Online supplement 4.

Internal Validation

Disease-specific VOCs or VOC patterns (breathprints) are usually identified in cross-sectional studies with internal validation [1,2,3]. Although internal validation is considered an adequate method to limit over fitting when training and validation sets are from the same population are used, it has implicit limitations in predicting how a model will fit in an independent population. Especially when this technique is used for feature selection for the model, as is often the case in VOCs studies. Examples of internal validation methods are: split-half, cross-validation, x-fold analyses, x-fold cross validation and bootstrapping [4-6].

External Validation

Definitive confirmation of disease-specific VOCs or VOC patterns, fit for inclusion in prospective diagnostic trials, can only be provided by validation in entirely independent patient cohorts. As breathomic profile (the "breathome") may potentially be influenced by genetics [7], environment, and diet, these validation cohorts should be preferably multi-centre and multi-national.

Whilst there have been many exploratory studies performed in breath analysis, often with internal cross validation (usually overestimating accuracy), there have been far fewer external validation studies, due to the need for multi-centre involvement and the attendant increase in financial and administrative requirements.

External validation has been provided in single centre and multi-centre settings [8, 9-13]. This has shown external validity of breath profiles that discriminated asthma (including those with fixed airflow obstruction) from COPD by using a validation dataset from a different hospital to that used for the model generation [8]. Likewise, a discriminatory model for COPD vs. controls based on 6 VOCs (GC-MS) was externally validated with patients from a different hospital [9]. In acute respiratory distress syndrome (ARDS) GC-MS and e-Nose data have been successfully externally validated at a single site [11,12]. Recently published data

showing than diagnostic accuracy cannot be estimated precisely from the training set, not even after internal validation, external validation is recommended for such studies [13].

References

1. Bikov A, Lázár Z, Horvath I. Established methodological issues in electronic nose research: how far are we from using these instruments in clinical settings of breath analysis? J Breath Res 2015: 9(3):034001.

2. van de Kant KD, van der Sande LJ, Jobsis Q, van Schayck OC, Dompeling E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. Respir Res 2012: 13: 117.

3. Fens N, van der Schee MP, Brinkman P, Sterk PJ. Exhaled breath analysis by electronic nose in airways disease. Established issues and key questions. Clin Exp Allergy 2013: 43(7): 705-715.

4. Sung J, Wang Y, Chandrasekaran S, Witten DM, Price ND. Molecular signatures from omics data: from chaos to consensus. Biotechnol J 2012: 7(8): 946-957.

5. Broadhurst D, Kell D. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2006: 2(4): 171-196.

6. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, Mesirov JP, Polley MYC, Kim KY, Tricoli JV, Taylor JMG, Shuman DJ, Simon RM, Doroshow JH, Conley BA. Criteria for the use of omics-based predictors in clinical trials: Explanation and elaboration. BMC Medicine 2013: 11(220): 1-15.

7. Tarnoki DL, Bikov A, Tarnoki AD, Lazar Z, Szilagyi BK, Korosi BZ, Horvath T, Littvay L, Losonczy G, Horvath I. Lack of heritability of exhaled volatile compound pattern: an electronic nose twin study. J Breath Res 2014;8(1):016001.

8. Fens N, Roldaan AC, van der Schee MP, Boksem RJ, Zwinderman AH, Bel EH, Sterk PJ. External validation of exhaled breath profiling using an electronic nose in the discrimination of asthma with fixed airways obstruction and chronic obstructive pulmonary disease. Clin Exp Allergy 2011: 41(10): 1371-1378.

9. Van Berkel JJ, Dallinga JW, Moller GM, Godschalk RW, Moonen EJ, Wouters EF, Van Schooten FJ. A profile of volatile organic compounds in breath discriminates COPD patients from controls. Respir Med 2010: 104(4): 557-563.

10. Bikov A, Pako J, Kovacs D, Tamasi L, Lazar Z, Rigo J, Losonczy G, Horvath I. Exhaled breath volatile alterations in pregnancy assessed with electronic nose. Biomarkers 2011: 16(6): 476-484.

11. Bos LD, Weda H, Wang Y, Knobel HH, Nijsen TM, Vink TJ, Zwinderman AH, Sterk PJ, Schultz MJ. Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome. Eur Respir J 2014: 44(1): 188-197.

12. Bos LD, Schultz MJ, Sterk PJ. Exhaled breath profiling for diagnosing acute respiratory distress syndrome. BMC Pulm Med 2014: 14: 72.

13. Leopold JH, Bos LD, Sterk PJ, Schultz MJ, Fens N, Horvath I, Bikov A, Montuschi P, Di Natale C, Yates DH, Abu-Hanna A. Comparison of classification methods in breath analysis by electronic nose. J Breath Res 2015; 9 (4): 046002.