Detection of lung cancer using weighted digital analysis of breath biomarkers

Michael Phillips,⁎, Nasser Altorki, John H.M. Austin, Robert B. Cameron, Renee N. Cataneo, Robert Kloss, Roger A. Maxfield, Muhammad I. Munawar, Harvey I. Pass, Asif Rashid, William N. Rom, Peter Schmitt, James Wai

⁎ Corresponding author. Menssana Research, Inc., 1 Horizon Road, Suite 1415, Fort Lee, NJ 07024, United States. E-mail address: mphillips@menssanaresearch.com (M. Phillips).
can be readily estimated e.g., hypoglycemia can be distinguished from normoglycemia by measuring the blood glucose concentration, and then determining whether it is above or below a designated cutoff value. However, clinical diagnosis is usually more difficult because assignment to one group or the other generally requires a combination of several different criteria. For example, a patient with acute streptococcal pneumonia may present with several different symptoms and signs, including fever, chills, productive cough, herpes labialis, vocal fremitus, and localized crackles in the chest. As an additional complication, not all of these features are required for the diagnosis, and some are more important than others. An experienced physician takes these difficulties into account by intuitively assigning a different diagnostic weight to each finding. For example, fever has a comparatively low diagnostic weight because it may be absent in elderly or immune suppressed patients, but the diagnostic weight of chest crackles is much higher. This process of intuitively weighing all of these relative values and then incorporating them into a binary prediction is often described as “diagnostic clinical judgment”.

Since the exercise of diagnostic clinical judgment is an intuitive process, its outcome must necessarily vary with an individual’s skill and experience. As a consequence, physicians have sought to improve the accuracy and reproducibility of clinical judgment by employing formal algorithms to assign patients to appropriate diagnostic groups. As early as 1944, Jones proposed an algorithm for the diagnosis of rheumatic fever that employed a combination of “high weight” major criteria (including carditis and polyarthritis) and “low weight” minor criteria (including fever and arthralgia). As evidence of its clinical value, a modified version of this algorithm is still in clinical use >60 y later [1]. The accuracy and reproducibility of diagnostic laboratory tests can also be improved in the same fashion. A predictive algorithm employing the relative diagnostic weights of two or more biomarkers of lung cancer in combination can predict disease with greater accuracy than a single biomarker employed alone [2].

Non-linear multivariate statistical analysis provides a useful tool for determining the relative weights of clinical markers of disease and incorporating them into new diagnostic algorithms. We have previously reported that biomarkers in the breath predict lung cancer [3,4], breast cancer [5], pulmonary tuberculosis [6], and heart transplant rejection [7]. All of these tests assigned a relative weight to a number of different biomarkers – volatile organic compounds (VOCs) in the breath – and incorporated them into a predictive algorithm with a binary outcome i.e. disease or no disease. A non-linear multivariate model employing fuzzy logic predicted lung cancer with greater accuracy than multilinear analysis [8].

We report here a new method for non-linear modeling. Weighted digital analysis (WDA) determines the relative weights of a set of VOCs as biomarkers of disease, and incorporates them into an algorithm to predict the presence or absence of disease. We present evidence for
the effectiveness of WDA in a reanalysis of data obtained from a previous study of breath VOC biomarkers of lung cancer.

2. Methods

2.1. Clinical study

This has been previously reported [8]. In summary, breath samples were collected from 404 patients: 193 with untreated primary lung cancer and 211 controls with no evidence of cancer on chest CT. VOCs in alveolar breath and ambient air were analyzed by gas chromatography and mass spectrometry. The data set comprised the breath VOCs in patients with untreated primary lung cancer and the controls.

2.2. Determination of alveolar gradients

The alveolar gradient is the difference between the abundance of a VOC in breath and air, and the method for its determination has been described [9].

2.3. General principles of WDA

WDA is a mathematical method for developing a diagnostic algorithm that generates a discriminatory function score. The value of this score predicts membership in 1 of 2 groups e.g., disease or no disease. Every diagnostic variable (e.g., the numerical value of a laboratory test result) that is employed in the algorithm has three parameters.

a. Sign. This may be positive (+1) or negative (-1). If the sign is positive, a higher value of the diagnostic variable indicates that disease is more likely. Conversely, if its sign is negative, a lower value of the variable indicates that disease is more likely.

b. Weight. This value indicates the relative contribution of each diagnostic variable to the discriminatory function, i.e., the higher its weight, the greater is its relative importance as a predictor of disease.

c. Cutoff value. This determines whether or not a diagnostic variable contributes to the discriminatory function score. The contribution of a diagnostic variable (weight) is added to the discriminatory function score only if that value exceeds the cutoff value.

For each variable, the value of (gradient X sign) is determined, and if it exceeds the cutoff value, the variable is included as a diagnostic variable in the discriminatory function score. The discriminatory function score (df) for a given patient (subscript i) is calculated as:

\[ df_i = \sum y_{kc} \text{ with } y_{kc} = \begin{cases} \text{Weight} & \text{if } x_{ki} \cdot \text{Sign}_i > \text{Cutoff}_i \cdot \text{Sign}_i \\ 0 & \text{otherwise} \end{cases} \]

The subscript c incorporates all contributing VOCs. The alveolar gradient of each VOC was employed to create a receiver operating characteristic (ROC) curve, demonstrating its ability to distinguish cancer patients from controls. As an example, the ROC curve for a breath VOC tentatively identified as isopropyl alcohol is shown in Fig. 1. From this ROC curve, the area under curve (AUC), sign (if AUC is < 0.5, +1 if AUC is > 0.5), and cutoff value C of the alveolar gradient where [sensitivity+(1–specificity)] is a maximum were determined. This enables the determination of the cutoff (C), weight = (AUC–0.5)/0.5, and sign = +1 if AUC is > 0.5; +1 if AUC is < 0.5. The WDA algorithm may be readily employed in a computerized spreadsheet program. It is currently implemented as a macro in Microsoft Excel.

2.4. Selection of VOCs for inclusion in WDA

ROC curves were determined in this fashion for all breath VOCs, and only those VOCs with weight >0.6 were selected for further analysis. Breath VOCs were edited to combine duplicates i.e., ROC curves were constructed from the discriminatory function generated by employing all of the predictor variables in the model.

2.5. Robustness of the model

We defined robustness as the number of breath VOCs that may be lost from the model without incurring a significant deterioration in its predictive accuracy. In this context, a "lost" VOC is either absent or else it is generating random results. We

Table 1

<table>
<thead>
<tr>
<th>Breath VOC</th>
<th>CAS #</th>
<th>NIST #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isopropyl alcohol</td>
<td>67-63-0</td>
<td>229015</td>
</tr>
<tr>
<td>4-Penten-2-ol</td>
<td>625-31-0</td>
<td>235565</td>
</tr>
<tr>
<td>Ethane, 1,1,2-trichloro-1,2,2-trifluoro-</td>
<td>76-13-1</td>
<td>233813</td>
</tr>
<tr>
<td>Propane, 2-methoxy-2-methyl-</td>
<td>1634-04-4</td>
<td>229277</td>
</tr>
<tr>
<td>1-Propene, 1-(methyliethyo)-, (E)-</td>
<td>42848-06-6</td>
<td>26402</td>
</tr>
<tr>
<td>2,3-Hexanediol</td>
<td>3848-24-6</td>
<td>291466</td>
</tr>
<tr>
<td>5,5-Dimethyl-1,3-hexadiene</td>
<td>1515-79-3</td>
<td>113453</td>
</tr>
<tr>
<td>3-Hexanone, 2-methyl-</td>
<td>7379-12-6</td>
<td>231728</td>
</tr>
<tr>
<td>1H-Indene, 2,3-dihydro-4-methyl-</td>
<td>824-22-6</td>
<td>2991</td>
</tr>
<tr>
<td>Camphor</td>
<td>21368-68-3</td>
<td>7361</td>
</tr>
<tr>
<td>Bicyclo[2.2.1]heptan-2-one, 1,7,7-trimethyl-,(1S)-</td>
<td>464-48-2</td>
<td>114698</td>
</tr>
<tr>
<td>3-Cyclohexene-1-methanol, aâ`-trimethyl-</td>
<td>98-55-5</td>
<td>231634</td>
</tr>
<tr>
<td>n-Pentanol-1-en-8-ol</td>
<td>N/A</td>
<td>151924</td>
</tr>
<tr>
<td>5-Isopropenyl-2-methyl-7-oxacyclo[4.1.0]heptan-2-ol</td>
<td>N/A</td>
<td>185009</td>
</tr>
<tr>
<td>â¨ Isopropyl isonone</td>
<td>127-51-5</td>
<td>196736</td>
</tr>
<tr>
<td>2,2,7,7-Tetramethyltricyclo[6.2.1.(1,6)]undec-4-en-3-one</td>
<td>N/A</td>
<td>189499</td>
</tr>
<tr>
<td>2,2,4,Trimethyl-1,3-pentanediol disobutyrate</td>
<td>6846-50-0</td>
<td>151177</td>
</tr>
<tr>
<td>Benzoc acid, 4-ethylxy, ethyl ester</td>
<td>23766-09-7</td>
<td>107721</td>
</tr>
<tr>
<td>Bicyclo[3.2.2]jonene-1,5-bicycloxylic acid, 5-ethyl ester</td>
<td>24238-73-1</td>
<td>260408</td>
</tr>
<tr>
<td>Pentaoc acid, 2,2,4-trimethyl-3-carboxyisopropyl, isobutyl ester</td>
<td>N/A</td>
<td>140775</td>
</tr>
<tr>
<td>Propanoic acid, 2-methyl-, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester</td>
<td>74381-40-1</td>
<td>59556</td>
</tr>
<tr>
<td>1,2,4,5-Tetroxane, 3,3,6,6-tetraphenyl-</td>
<td>16204-36-7</td>
<td>11836</td>
</tr>
<tr>
<td>Benzophenone</td>
<td>119-61-9</td>
<td>118652</td>
</tr>
<tr>
<td>2,5-Cyclohexadien-1-one, 2,6-bis,(1,1-dimethylethyl)-4-ethylidene</td>
<td>6738-27-8</td>
<td>215417</td>
</tr>
<tr>
<td>Furan, 2-[2-ethoxy-3,4-dimethyl-2-cyclohexen-1-ylidene(methyl)]</td>
<td>551-62-9</td>
<td>47619</td>
</tr>
<tr>
<td>Benzcen, 1,1-(1,2-cyclobutane)yl)bis,-cis-</td>
<td>7694-30-6</td>
<td>62825</td>
</tr>
<tr>
<td>Benzene, 1,1-[1-ethylthio]propyldiene]bis</td>
<td>53699-80-2</td>
<td>149972</td>
</tr>
<tr>
<td>Anthracene, 1,2,3,4-tetrachydro-9-propyl</td>
<td>101580-33-0</td>
<td>155542</td>
</tr>
<tr>
<td>9,10-Anthracenediol, 2-ethyl-</td>
<td>839-73-6</td>
<td>153923</td>
</tr>
<tr>
<td>Benzene, 1,1-ethylidenbis[4-ethyl]</td>
<td>10224-91-6</td>
<td>11431</td>
</tr>
</tbody>
</table>

WDA selected these 30 breath VOCs as candidate biomarkers of primary lung cancer because the AUC of each ROC curve exceeded 0.6. VOCs are ranked by their chromatographic retention times. CAS and NIST numbers are shown, where available. N/A = not available. In a previous report [6], fuzzy logic selected a different set of lung cancer biomarkers from the same data set, possibly reflecting that fuzzy logic and WDA are fundamentally different techniques of multivariate analysis. Two sets of duplicate VOCs were identified in this list (10, 11 and 12, 13, highlighted). This was a consequence of the mass spectra library assigning different synonyms (with different CAS and NIST numbers) to the same VOC on different occasions. This did not affect the outcome of multivariate analysis with WDA.
employed the following algorithm to evaluate robustness: select the candidate VOCs in
the model, rank these VOCs by AUC of the ROC curve:

1. Determine their cumulative AUC.
2. Randomly remove one VOC at a time until the cumulative AUC falls below 90% of
the original value.
3. Repeat this step n (e.g., 30) times.
4. The average number of removed VOCs is termed the robustness for 10% degradation.

2.6. Detection of lung cancer

We employed WDA to analyze alveolar gradients of breath VOCs in the entire data
set of 404 subjects (untreated primary lung cancer and cancer-free controls) employing
the method described above. In addition, we cross-validated algorithms in randomly
split subset groups [10]. Subjects were randomly assigned to a training set or to a
prediction set in a 2:1 ratio. WDA was performed employing multiple (n=20) randomly
selected unique training and prediction sets.

3. Results

3.1. Candidate biomarkers of lung cancer

Fig. 1 displays the accuracy of a typical single breath VOC employed
as a biomarker of lung cancer. Note that this VOC alone did not
significantly distinguish between cancer patients and the control
group; it required the combined results of several VOCs to generate a
discriminant function with a significant AUC value. Fig. 2 displays
the effect of random patient assignment on predictive accuracy of
individual breath VOCs, and how these results were employed to
identify the optimal candidate biomarkers of lung cancer, with AUC of
ROC curve >0.6. This cutoff value was employed because no VOC had
an AUC >0.6 when the diagnosis was randomly assigned (Table 1).

3.2. WDA discriminatory function in all cases of lung cancer

Fig. 3 displays WDA discriminatory function values in controls
and lung cancer patients. The area under curve (AUC) of the resulting
receiver operating characteristic (ROC) curve was 0.90. Employing
multilinear regression of the same data set, AUC of ROC curve= 0.74.

3.3. WDA discriminatory function in lung cancer stratified by TNM
(tumor, node, metastasis) stage

Fig. 3 displays mean discriminatory function values in controls
and lung cancer patients stratified by TNM stages 1 to 4. Fig. 4 displays the
ROC curves obtained from these data. Test accuracy did not vary appreciably with TNM stage of disease.

3.4. WDA discriminatory function in lung cancer stratified by tobacco
smoking

Fig. 5 displays the ROC curves obtained when WDA data in Fig. 3
was stratified according to whether subjects were current smokers
(AUC=0.92) or former smokers (AUC=0.90). By inspection, the simi-
larity of the two ROC curves demonstrated that the WDA discrimi-
natory function was not affected by whether subjects were current
smokers or ex-smokers.

3.5. Effect of the number of VOCs in model on WDA discriminatory
function

Fig. 6 displays the effect of the number of VOCs employed in the
model on the AUC of the ROC curve for all patients with lung cancer. VOC
biomarkers of lung cancer shown in Table 1 were added to the model
one by one, commencing with the highest weight VOC. The WDA
discriminatory function required only ten VOCs in order to identify lung
cancer with near maximal accuracy. However, more VOCs were added to
the algorithm in order to enhance the robustness of the analysis.

3.6. Cross validation in random split subsets

Fig. 7 displays the training and validation ROC curves (mean of 20
random splits). The same 30 VOCs employed in the final model were
also employed in these split data sets; however the cutoff points,
signs, and weight, were adjusted for each split based on the results
in their respective training sets.

3.7. Robustness of the model

Fig. 8 displays variation in the robustness with the number of VOCs
in the model, as well as “robustness ~ 3 Sigma”, which indicates the
number of VOCs that can be lost with 99.7% probability the AUC will
not degrade by >10%. If, for example, the model employed 30 VOCs, a
third of these VOCs could be lost without reducing the accuracy of the
breath test by >10% at 99.7% CI.
3.8. Effect of random assignment of diagnosis

Fig. 7 (lower panel) displays the effect of random reversal of assignment of patient diagnosis, prior to determination of the WDA discriminatory function scores. The accuracy of the WDA model progressively deteriorated with the declining integrity of the breath VOC data, supporting the conclusion that the undegraded WDA model identified lung cancer by extracting a signal of disease from the breath VOC data.

4. Discussion

A test employing WDA of a combination of breath VOCs accurately identified patients with lung cancer. The accuracy of the breath test may be directly compared to that of chest CT. In a large population screening study, chest CT detected lung cancer with 55% sensitivity and 95% specificity [11]. As the ROC curve in Fig. 4 demonstrates, at the point where the breath test sensitivity was 55%, its specificity was approximately 93% i.e. close to the same as chest CT.

Table 1 lists the breath VOCs identified as candidate biomarkers of primary lung cancer. Although we observed a strong statistical association between lung cancer and a set of apparent VOC biomarkers in the breath, the biological mechanism linking lung cancer with these breath VOCs has not yet been identified with certainty. Tumor markers are conventionally regarded as downstream products that are manufactured in cancer cells and discharged into the blood. Examples include CA125 in ovarian cancer, PSA in prostate cancer and CEA in ovarian and breast cancers. However, we have previously described four important points of difference between typical downstream tumor markers and breath VOC biomarkers of lung cancer [8].

4.1. Biological significance and variation

Few of the breath VOCs associated with lung cancer have known biological significance in lung disease. Also, the set of breath VOCs associated with lung cancer has been found to vary from study to study, and also within an individual study when different techniques of multivariate analysis were employed.

4.2. Effect of tumor mass

Serum levels of a tumor marker may increase as a tumor grows larger [13]. However, tumor mass did not affect the abundance of breath VOC biomarkers of lung cancer; these remained relatively constant, as shown by the similarity of ROC curves in TNM stages 1 through 4 (Fig. 4).

Fig. 4. Breath biomarkers of lung cancer stratified by TNM stage: These figures display the ROC curves obtained by stratifying the WDA data in Fig. 4 according to the TNM stage of lung cancer. The AUC was high in TNM1 lung cancer, and a similar performance was maintained at all other stages. (Since the overall AUC of the total set is high (around 0.9) it is to be expected that the AUC of any of the subsets stratified by TNM stage will have a similarly high value.)
4.3. Effect of surgery

Ablation of the prostate reduces serum levels of PSA [14], and concentrations of serum tumor marker are generally reduced by excision of the cancer. However, we previously reported that the outcome of the breath test was unchanged in most patients with lung cancer following thoracotomy with resection of the tumor.

4.4. Abundance

Serum tumor markers are consistently increased in patients with cancer; however, we observed a combination of decreased as well as increased abundance of breath VOC biomarkers in lung cancer.

For these reasons, we concluded that the downstream model of tumor marker production did not provide a satisfactory explanation of the observed breath VOC biomarkers of lung cancer. We therefore proposed an alternative biological mechanism: an upstream model hypothesis, in which the pathophysiologic process that results in lung cancer may also modulate the abundance of VOCs in breath, so that carcinogenesis and altered breath VOCs are 2 concurrent but independent phenomena. Fig. 8 displays a pathophysiologic hypothesis in which activation of lethal cytochrome p450 mixed oxidases may lead to lung cancer while independently altering the catabolism of VOCs. This model provides a rational explanation for the observed points of difference described above:

4.5. Biological significance and variation

More than 3000 different VOCs have been observed in normal human breath [9], all of them with low molecular weights (<600), unlike protein serum tumor markers which have molecular weights of several kilodaltons [15,16]. Induced cytochrome p450...
mixed oxidase activity could potentially modulate the catabolism of many of these breath VOCs, and thereby account for the large and diverse sets of candidate breath biomarkers associated with lung cancer. Cytochrome p450 enzymes catabolize most of the VOCs listed in Table 1, including isopropy alcohol [17], hexanedione [18], camphor [19], benzophenone [20] and derivatives of tetroxane [21], benzene [22], benzoic acid [23], furan [24] and ionone [25] (this list is not exhaustive). The resulting diversity of candidate biomarkers constitutes a major strength of breath VOC analysis, because the accuracy of detection fell with the declining integrity of the signal.

4.6. Effects of tumor mass and surgery

The model shown in Fig. 8 predicts that tumor mass will have no effect on breath biomarkers of lung cancer, since induction of cytochrome p450 activity would alter the composition of breath VOCs independently of the growth of lung cancer cells. Similarly, resection of the lung cancer should have no effect on the breath signal.

4.7. Effects on breath VOC abundance

This model predicts a combination of decreased as well as increased abundance of breath VOC, since VOC precursors will be depleted by catabolism, while their metabolites will be increased.

Table 2 lists the advantages of WDA compared to traditional multilinear analysis, and newer methods of multivariate data analysis such as fuzzy logic, pattern recognition, or neural networks. WDA identified a set of VOC biomarkers of lung cancer that were similar, though not identical, to those identified in a previous study employing fuzzy logic analysis of data [8]. Other results were also similar to those previously reported [8]. The accuracy of the breath test was not affected by the TNM stage of lung cancer, nor by current or former tobacco smoking.

There are risks as well as benefits to adding more variables to a multivariate predictive algorithm. While the main benefit is improved accuracy of the algorithm, the risk is that some of this improvement may be illusory, because the inclusion of variables with poorer correlations can degrade predictive accuracy. Consequently, the number of variables in an algorithm is usually a compromise between two conflicting demands: there must be a sufficient number to ensure accuracy, yet not so many as to introduce spurious results. We identified 69 candidate VOC biomarkers of lung cancer in the breath (Fig. 2), and then ranked them according to the AUC of their ROC curves. We selected the top 30 VOCs for inclusion in the predictive algorithm based solely on their AUC values, without any knowledge or consideration of their potential biological significance. This approach yielded three main advantages: first, the algorithm was highly sensitive and specific for lung cancer – the AUC of the ROC curve was close to 0.9 (Fig. 6, upper panel). Second, there was no evidence of spurious results caused by poorly correlated variables because the cross-validated ROC curve exhibited virtually no degradation (Fig. 6, lower panel). Third, the WDA model was highly robust. When one third of the VOCs were randomly removed from a 30 VOC model, the accuracy of the breath test was degraded by less than 10% at the 99.7% confidence level (Fig. 7, upper panel). This provides evidence for high redundancy in breath testing for lung cancer: Even when part of the information in the breath VOC signal was unavailable, the WDA model delivered highly accurate information about the presence or absence of disease.

The weight of each predictor variable was the area under the curve (AUC) of its receiver operating characteristic (ROC) curve. We employed only VOCs with AUC > 0.6, and subtracted a fixed offset of 0.55 from these values in order to increase the relative differences between them. For example, if VOC1 AUC = 0.6 and VOC2 AUC = 0.65, then VOC2 has approximately 10% higher weight without subtraction. However, with subtraction of 0.55, the relative AUC of VOC2 to VOC1 is doubled. Most of the AUCs in this data set fell between 0.6 and 0.7, so that their relative differences were markedly increased by this procedure.

Lung cancer causes more deaths than any other malignancy in the U.S. [26]. Since annual screening with chest computed tomography (CT) can detect early stage lung cancer that is likely to be curable [27], there is hope that early detection could increase 5-y survival. However, chest CT screening is costly and the hazards of associated radiation may outweigh its potential benefits [28]. These concerns have led researchers to seek biomarkers of lung cancer that could provide...
an early warning of potentially curable disease. However, evidence from randomized trials is not yet available concerning morbidity and mortality following early detection of lung cancer. Screening for lung cancer with the breath test employing the WDA algorithm reported in this study could provide an early warning test that is safe, accurate, non-invasive, rapid, simple, and inexpensive.

WDA appears to provide a useful new technique for non-linear multivariate analysis of data. In this study, the algorithm identified dependencies beyond traditional linear approaches. WDA is a completely digital approach in the sense that it employs hard cutoff values. In future studies, multivariate modeling of clinical data could potentially produce superior results by employing a combination of linear and digital methods. A natural extension of WDA would be to investigate combinations of linear and non-linear discriminatory functions and evaluate their applicability. We conclude that WDA of breath VOCs provided a rational and accurate predictor of primary lung cancer. Because this test identifies persons at high probability for having lung cancer, it is tempting to speculate that those individuals would especially benefit from subsequent chest CT screening.

Acknowledgements
This research was supported by SBIR award 5R44HL070411-03 from the National Heart Lung and Blood Institute of the National Institutes of Health. William N Rom was supported by Early Detection Research Network grant U01 CA086137. Michael Phillips is President and CEO of Menssana Research, Inc. All coauthors had full access to all of the data in the study and they take responsibility for the integrity of the data and the accuracy of the data analysis. We thank James J. Grady DrPH for reviewing the statistical analysis of data.

References


