# Identifying Non-Biological Variance in Untargeted Analysis in Breath VOCs


ASU School of Life Sciences
Arizona State University

Darakshan Zabin[1,2], Trenton J. Davis[2], Heather D. Bean[1,2]

[1]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

[2]Center for Fundamental and Applied Microbiomics, The Biodesign Institute, Tempe, AZ 85287, USA

ASU Biodesign Institute
Arizona State University


## Abstract


**Background:** When performing untargeted analyses for breath biomarker discovery, the goal is to identify biological signals in the breath chemistry that can help detect disease states, probe microbial interactions, and track environmental exposures. However, untargeted metabolomics and downstream data analysis workflows are vulnerable to signal contamination by non-biological variance associated with sample collection and handling and instrumental drift, which can obscure meaningful biological patterns.

**Objective/Aims:** The goal of this study is to identify and quantify non-biological variance in untargeted volatile metabolomics arising from sample collection procedures and instrumental drift.


## Methods

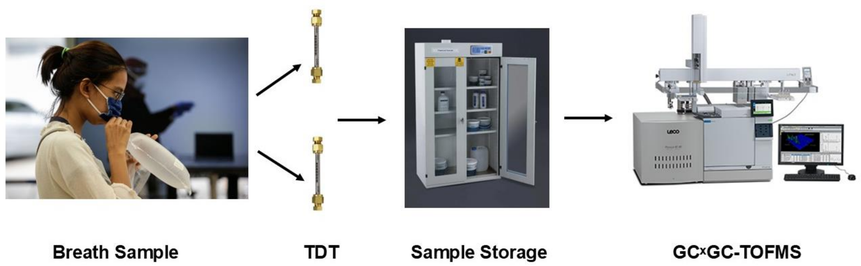


**Fig. 1: Workflow of data collection.** Breath samples were collected and transferred onto Thermal Desorption Tubes (TDTs) in technical duplicates. TDTs were stored at 4˚C and analyzed using two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC×GC–TOFMS).

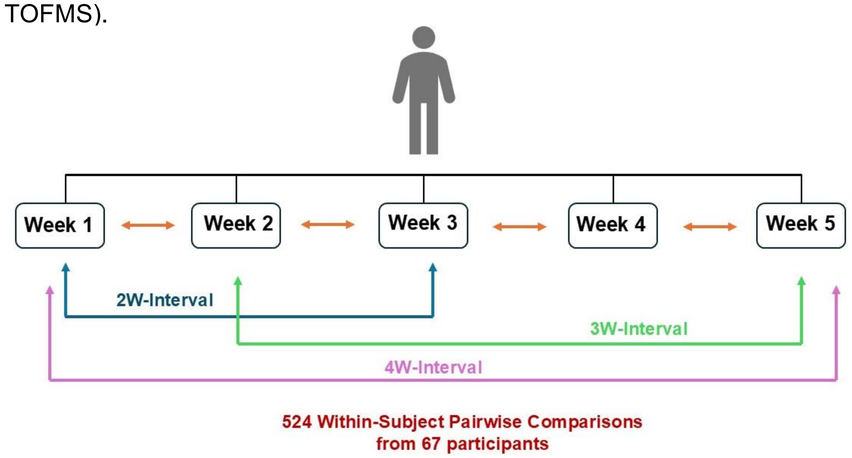


**Fig. 2: Within-subject study design.** Participants provided breath samples over 1–5 weeks, enabling within-subject comparisons across 1- to 4-week intervals to test whether breath profiles become more dissimilar as time between collections increases.

## Methods

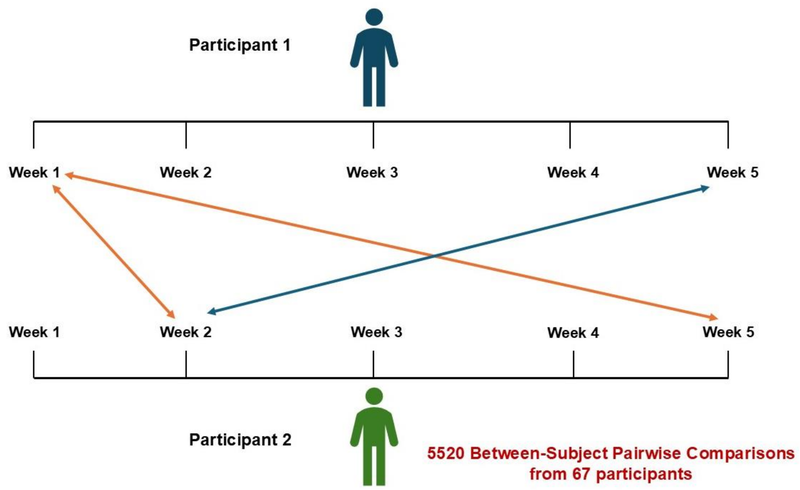


**Fig. 3: Between-subject study design.** All participants gave breath samples across five weeks. We compared each breath profile to all other participants' breath profiles to evaluate how between-subject dissimilarity changed as the collection time gap increased.

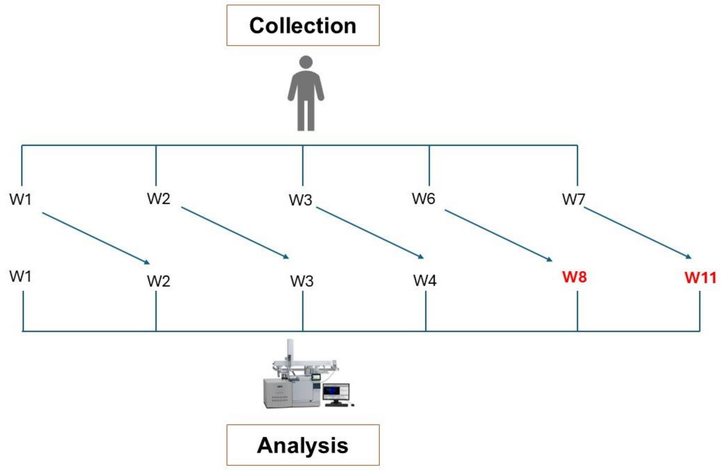


**Fig. 4:** Breath samples were typically collected and analyzed one week apart, but instrument availability created longer analysis gaps (up to 11 weeks).

## Conclusions

Instrument drift and other non-biological factors can strongly affect metabolomics data. Consistent, time-dependent patterns of dissimilarity highlight the need to identify and control these sources of variance to ensure reliable biomarker discovery.

## Results

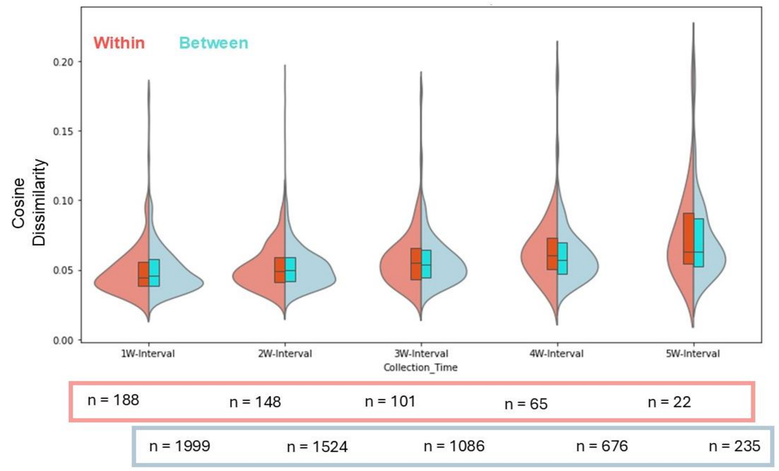


**Fig. 5: Breath dissimilarity as a function of collection time gaps.** Cosine dissimilarity increased with longer collection intervals for both within-subject and between-subject comparisons, suggesting that technical variance outweighs biological variance in this dataset.

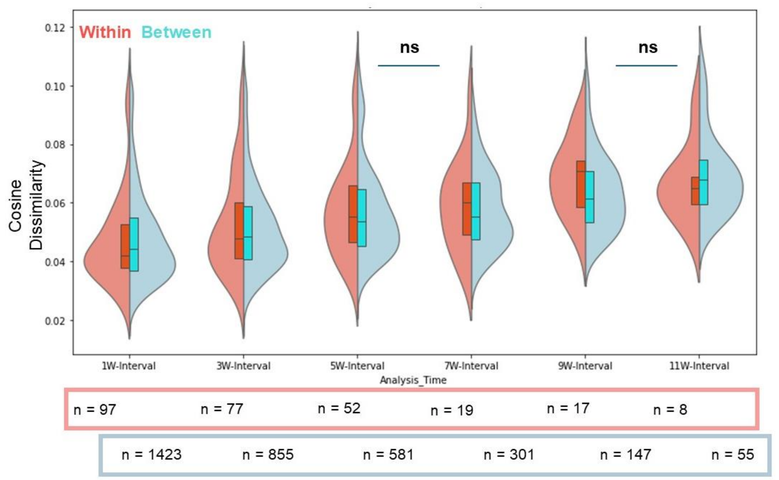


**Fig. 6: Breath dissimilarity as a function of analysis time gaps.** Breath profile dissimilarity increased as analysis time gaps grew (up to 11 weeks). The increase in both within-subject and between-subject dissimilarity across analysis time supports the conclusion that instrumental drift strongly influence observed dissimilarities (498 within-subject and 5,714 between-subject comparisons).

## Funding & Disclaimer


This work was supported by the Defense Advanced Research Projects Agency (DARPA) Fatigue Assessment via Breath (FAB) study (Cooperative Agreement HR00112220040). The views, opinions, and/or findings contained in this material are those of the authors/presenters and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. No official endorsement should be inferred.