

## Supplementary material: Popularity and prevalence of gas exchange data processing methods: a semi-automated scoping review

### Supplemental Methods

#### Information Sources and Search

Scopus search:

( TITLE-ABS-KEY ( "oxygen consumption" OR "oxygen uptake" OR vo2 ) AND TITLE-ABS-KEY ( cycli\* OR bicycl\* OR run OR runn\* OR treadmill\* OR swim\* OR ski OR skie\* OR skiing OR ergometer\* OR row\* ) AND TITLE-ABS-KEY ( exer\* ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ) AND ( LIMIT-TO ( SRCTYPE , "j" ) ) AND ( LIMIT-TO ( EXACTKEYWORD , "Human" ) )

Web of Science Search

((ALL=("oxygen consumption" or "oxygen uptake" or VO2)) AND ALL=(cycli\* or bicycl\* or run or runn\* or treadmill\* or swim\* or ski or skie\* or skii or ergometer\* or row\*)) AND ALL=(exer\*) and English (Languages) and Articles (Document Types)

Ovid-MEDLINE

- ("oxygen consumption" or "oxygen uptake" or VO2).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

(cycli\* or bicycl\* or run or runn\* or treadmill\* or swim\* or ski or skie\* or skii or ergometer\* or row\*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
1. word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
2. word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
3. 1 and 2
- 3 and exer\*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
4. 3 and "Humans".sa\_suba.
5. 4 and "Journal Article".sa\_pub.
6. 5 and "Review" or "Systematic Review" or "Meta-Analysis".sa\_pub.

### *Text Analysis & Screening*

We principally employed two machine learning (ML) models using a random forest classifier through the Python sklearn package [1]. One model predicted if the article was original peer-reviewed research vs. a review, meta-analysis, case study, protocol registration, etc. The other predicted if the subjects were human or non-human.

We initially built the training data for these models while manually reading articles to construct our regular expressions (RegExs). We then used the models to predict the eligibility of the remaining unread articles. Given that our electronic search primarily obtained articles that matched our search criteria, a relatively small fraction of the unread articles were predicted to be ineligible. We manually verified the eligibility for this small number of articles. Our classifiers also calculated an eligibility probability. We also manually reviewed articles predicted eligible within ~15% of the decision boundary.

We then iteratively rebuilt our classifiers with more data from the manual eligibility verification and from other ineligible articles found incidentally while reading their full text for context when charting our data items. Through rebuilding larger models, we discovered additional ineligible articles missed by previous classifier iterations.

Both models were assessed using repeated stratified 5-fold cross-validation as our training data favored eligible articles. The original peer-reviewed research model contained 1,505 samples while the human model 1,919. The mean accuracy for the original research and human models were 83.9% and 93.4%, respectively.

### **References**

- [1] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 2011; 12: 2825–2830