Full text only available in non-English language
B*	no primary research	research was no original investigation or only a reanalysis of data
C*	research not in humans	research was conducted in animals
D*	$\dot{V}O_{2max}$ only estimated	$\dot{V}O_{2max}$ was only approximated by means of a predictive equation
E	no appropriate test protocol	Protocol for $\dot{V}O_{2max}$ testing did either not include exercise to voluntial exhaustion or was to long (>20 min) for a reliable estimate
F	no information regarding the exclusion criteria	crucial information on $\dot{V}O_{2max}$ testing that allowed the evaluation of the other exclusion criteria were missing

After the abstract screening I retrieved the full texts for all articles remaining in the review. The full texts were again independently scanned by two researchers to include only those articles that measured $\dot{V}O_{2max}$ using an appropriate testing procedure in humans (see Table 3 for the detailed full-text exclusion criteria). Conflicts were resolved by discussion.

All data exclusion steps are documented in an Markdown script on GitHub.

2.1.2 Data Extraction

I 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 $\dot{V}O_{2max}$ (primary, secondary, other)
- data preprocessing (e.g., filtering)
- data processing software
- interpolation procedure
- data processing/determination of $\dot{V}O_{2max}$:
  - type (time average, breath average, digital filtering)
  - alignment (rolling, binned, …)
  - interval (in seconds or breaths, parameters for filtering)
- reference for the used data processing strategy

The criteria ‘type of outcome’ and ‘reference’ were added to the extraction list after the abstracts had been scanned, thus they were not stated in the preregistration. All extracted data is available as a csv file on GitHub.

2.1.3 Data Synthesis

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

2.2 Experimental Comparison

To determine the influence the most common data processing strategies have on the determination of $\dot{V}O_{2max}$, I compared them on a set of already collected gas exchange data from ramp tests in running.

2.2.1 Data Source

A total of N = 72 exercise tests were analysed for this study. Due to a miscalculation, the preregistration had incorrectly stated a number of 76 tests. The data was from previous research on the metabolic profile of endurance runners (Quittmann et al., under review, unpubl.). The tested individuals were experienced distance runners (15 female, 54 male; three of the males participated in both studies). The $\dot{V}O_{2max}$ tests were conducted in March to September 2019 (Quittmann et al., under review) and March to October 2021 (Quittmann et al., unpubl.) using an identical exercise protocol. Participants run on a treadmill (saturn 300/100, h/p/cosmos sports & medical 127 GmbH, Nussdorf-Traunstein, Germany) with 1% inclination for ten minutes at a velocity of 2.8 $m \cdot s^{-1}$ as a warm-up. After preparing the gas exchange measures, they started a ramp protocol with an initial speed of 2.8 $m \cdot s^{-1}$ for two minutes and subsequently increased velocity by 0.15 $m \cdot s^{-1}$ 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 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% O_2, 6% CO_2) before each measurement. The measured breath-by-breath data is available on GitHub.

2.2.2 Data Processing

The spiro Package Version XXX for R (Nolte, 2022) processed the raw gas exchange data. It includes various algorithms to calculate $\dot{V}O_{2max}$ with user-defined parameters on given data. The full analysis script is available on GitHub.

2.2.2.1 Time based averaging

Moving time based averages were calculated by initially linearly interpolating the breath-by-breath data to seconds. Subsequently a (center aligned) moving average was calculated over a defined timespan. These processing steps are implemented in the spiro package (Nolte, 2022). For calculating the moving average over 30 seconds for example, I used the functions `spiro(data) |> spiro_max(30)`.

Binned time averages were calculated using a custom function (available in the analysis script on GitHub). Breath-by-breath data was initially interpolated to full seconds and then binned into consecutive interval of a constant length. 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.

2.2.2.2 Breath based averaging

Breath based moving averages were calculated on the raw data. As this functionality is implemented in the spiro package, I used the functions `spiro(data) |> spiro_max("30b")` for an exemplary 30-breath long averaging interval.

2.2.2.3 Butterworth filtering

Butterworth filters are a class of recursive digital filters. Robergs et al. (2010) and Weir et al. (2004) argue that these more advanced processing strategies are superior to the traditional moving or binned average approach for analysing gas exchange data. However Robergs et al. (2010) missed to account for the time lag introduced by Butterworth filters. Therefore I applied a zero-phase Butterworth filter by means of a forward-backward filtering. While this effectively zeroes out the time lag, the resulting filter is non-causal, which means it can non be used online (i.e. in real time). However, for the present application, an offline filter is sufficient. The forward-backward filtering also introduces transients at both ends of the signal (Gustafsson, 1996). I therefore padded the reverted signal at both the start and the end of the array (identical to the 'even padtype' in Python's `scipi.signal.filtfilt()` function (Virtanen et al., 2020)). I used 3 as the order parameter and 0.04 as the cut-off frequency for each filter (Robergs et al., 2010). Note that due to the forward-backward approach both the overall order and overall cut-off frequency are different from these values. The described approach is implemented in the spiro package (Nolte, 2022), and can be used by calling, for example, the functions `spiro(data) |> spiro_max("0.04fz3")`.

2.2.3 Comparison of methods

To compare different processing methods within and between individuals, I choose to express the $\dot{V}O_{2max}$ normalized to a reference procedure. The reference procedure was chosen as being the most commonly applied in current literature as determined by the systematic review. Individual $\dot{V}O_{2max}$ values were expressed in reference to this procedure, where a value of 1 means that the processing method yields exactly the same $\dot{V}O_{2max}$ value as the reference method. I calculated the data for all integer parameter values within the range of the values found in the literature during the review. On a group level I calculated the median and 10%- and 90%-quantiles of each processing strategy.

In an additional exploratory analysis I investigated the respiratory rate (number of breaths per minute) during the ramp tests. This may help to understand how breath-based and time-based data processing methods relate to each other.

3 Results

3.1 Systematic Scoping Review

Initial search yielded 7529 results of which 4364 remained after automated filtering and removal of duplicates (see flow diagram in Figure 2). Out of the random sample (n = 500), 242 articles were included in the final analysis.

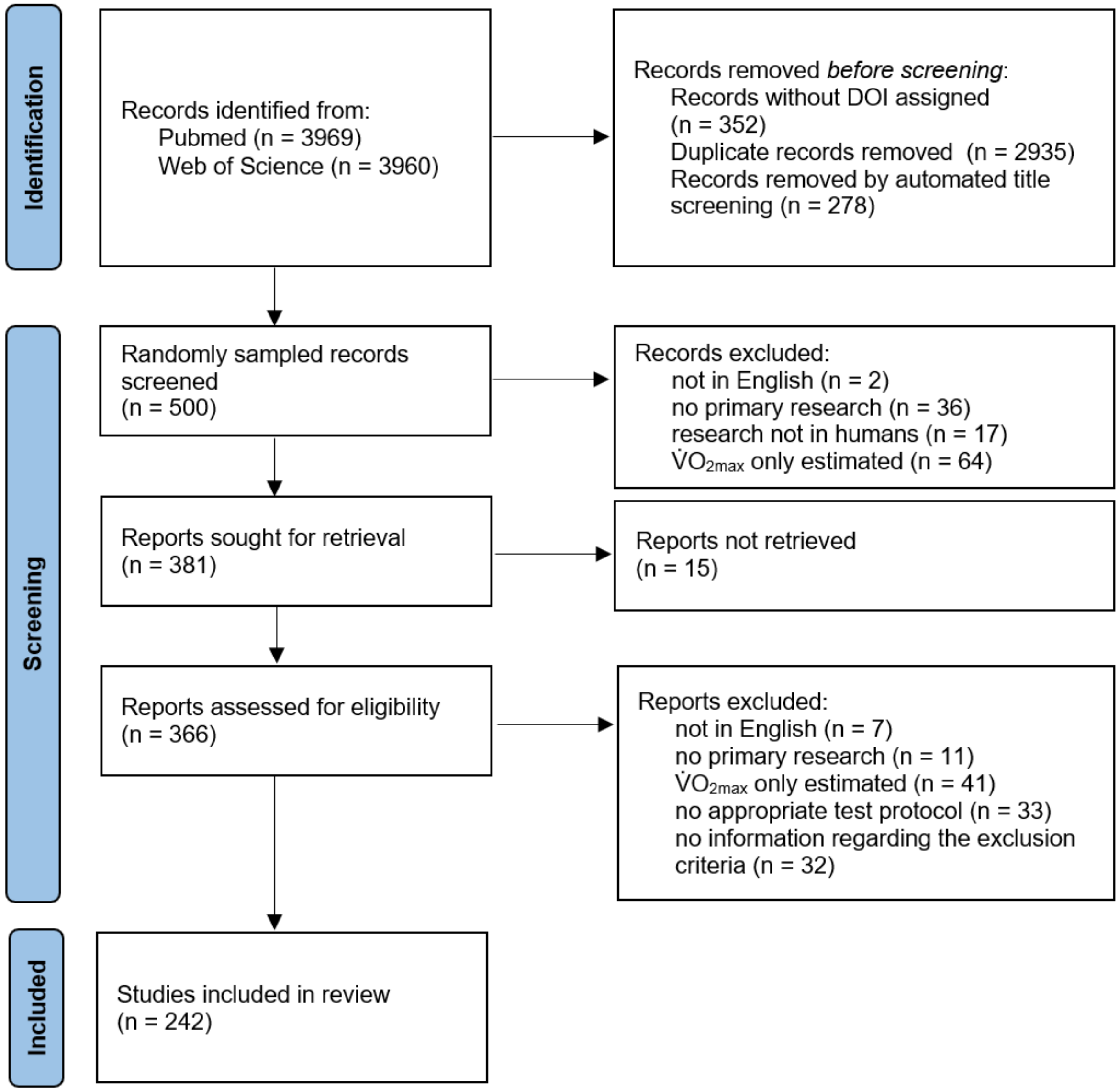

Figure 2: Flow diagram for the systematic scoping review in accordance with the PRISMA 2020 Statement (Page et al., 2021)

Reporting practices of the methodology of gas exchange measures differed widely across the literature (see Table 4). More than half (51.8%) of the articles did not report any information regarding their data processing strategy. One in twenty articles (5.4%) provided a rationale for their used strategy. Only a single article (Maturana et al., 2021) reported information regarding all the investigated criteria.

Table 4: Percentage of studies that provided details on the different characteristics of oxygen uptake data processing. *only examined within the subgroup of studies using breath-by-breath measurements

Metabolic cart	Preprocessing	Software	Processing Strategy	Reference
88.0%	5.6%*	15.0%*	55.8%	5.8%

Out of the authors that provided information and collected breath-by-breath measurements, most (79.5%) utilized binned averages to determine $\dot{V}O_{2max}$. Moving time averages, or breath-based averages were uncommon (see Figure 3). No study used methods of digital filtering to determine $\dot{V}O_{2max}$.

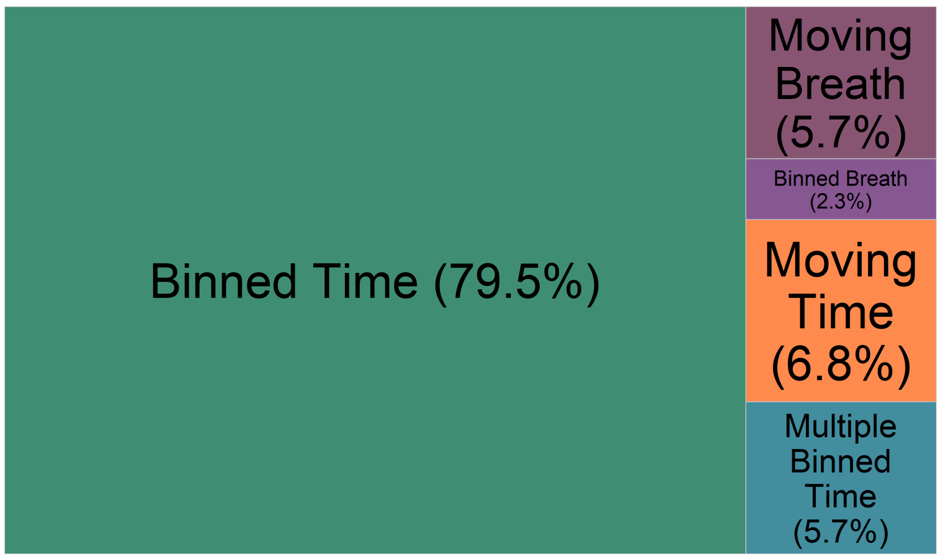

Figure 3: Data strategies for processing breath-by-breath data in the reviewed literature (n = 88).

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 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 interval for time-based averages of mixing chamber and breath-by-breath devices ranged from 5 to 60 seconds (see Figure 4). 30 second length intervals were most common to define $\dot{V}O_{2max}$, while authors also often employed shorter (10-20 s) and longer (60 s) periods.

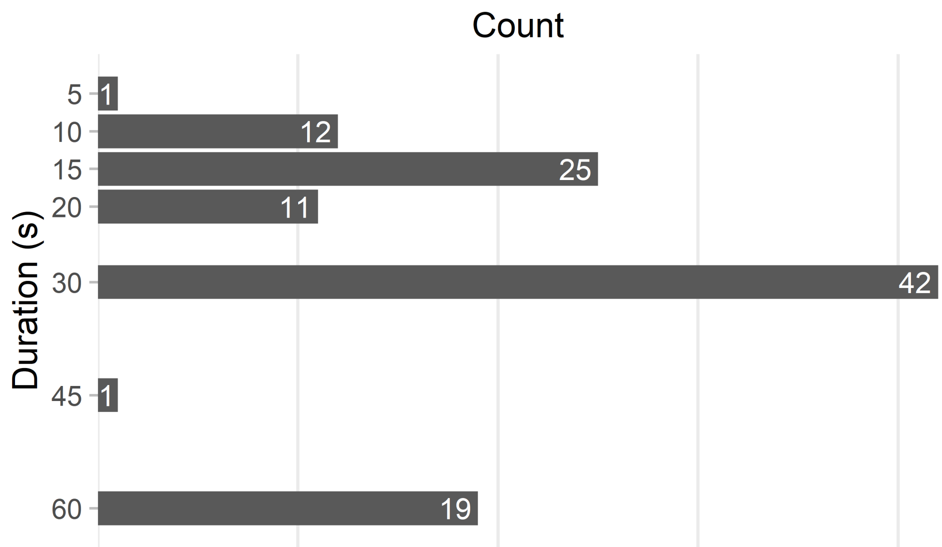

Figure 4: Total durations of the calculation interval of $\dot{V}O_{2max}$ in the reviewed studies.

3.2 Experimental Comparison

The average $\dot{V}O_{2max}$ as determined by a binned 30-second average was 62.2±6.3 $ml \cdot min^{-1} \cdot kg^{-1}$ (mean ± standard deviation). Applying different data processing strategies for $\dot{V}O_{2max}$ determination leads to different outcome values (see Figure 5).

Binned time averages systematically generate $\dot{V}O_{2max}$ values compared to moving averages. Decreasing the calculation interval to 5 seconds leads to a 3-4% median increase of $\dot{V}O_{2max}$ 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 $\dot{V}O_{2max}$ over all calculation intervals. This corresponds to the trained athletes reaching respiratory rates around 60 min^{-1} in the final minute of the exercise test (see Figure 6).

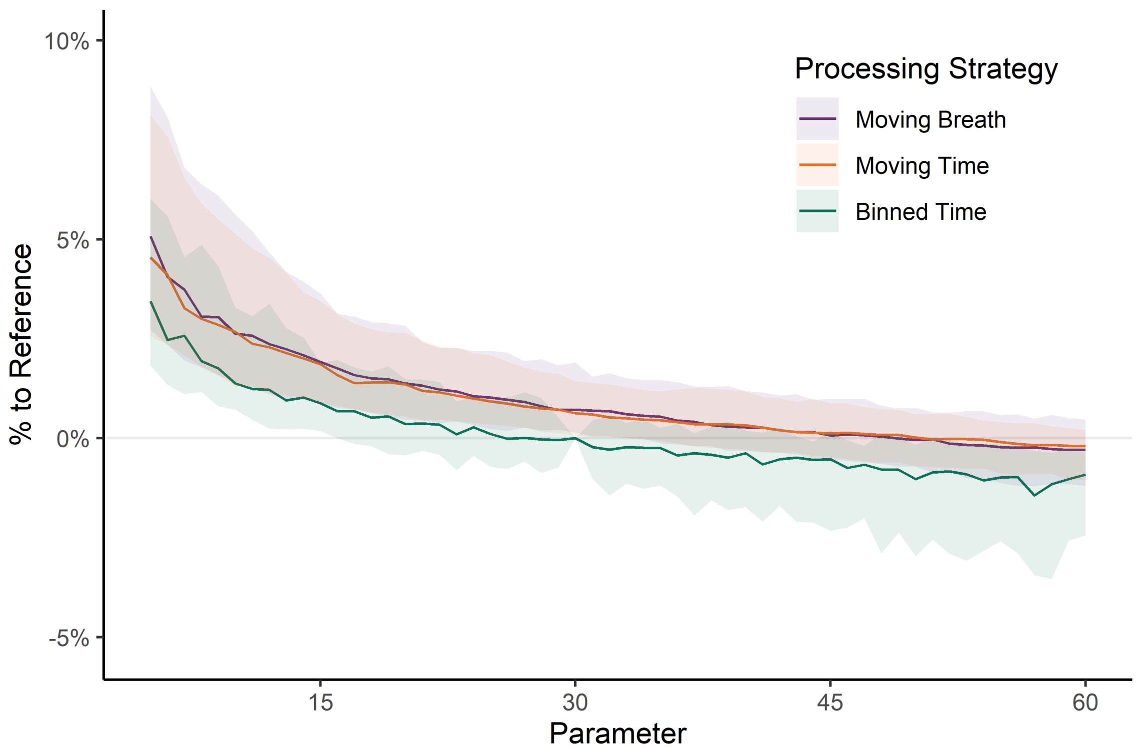

Figure 5: $\dot{V}O_{2max}$ varies by data processing strategy. Values are expressed relative to the $\dot{V}O_{2max}$ from a 30-second binned average — the most common strategy as determined by the review. Solid lines display the median, the shaded area marks the interval between 10th and 90th percentile. Using moving average leads to systematically higher $\dot{V}O_{2max}$ values compared to binned time averages. Changing the averaging interval (in seconds or breaths) can lead to median changes in $\dot{V}O_{2max}$ as large as 5%,.

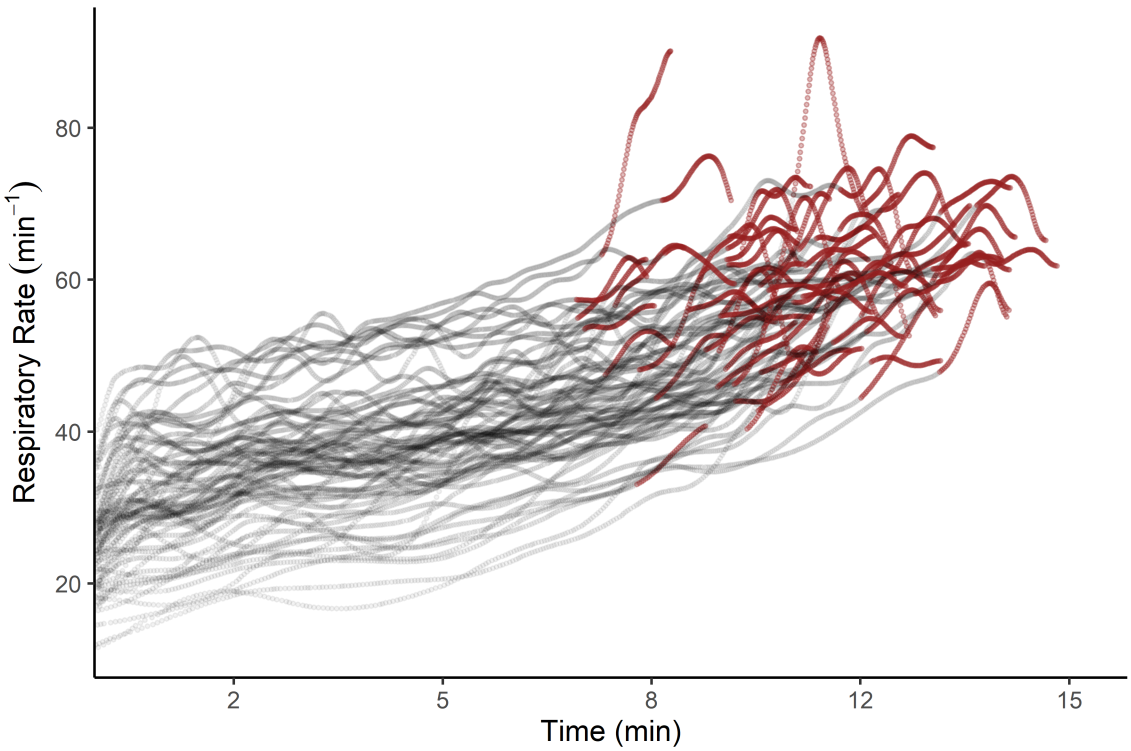

Figure 6: Respiratory Rates peak around 60 min^{-1} in the ramp tests. The red segments correspond to the last minute before exhaustion of each individual (n = 72).

4 Discussion

The collected data shows, that current research uses a variety of strategies to determine $\dot{V}O_{2max}$, which directly influences the values obtained. Many articles only incompletely report on their methods and use outdated strategies. These practices hinder the validity and reproducibility of $\dot{V}O_{2max}$ measurement.

4.1 Current state of data processing

Despite calls to use moving averages (Robergs et al., 2010; Robergs & Burnett, 2003b), binned time averages remain the most commonly used data processing strategy to determine $\dot{V}O_{2max}$ in the reviewed literature (see Figure 3). These numbers are generally in line with the findings of the non-systematic search by Midgley et al. (2007) and the survey by Robergs et al. (2010). It is somewhat surprising, that practices have not changes in recent years despite the publication of recommendations, that highly discourage researchers from using binned averages (Robergs et al., 2010). Using binned time averages leads to systematically lower $\dot{V}O_{2max}$ values as compared to moving averages (see Figure 5). The peak in oxygen uptake may fall in between two average intervals, resulting in an underestimation of $\dot{V}O_{2max}$ (see Figure 1 for an example). Binned time average 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, my review demonstrates that they remain very common in scientific literature.

Breath-based averages seem to be more common (≈ 8%) than reported previously (< 1%) (Midgley et al., 2007), but less common than assessed in self-reporting (≈ 17%) (Robergs et al., 2010). The increasing proportion of breath-based averages may be explained by publications in the recent years advocating for them (Robergs et al., 2010; Robergs & Burnett, 2003b). However the proportion of breath-based averages is much smaller than that of binned time averages, which have not been recommended in such a way.

The length of the calculation intervals is highly diverse within the literature (see Figure 4). This may reflect contradictory recommendation (Howley et al., 1995; Robergs & Burnett, 2003b). As long as there is no consensus on the arguments speaking for shorter or longer calculation intervals, there appears to be no optimal interval duration. As different interval duration can heavily influence $\dot{V}O_{2max}$ by more than 5% on a median level (see Figure 5), the exact reporting of the data processing strategy remains essential for interpretation.

The exact reasons for exercise scientist ignoring most recommendations by using binned time averages remain unknown. Some researchers may simply not be aware of the impact different data processing strategies have on the $\dot{V}O_{2max}$, but past publications on this issues have been widely cited (Astorino, 2009; Midgley et al., 2007; Robergs et al., 2010) and should be known to most scientists in this field. Researchers may also use binned averages for traditional reasons. Douglas bags, as well as many mixing chamber devices, measure the oxygen uptake over fixed time intervals, producing data appeerently similar to those by a binned time average of breath-by-breath data. But comparability with older data should only be an issue when using data acquired by different measurement methods within on analysis. Current studies using new breath-by-breath data do not reasonably need to rely on outdated methods of data processing.

A major source for choosing suboptimal processing strategies may be limitations by analysis software. My review shows that most researchers use the software of the metabolic cart's manufacturer to analyse 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 is also an reasonable explanation why digital filtering has — despite recommended by Robergs et al. (2010) — 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 malpractices of data processing are likely to be both attributed to a lack of awareness and a lack of easy-to-use software solutions.

4.2 Impact of different data processing

The different data processing strategies found in the literature directly bias the $\dot{V}O_{2max}$ determination (see Figure 5), and as such can influence the classification of individuals, evaluation of training success, and assessment of $\dot{V}O_{2max}$ attainment. In accordance with previous findings (Astorino, 2009; Johnson et al., 1998; Sell et al., 2021) — and pure logic — longer calculation intervals lead to lower $\dot{V}O_{2max}$ values (see Figure 5). The analysed data shows median differences up to 5-7% between processing strategies, which is in accordance with previous research. Some studies reported even higher mean differences up to 20% (Myers et al., 1990), but only when including raw breath-by-breath data in the comparison. The evaluation of unprocessed raw data for its maximum is highly erroneous (see the individual breath data points in Figure 1) and as such not performed in research (see Figure 4); so there is no reason to include it in a comparison to other strategies. While previous research was often conducted in sedentary or recreational individuals, this thesis now presents evidence, that a similar effect of data processing strategies on $\dot{V}O_{2max}$ exists in well-trained athletes.

Binned time averages lead to systematically lower $\dot{V}O_{2max}$ values compared with moving averages, for the reasons explained above. While this general trend has been acknowledged previously (Martin-Rincon et al., 2019), it has not been quantified. The data presented in this thesis suggest a ≈-1% lower median $\dot{V}O_{2max}$ 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 is systematic and as such may bias the evaluation in scenarios where small changes in $\dot{V}O_{2max}$ are important (e.g. in elite sports).

Moving time and moving breath averages lead to almost identical $\dot{V}O_{2max}$ values on a median level (Figure 5). This is hardly surprising, as the athletes in this study reached respiratory rates around 60 min^{-1} (see Figure 6), resulting in equivalent time- and breath-based interval lengths. For an athletic population, $\dot{V}O_{2max}$ values obtained by moving time and moving breath average can approximately used interchangeable. Given that less trained individual will likely display lower respiratory rates during exercise tests to exhaustion (Blackie et al., 1991), this finding will not translate to a sedentary population.

The exact impact of data processing strategies on the $\dot{V}O_{2max}$ is highly individual. Most research did present only comparisons of mean values, with results in accordance with those found here (Martin-Rincon et al., 2019). On an individual level, data processing strategies may impact the $\dot{V}O_{2max}$ in different severity. For example, for 10% of the investigated athletes a binned time average of 5 seconds lead to a $\dot{V}O_{2max}$ <3% higher than by a 30 second average, while for another 10% the $\dot{V}O_{2max}$ was >6% higher (see Figure 5). Current values reported and equation derived compare strategies on a group level (Martin-Rincon et al., 2019), which improves comparability of group results for meta-analyses or group classification. However on a individual level such work does not help. Changes in the impact of different data processing strategies on $\dot{V}O_{2max}$ range from 1-2% in some to more than 10% in other individuals. Hence when evaluating $\dot{V}O_{2max}$ 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 $\dot{V}O_{2max}$ 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.

4.3 Guidelines for data processing and reporting

This thesis focussed on the occurrence and impact of different data processing strategies, but did not investigate their validity in the first place. However the results of the scoping review and experimental comparison allow to specify existing recommendation for data processing and set new guidelines for data reporting.

I highly disregard researchers from continuing to use binned time averages to determine $\dot{V}O_{2max}$. The main reason for such procedure is pure tradition, as it reduces the breath-by-breath data in an inappropriate way leading to a small, yet systematic, underestimation of $\dot{V}O_{2max}$. Moving time or moving breath averages are preferable to binned averages. In a athletic population they may be used equivalent, but in sedentary individuals breath-based averages will likely lead to lower $\dot{V}O_{2max}$ values than time-based averages with the same interval length parameter. While using a constant breath interval will lead to a similar degree of data smoothing regardless of the training status, using a constant time interval will lead to the same physiological time-frame for determining $\dot{V}O_{2max}$ (but with a different degree of data smoothing). From a data processing perspective, I prefer using breath-based moving averages, while keeping in mind, that this may underestimate true $\dot{V}O_{2max}$ in individuals with a low respiratory rate in the end of the exercise test.

Digital filtering seems to be a promising method, as it markedly reduces the variability moving averages have due to single data outliers (see Figure 1). I therefore agree with Robergs et al. (2010), who recommends digital filters over any classical averaging procedure. The exact type of filter and values of filter parameters to be used have not yet been systematically investigated, and it is unclear whether it is possible at all to determine criteria for identifying an optimal filter. However, a Butterworth filter, as proposed by Robergs et al. (2010), seems to produce reasonably smoothed data. Note, that a single Butterworth filter comprises a time lag, a fact highlighted by Weir et al. (2004) that was not acknowledged by Robergs et al. (2010). To allow for a correct phasing of data, as well as for a correct $\dot{V}O_{2max}$ determination when the measurement is shortly terminated after reaching exhaustion, a zero-phase filter is needed. A zero-phase forward-backward Butterworth filter (as suggested by Weir et al. (2004)) seems to produce reasonable smoothing when used with the parameters suggested by Robergs et al. (2010) (low-pass cut-off frequency: 0.04 Hz; third order; see Figure 1). It should be noted, that despite using the same parameters the degree of filtering varies from that by Robergs et al. (2010), as the filter is applied twice, changing the overall magnitude of parameters. Other filtering parameters may be as reasonable as this, but it is unclear how to objectively compare them.

I recommend to use a zero-phase forward-backward Butterworth filter on the raw breath-by-breath data. The $\dot{V}O_{2max}$ is then defined as the highest single filtered data point. An filter order of 3 and a low cut-off frequency of 0.04 Hz appear to be reasonable parameters for each filter. While such data processing has advantages over traditional data processing forms, it requires specialized software.

To compare and evaluate $\dot{V}O_{2max}$ values from different studies, knowledge of the underlying data processing strategy is crucial. The review presented here shows that almost half of the studies that measured $\dot{V}O_{2max}$ did not describe their processing strategy. Other aspects of the data processing, such as outlier filtering or rationale for the chosen procedure were only in rare instances reported (see Table 4). Table 5 lists 7 item that should be reported to provide sufficient information on the data processing strategy used to determine $\dot{V}O_{2max}$. These items may be reported in form of a check-list, as an in-text enumeration or in sentence form. An example paragraph containing all the relevant information for the current study would be:

> *"We measured breath-by-breath data during the ramp tests with a ZAN 600 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 (Nolte, 2022). This strategy produces similar results as that recommended by Robergs et al. (2010), but does not produce a time lag."*

Table 5: Recommendations for reporting data processing strategies to determine $\dot{V}O_{2max}$.

reporting item	description
Metabolic Cart	State the exact device model and manufacturer

reporting item	description
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 $\dot{V}O_{2max}$ (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)

Note that the correct reporting of an exercise test to determine $\dot{V}O_{2max}$ 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, these 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 $\dot{V}O_{2max}$ values.

The results of this thesis suggest that comprehensive reporting facilitates approximate comparisons of $\dot{V}O_{2max}$ 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 may vary between individuals and are potentially influenced by training status. Sharing of the raw metabolic exchange data is the solution to this problem. It allows any researcher to recalculate the $\dot{V}O_{2max}$ using their preferred data processing strategy. Anonymization of raw gas exchange data files should not be a big deal for researchers, since most of these files are structured in a simple way and allow to easily remove any personal information (if this has not been done in the metabolic cart's system before). In terms of reproducibility of $\dot{V}O_{2max}$ determination, an even superior way to correct reporting and raw data sharing is to additionally share data analysis code. This allows any researcher to independently reproduce the $\dot{V}O_{2max}$ processing conducted within a study, but requires the data analysis to take place in a programming, or at least code generating, environment. Luckily such programs for the purpose of analyzing gas exchange data exist as free open-source software (Maturana, 2022; Nolte, 2022).

4.4 Limitations

Due to the sheer quantity of the publications investigating $\dot{V}O_{2max}$, it was not possible to perform an exhaustive review of the literature. The scoping review therefore relies on a random sample which not necessarily exactly captures trends of the whole 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 described 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 durations (e.g., 4x15 second), without describing this correctly. Moreover the exact definitions for building binned averages varied within the literature: While most studies defined the binning periods from the beginning of the exercise, some may have defined them from the endpoint. Additionally some studies reviewed did not define the maximum, but a predefined binned average period as their $\dot{V}O_{2max}$ (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. I did not seperately consider such sub-categories of data processing strategies, as they may not be very common and are often hard to precisely investigate due to ambiguity in reporting.

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 be transferable to different settings, such as individuals with less training background. I did not conduct an formal analysis of the validity or reliability of different data processing strategies, so recommendations regarding their use rely on theoretical derivations and prior research. Since the reliability of different strategies appears to be similar (Martin-Rincon et al., 2019; Midgley et al., 2007) and currently no accepted methods to quantify validity of $\dot{V}O_{2max}$ exist, the presented approach is sufficient to derive recommendations for data processing strategies to determine $\dot{V}O_{2max}$.

5 Conclusion

6 Bibliography

Akker, O. van den, Peters, G.-J., Bakker, C., Carlsson, R., Coles, N. A., Corker, K. S., Feldman, G., Mellor, D. T., Moreau, D., Nordström, T., Pfeiffer, N., Pickering, J., Riegelman, A., Topor, M., Veggel, N. van, & Yeung, S. K. (2020). *Inclusive systematic review registration form*. https://doi.org/10.31222/osf.io/3nbea

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., & Dervieux, C. (2022). *Quarto*. https://doi.org/10.5281/zenodo.5960048

Astorino, T. A. (2009). Alterations in VO2max and the VO2 plateau with manipulation of sampling interval. *Clinical Physiology and Functional Imaging*, *29*(1), 60–67. https://doi.org/10.1111/j.1475-097X.2008.00835.x

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

Bassett, D. R. J. E. T. H. (2000). Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine & Science in Sports & Exercise*, *32*(1), 70. https://journals.lww.com/acsm-msse/Fulltext/2000/01000/Limiting_factors_for_maximum_oxygen_uptake_and.12.aspx

Beijst, C., Schep, G., Breda, E. van, Wijn, P. F. F., & Pul, C. van. (2013). Accuracy and precision of CPET equipment: A comparison of breath-by-breath and mixing chamber systems. *Journal of Medical Engineering & Technology*, *37*(1), 35–42. https://doi.org/10.3109/03091902.2012.733057

Blackie, S. P., Fairbarn, M. S., McElvaney, N. G., Wilcox, P. G., Morrison, N. J., & Pardy, R. L. (1991). Normal values and ranges for ventilation and breathing pattern at maximal exercise. *CHEST*, *100*(1), 136–142. https://doi.org/10.1378/chest.100.1.136

Buchfuhrer, M. J., Hansen, J. E., Robinson, T. E., Sue, D. Y., Wasserman, K., & Whipp, B. J. (1983). Optimizing the exercise protocol for cardiopulmonary assessment. *Journal of Applied Physiology*, *55*(5), 1558–1564. https://doi.org/10.1152/jappl.1983.55.5.1558

Costill, D. L., Thomason, H., & Roberts, E. (1973). Fractional utilization of the aerobic capacity during distance running. *Medicine & Science in Sports & Exercise*, *5*(4), 248252. https://journals.lww.com/acsm-msse/Abstract/1973/00540/Fractional_utilization_of_the_aerobic_capacity.7.aspx

Decroix, L., Pauw, K. D., Foster, C., & Meeusen, R. (2016). Guidelines to Classify Female Subject Groups in Sport-Science Research. *International Journal of Sports Physiology and Performance*, *11*(2), 204–213. https://doi.org/10.1123/ijspp.2015-0153

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association*, *105*(2), 203–206. https://doi.org/10.5195/jmla.2017.88

Green, S., & Askew, C. (2018). $\dot{V}o2$peak is an acceptable estimate of cardiorespiratory fitness but not $\dot{V}o2$max. *Journal of Applied Physiology*, *125*(1), 229–232. https://doi.org/10.1152/japplphysiol.00850.2017

Gustafsson, F. (1996). Determining the initial states in forward-backward filtering. *IEEE Transactions on Signal Processing*, *44*(4), 988–992. https://doi.org/10.1109/78.492552

Hill, D. W., Stephens, L. P., Blumoff-Ross, S. A., Poole, D. C., & Smith, J. C. (2003). Effect of sampling strategy on measures of $\dot{V}O2$peak obtained using commercial breath-by-breath systems. *European Journal of Applied Physiology*, *89*(6), 564–569. https://doi.org/10.1007/s00421-003-0843-1

Holloszy, J. O., & Coyle, E. F. (1984). Adaptations of skeletal muscle to endurance exercise and their metabolic consequences. *Journal of Applied Physiology*, *56*(4), 831–838. https://doi.org/10.1152/jappl.1984.56.4.831

Howley, E. T., Bassett, D. R., & Welch, H. G. (1995). Criteria for maximal oxygen uptake:

Review and commentary. *Medicine & Science in Sports & Exercise*, *27*(9), 12921301. https://journals.lww.com/acsm-msse/Abstract/1995/09000/Criteria_for_maximal_oxygen_uptake__review_and.9.aspx

Johnson, J. S., Carlson, J. J., VanderLaan, R. L., & Langholz, D. E. (1998). Effects of sampling interval on peak oxygen consumption in patients evaluated for heart transplantation. *CHEST*, *113*(3), 816–819. https://doi.org/10.1378/chest.113.3.816

Katch, V. L., Sady, S. S., & Freedson, P. (1982). Biological variability in maximum aerobic power. *Medicine & Science in Sports & Exercise*, *14*(1), 2125. https://journals.lww.com/acsm-msse/Abstract/1982/01000/Biological_variability_in_maximum_aerobic_power.4.aspx

Macfarlane, D. J. (2001). Automated Metabolic Gas Analysis Systems. *Sports Medicine*, *31*(12), 841–861. https://doi.org/10.2165/00007256-200131120-00002

Mancini, D. M., Eisen, H., Kussmaul, W., Mull, R., Edmunds, L. H., & Wilson, J. R. (1991). Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. *Circulation*, *83*(3), 778–786. https://doi.org/10.1161/01.cir.83.3.778

Martin-Rincon, M., González-Henríquez, J. J., Losa-Reyna, J., Perez-Suarez, I., Ponce-González, J. G., de La Calle-Herrero, J., Perez-Valera, M., Pérez-López, A., Curtelin, D., Cherouveim, E. D., Morales-Alamo, D., & Calbet, J. A. L. (2019). Impact of data averaging strategies on V̇O2max assessment: Mathematical modeling and reliability. *Scandinavian Journal of Medicine & Science in Sports*, *29*(10), 1473–1488. https://doi.org/10.1111/sms.13495

Matthews, J. I., Bush, B. A., & Morales, F. M. (1987). Microprocessor Exercise Physiology Systems vs a Nonautomated System: A Comparison of Data Output. *Chest*, *92*(4), 696–703. https://doi.org/10.1378/chest.92.4.696

Maturana, F. M. (2022). *Whippr: Tools for manipulating gas exchange data*. https://github.com/fmmattioni/whippr

Maturana, F. M., Schellhorn, P., Erz, G., Burgstahler, C., Widmann, M., Munz, B., Soares, R. N., Murias, J. M., Thiel, A., & Nieß, A. M. (2021). Individual cardiovascular responsiveness to work-matched exercise within the moderate- and severe-intensity domains. *European Journal of Applied Physiology*, *121*(7), 2039–2059. https://doi.org/10.1007/s00421-021-04676-7

Midgley, A. W., Bentley, D. J., Luttikholt, H., McNaughton, L. R., & Millet, G. P. (2008). Challenging a Dogma of Exercise Physiology. *Sports Medicine*, *38*(6), 441–447. https://doi.org/10.2165/00007256-200838060-00001

Midgley, A. W., Marchant, D. C., & Levy, A. R. (2018). A call to action towards an evidence-based approach to using verbal encouragement during maximal exercise testing. *Clinical Physiology and Functional Imaging*, *38*(4), 547–553. https://doi.org/10.1111/cpf.12454

Midgley, A. W., McNaughton, L. R., & Carroll, S. (2007). Effect of the o2 time-averaging interval on the reproducibility of o2max in healthy athletic subjects. *Clinical Physiology and Functional Imaging*, *27*(2), 122–125. https://doi.org/10.1111/j.1475-097X.2007.00725.x

Miles, D. S., Cox, M. H., & Verde, T. J. (1994). Four commonly utilized metabolic systems fail to produce similar results during submaximal and maximal exercise. *Sports Medicine, Training and Rehabilitation*, *5*(3), 189–198. https://doi.org/10.1080/15438629409512016

Myers, J., Buchanan, N., Walsh, D., Kraemer, M., McAuley, P., Hamilton-Wessler, M., & Froelicher, V. F. (1991). Comparison of the ramp versus standard exercise protocols. *Journal of the American College of Cardiology*, *17*(6), 1334–1342. https://doi.org/10.1016/S0735-1097(10)80144-5

Myers, J., Walsh, D., Sullivan, M., & Froelicher, V. (1990). Effect of sampling on variability and plateau in oxygen uptake. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *68*(1), 404–410. https://doi.org/10.1152/jappl.1990.68.1.404

Nolte, S. (2022). *Spiro: Manage data from cardiopulmonary exercise testing*. https://doi.org/10.5281/zenodo.5816170

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pauw, K. D., Roelands, B., Cheung, S. S., Geus, B. de, Rietjens, G., & Meeusen, R. (2013). Guidelines to Classify Subject Groups in Sport-Science Research. *International Journal of Sports Physiology and Performance*, *8*(2), 111–122. https://doi.org/10.1123/ijspp.8.2.111

Poole, D. C., & Jones, A. M. (2017). Measurement of the maximum oxygen uptake V̇o2max: V̇o2peak is no longer acceptable. *Journal of Applied Physiology*, *122*(4), 997–1002. https://doi.org/10.1152/japplphysiol.01063.2016

Poole, D. C., Wilkerson, D. P., & Jones, A. M. (2008). Validity of criteria for establishing maximal O2 uptake during ramp exercise tests. *European Journal of Applied Physiology*, *102*(4), 403–410. https://doi.org/10.1007/s00421-007-0596-3

Quittmann, O. J., Foitschik, T., Vafa, R., Freitag, F., Spearmann, N., Nolte, S., & Abel, T. (under review). *Augmenting the metabolic profile in endurance running by maximal lactate accumulation rate*.

Quittmann, O. J., Schwarz, Y. M., Nolte, S., Fuchs, M., Gehlert, G., Slowig, Y., Schiffer, A., Foitschik, T., & Abel, T. (unpubl.). *Relationship between physiological parameters and time trial performance over 1, 2 and 3 km in trained runners*.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Reaburn, P., & Dascombe, B. (2008). Endurance performance in masters athletes. *European Review of Aging and Physical Activity*, *5*(1), 31–42. https://doi.org/10.1007/s11556-008-0029-2

Robergs, R. A., & Burnett, A. (2003b). Methods used to process data from indirect calorimetry and their application to VO2max. *Journal of Exercise Physiology Online*, *6*(2), 44–57. http://connection.ebscohost.com/c/articles/21794702/methods-used-process-data-from-indirect-calorimetry-their-application-vo2max

Robergs, R. A., & Burnett, A. (2003a). Methods used to process data from indirect calorimetry and their application to VO2max. *Journal of Exercise Physiology Online*, *6*(2), 44–57. http://connection.ebscohost.com/c/articles/21794702/methods-used-process-data-from-indirect-calorimetry-their-application-vo2max

Robergs, R. A., Dwyer, D., & Astorino, T. (2010). Recommendations for Improved Data Processing from Expired Gas Analysis Indirect Calorimetry. *Sports Medicine*, *40*(2), 95–111. https://doi.org/10.2165/11319670-000000000-00000

Roecker, K., Prettin, S., & Sorichter, S. (2005). Gas Exchange Measurements with High Temporal Resolution: The Breath-by-Breath Approach. *International Journal of Sports Medicine*, *26*(S 1), S11–S18. https://doi.org/10.1055/s-2004-830506

RStudio Team. (2022). *RStudio: Integrated development environment for r*. RStudio, PBC. http://www.rstudio.com/

Scheadler, C. M., Garver, M. J., & Hanson, N. J. (2017). The gas sampling interval effect on v·O2peak is independent of exercise protocol. *Medicine & Science in Sports & Exercise*, *49*(9), 19111916. https://doi.org/10.1249/MSS.0000000000001301

Sell, K. M., Ghigiarelli, J. J., Prendergast, J. M., Ciani, G. J., Martin, J., & Gonzalez, A. M. (2021). Comparison of V̇o2peak and V̇o2max at Different Sampling Intervals in Collegiate Wrestlers. *Journal of Strength and Conditioning Research*, *35*(10), 2915–2917. https://doi.org/10.1519/JSC.0000000000003887

Shephard, R. J. (2017). Open-circuit respirometry: a brief historical review of the use of Douglas bags and chemical analyzers. *European Journal of Applied Physiology*, *117*(3), 381–387. https://doi.org/10.1007/s00421-017-3556-6

Tanaka, K., Takeshima, N., Kato, T., Niihata, S., & Ueda, K. (1990). Critical determinants of endurance performance in middle-aged and elderly endurance runners with heterogeneous training habits. *European Journal of Applied Physiology and Occupational Physiology*, *59*(6), 443–449. https://doi.org/10.1007/BF02388626

Taylor, H. L., Buskirk, E., & Henschel, A. (1955). Maximal oxygen intake as an objective measure of cardio-respiratory performance. *Journal of Applied Physiology*, *8*(1), 73–80. https://doi.org/10.1152/jappl.1955.8.1.73

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J. van der, Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Mulbregt, P. van. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Weir, J., Koerner, S., Mack, B., Masek, J., Vanderhoff, D., & Heiderscheit, B. (2004). VO2 plateau detection in cycle ergometry. *Journal of Exercise Physiology Online*, *7*(2), 55–62.

Winkert, K., Kirsten, J., Kamnig, R., Steinacker, J. M., & Treff, G. (2021). Differences in V·O2max Measurements Between Breath-by-Breath and Mixing-Chamber Mode in the COSMED K5. *International Journal of Sports Physiology and Performance*, *16*(9), 1335–1340. https://doi.org/10.1123/ijspp.2020-0634

Yoon, B.-K., Kravitz, L., & Robergs, R. (2007). V·O2max, protocol duration, and the v·O2 plateau. *Medicine & Science in Sports & Exercise*, *39*(7), 11861192. https://doi.org/10.1249/mss.0b13e318054e304

A Appendix

A.1 Transparent Changes

A.2 Prisma Reporting Checklist

A.3 Blinded abstract example

Development in Adolescent Middle-Distance Athletes: A Study of Training Loadings, Physical Qualities, and Competition Performance

Sampling ID: 030

Jones, TW, Shillabeer, BC, Ryu, JH, and Cardinale, M. Development in adolescent middle-distance athletes: a study of training loadings, physical qualities, and competition performance. J Strength Cond Res 35(12S): S103-S110, 2021-The purpose of this study was to examine changes in running performance and physical qualities related to middle-distance performance over a training season. The study also examined relationships between training loading and changes in physical qualities as assessed by laboratory and field measures. Relationships between laboratory and field measures were also analyzed. This was a 9-month observational study of 10 highly trained adolescent middle-distance athletes. Training intensity distribution was similar over the observational period, whereas accumulated and mean distance and training time and accumulated load varied monthly. Statistically significant ($p < 0.05$) and large effect sizes (Cohen's d) (0.80) were observed for improvements in: body mass (5.6%), 600-m (4.6%), 1,200-m (8.7%), and 1,800-m (6.1%) time trial performance, critical speed (7.1%), V?o(2)max (5.5%), running economy (10.1%), vertical stiffness (2.6%), reactive index (3.8%), and countermovement jump power output relative to body mass (7.9%). Improvements in 1,800 m TT performance were correlated with increases in V?o(2)max ($r = 0.810$, $p = 0.015$) and critical speed ($r = 0.918$, $p = 0.001$). Increases in V?o(2)max and critical speed were also correlated ($r = 0.895$, $p = 0.003$). Data presented here indicate that improvements in critical speed may be reflective of changes in aerobic capacity in adolescent middle-distance athletes.

ID: 130; PMID = 31809463

A.4 Technical Details

A.4.1 Session info

```
sessionInfo()
```

```
R version 4.1.2 (2021-11-01)
Platform: i386-w64-mingw32/i386 (32-bit)
Running under: Windows 10 x64 (build 22000)

Matrix products: default

locale:
[1] LC_COLLATE=German_Germany.1252  LC_CTYPE=German_Germany.1252
[3] LC_MONETARY=German_Germany.1252 LC_NUMERIC=C
[5] LC_TIME=German_Germany.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
 [1] rstudioapi_0.13  knitr_1.39       magrittr_2.0.3   munsell_0.5.0
 [5] tidyselect_1.1.2 colorspace_2.0-3 here_1.0.1       R6_2.5.1
 [9] rlang_1.0.2      fastmap_1.1.0    fansi_1.0.3      highr_0.9
[13] stringr_1.4.0    dplyr_1.0.8      tools_4.1.2      xfun_0.30
[17] utf8_1.2.2       cli_3.3.0        htmltools_0.5.2  ellipsis_0.3.2
[21] yaml_2.3.5       digest_0.6.29    rprojroot_2.0.3  tibble_3.1.6
[25] lifecycle_1.0.1  crayon_1.5.1     purrr_0.3.4      vctrs_0.4.1
[29] glue_1.6.2       evaluate_0.15    rmarkdown_2.14   stringi_1.7.6
[33] compiler_4.1.2   pillar_1.7.0     scales_1.2.0     generics_0.1.2
[37] jsonlite_1.8.0   pkgconfig_2.0.3
```

A.4.2 Used Packages

```
p_used <- unique(renv::dependencies(path = "../")$Package)
```

```
Finding R package dependencies ... Done!
```

```
p_inst <- as.data.frame(installed.packages())
out <- p_inst[p_inst$Package %in% p_used, c("Package", "Version")]
rownames(out) <- NULL
out
```

	Package	Version
1	colorspace	2.0-3
2	dplyr	1.0.8
3	ggplot2	3.3.5
4	ggtext	0.1.1
5	here	1.0.1
6	knitr	1.39
7	MetBrewer	0.2.0
8	purrr	0.3.4
9	readxl	1.4.0
10	rentrez	1.2.3
11	renv	0.15.4
12	rmarkdown	2.14
13	scales	1.2.0
14	shiny	1.7.1
15	spiro	0.0.3.9000
16	tidyr	1.2.0
17	treemapify	2.5.5
18	XML	3.99-0.9
19	grid	4.1.2