**Popularity and prevalence of gas exchange data processing**

**methods in peer-reviewed literature: a scoping review**

**Introduction:** Cardiopulmonary exercise testing (CPET) is the gold standard for non-invasively measuring cardiorespiratory fitness and is popular in many fields. CPET analyzes expired gases, usually generating one data point per breath (breath-by-breath). Consequentially, this data requires processing before later analysis because respiratory variability is much higher than the underlying metabolism it reflects. There is no universal method to process breath-by-breath data, and previous research shows that different approaches can affect downstream results. Moreover, there is little research on the popularity of different data processing methods and how often they are described in peer-reviewed literature. This scoping review aims to document the popularity and prevalence of CPET data processing methods.

**Methods:** We followed current scoping review methodology and searched three databases for papers whose methods included CPET with gas exchange. We screened those results using regular expressions and machine learning to identify papers where authors collected gas data breath-by-breath. Using regular expressions, we documented which, if any, gas data processing procedures were described in the methods section regarding outlier removal, interpolation, and averaging methods.

**Results:** Of the 7119 articles analyzed, 328 (4.6%) described outlier removal, 470 (6.6%) described interpolation, and an estimated 4366 (61.3%) described averaging methods. The most popular outlier cutoffs are mean ± 3 or 4 SD (40.9% and 50.3%, respectively). When documented, the dominating interpolation time frame and procedure were one second (94.7%) and linear interpolation (92.8%), respectively. Time-based bin averages (84.6%) were the most popular averaging methods.

**Discussion:** Few papers explicitly describe gas data processing regarding outlier removal and interpolation. The importance of these steps depends on the type of subsequent analysis, but it may be helpful to increase such documentation for reproducibility and study comparisons.