

# Experimental Evaluation of Gas Sensors Array for the Identification of Complex VOC Mixtures in Human Breath

Justin D.M. Martin\*, Anne-Claude Romain

Sensing of Atmospheres and Monitoring (SAM), University of Liège (ULiège) Arlon Campus Environnement, Avenue de Longwy 185, 6700 ARLON, Belgium.  
[Jdm.martin@uliege.be](mailto:Jdm.martin@uliege.be)

According to literature, a Gas Sensor Array (GSA) can identify lung cancer by sensing the volatile organic compounds (VOC) in the breath of patients. The aim of this paper is to report a first approach of experimental design to characterise GSAs' metrological characteristics, and compare the results with classical calibration. The VOCs used for both calibration and experimental design are selected on the basis of a literature review. The effect of humidity is considered, and the results between both methods are compared. This study was conducted as a part of the European PATHACOV project. The project's approach is to develop a prototype GSA with new type of sensors (designed by 4 partners) for identifying the footprint of breath from patients diagnosed with lung cancer for early screening purpose.

*Index terms – Experimental design, synthetic gas mixtures, gas sensor array, benchmark, volatile organic compound, metal oxide semiconductor.*

## 1. Introduction

In Belgium and everywhere else in Europe, lung cancer was the leading cause of death for men (25% of cancer fatalities, about 267 000 Europeans) and the second leading cause of death for women behind breast cancer in 2018 (15% of cancer fatalities, about 121 000 Europeans). Lung cancer is the most common form of cancer (Ferlay et al., 2018), and is responsible worldwide for more than 1,76 million fatalities in 2018 according to the WHO. The death toll of lung cancer is on the rise each year (Ferlay et al., 2015; Plummer et al., 2016). Most lung cancer diagnosis methods use high-end, slow, stationary, expensive devices with the need of trained personnel to operate them. It is not conceivable to use these methods (Pet-scan, IRM) for early lung cancer diagnostics in a wide asymptomatic population. This poses a significant problem since early diagnosis improve tremendously the survival rate: 80% of early identified lung cancer patients survive the next 5 years after the diagnosis, against 15% only if the lung cancer is found in its late stages (National Lung Screening Trial Research Team et al., 2011). Most patients do not consult a doctor in medicine unless they experience symptoms such as chest pain, chronic cough, weight loss, which mostly appear in later stages of the illness (Silvestri et al., 2016). Therefore, there is a rising need for a simple, portable, inexpensive, non-invasive new method to help in the screening of a large number of people. There's been a number of attempts to tackle the problem in the last decades, and one promising method resides in the analysis of volatile organic compounds (VOC) naturally emitted by the human body. By analysing the VOCs emitted by blood, urine, or in breath samples, several studies managed to identify characteristic patterns for various diseases (asthma, atherosclerosis, cirrhosis, diabetes, bacterial infections) including several forms of cancer (lung, mesothelioma, liver, neck, breast, stomach, colorectal...) (Righettoni et al., 2015). One promising method chosen by the Pathacov project (*Pathacov project*, 2020) supporting this paper's work is breath VOC pattern identification by gas sensor array. The prototype built for the project has to be extensively tested before field use, and this paper is in this line of work. In this paper, a comparison is made between a Design of Experiment (DoE) approach and a classical four-points calibration approach.

The goal is to find the optimal way to obtain the maximum amount of reliable information regarding the metal oxide semiconductor (MOS) sensors in as few operations as possible. Four VOCs and humidity in varying amounts are used to prepare synthetic gas mixes that are submitted to a homemade e-nose system, using both methods.

## 2. Material and methods

### 2.1 Gas sensor array

The tested Gas Sensor Array (GSA) was built using a “Teensy 3.5” board and six commercial metal oxide semiconductor (MOS) sensors. The sensors are inserted radially and evenly into a small size cylindrical PTFE chamber (external volume 47.1 mL, internal dead volume 7.5mL). The selected sensor models were the TGS T2603 (Figaro Engineering®), GGS G3530, G1430, G2530, G8530 (Umwelt Sensor Technik®), MP901 (Winsen®) and BME680 (Bosh®). This last sensor is a thin-film MOS digital sensor that also includes temperature and humidity measurements. The conductance of each sensor, the temperature and humidity are recorded on a microSD flash memory card and on a computer every second.

Commercial sensors were chosen as a first approach to test our benchmark method. The following step is the testing of the new chemical sensors developed by the Pathacov project partners. Therefore, the benchmark should give information regarding each sensor’s characteristics and the performance of the whole array for discrimination. This enables comparison of sensors between them, which is important for iterative improvement of both individual sensors and the array as a whole.

The GSA is equipped with a carbon dioxide sensor (GGS SprintIR®) for assessing the influence of CO<sub>2</sub> to the MOS sensors signal and for sample tracking. The sensor is also usable for capnography and breath fraction selection. The final method for collecting samples will use an automatic sampler, which uses capnography to sample the last fraction of the exhalation. The last fraction of the exhalation carries less contamination from the upper airways (mouth, oesophagus), and is preferably sampled.

The whole device (array, tubing, samples...) is kept at 40°C in an oven to prevent condensation in the tubings. MOS sensors are sensitive to humidity, and condensation can modify the concentration of hydrophilic compounds in the gas phase, which is of course to be avoided.

A simple electric pump and a rotameter are used to suck air out of the sensor chamber, the flow being regulated at 0.5L/min. The breath is stored at 40°C in a low volume gas sampling bag, and connected shortly after to the GSA’s intake. This way, contamination from the upper airways is moderated and flow is made constant in the sensor chamber. Constant flow is important to ensure quick and repeatable signal stabilisation.

### 2.2 Parameters and variables selection

Relevant testing of the characteristics of gas sensors requires the selection of parameters and variables from real field use conditions. The environmental conditions of the usage of the device can be known in advance (e.g. temperature, gas flow...), which is not true for the characteristics of the breath. Therefore, variables to be studied are chosen within the breath main characteristics: nature and concentration of VOCs, relative humidity level, CO<sub>2</sub> content. If the last two variables can be known relatively easily, the first one needs a comprehensive literature review to be approached.

About 42 articles from 1985 to 2019 were selected and synthesised into a grid grouping 218 VOCs by frequency of citation as potential biomarkers. A variety of putative biomarkers was chosen with an educated guess from the list, with the intention to use them during lab testing. Main choosing criterions were: short half-life in the body, not closely linked to cigarette usage, found as relevant for studies cumulating a large number of test subjects, not found to be exclusively exogenous, not highly correlated with physical activity.

For this experimental test of GSA, 4 VOCs have been picked from the previously listed compounds, as interesting because of their relevance as biomarkers and their variety in mass and chemical functions: 1-propanol, 2-butanone, heptanal, n-decane (in this article, referred to as Propanol, Butanone, Heptanal and Decane). 1-propanol, being very light, is not retained by sorbents such as Tenax®, and therefore cannot be used with techniques requiring sorption (see 2.3).

### 2.3 Gas sample preparation

Usual concentrations of VOC biomarkers are in parts per billion (ppb), often between 10 and 1000 ppb. However, the commercial sensors being tested as a first approach for technique development are announced with a minimum LOD of 1ppm. As a first approach, the tested concentration range was chosen with this information taken into account.

The chosen approach for dilution is to inject between 0.1 and 1 µL of pure liquid VOC into a gas sample bag filled with a known volume. The bag is then heated at 60°C for 30 minutes to ensure complete volatilisation. At this point, the content of the bag is of a few parts per million (ppm) in concentration, and need to be diluted

further. The obtained mix is then fed into a system using Mass Flow Controllers (MFC) and commercial analytical air canisters to dilute the mix further down to the ppb level. The gas volume filling the bags is made replicable using a timer and a MFC limiting the flow. Analytical grade air is humidified using three inline bubblers that give a controlled fraction of the flow (20-90% relative humidity at 40°C). Used commercial canisters are of two type: one is CO<sub>2</sub> enriched air, used as dilution air, and the second is pre-made VOC mixture (Westfalen®), used to provide background VOC interfering compounds (ppm level, to be diluted by MFCs). In this project, the final sample is stored for a short time in another gas sampling bag. Storage is made at elevated temperature (40°C) to avoid any condensation. The samples are prepared in 3L PFA Teflon bags (Hedetech®), chosen for their low VOC footprint.

Between each sample, the array is flushed for at least 4 minutes with humidified analytical air (40% relative humidity at 20°C, stored in a 35L nalophan bag at 40°C for at least one hour before usage).

Switching from blank washout air to sample is done by a manual three-point valve so that no contamination occurs from disconnection and reconnection of sampling bags.

Reproducibility of synthesized atmospheres has been assessed intra-day and inter-day across 10 replications of a mix of three of the four compounds (see 2.2: Heptanal, Pentanone, Decane), using Thermal Desorption Gas Chromatography Mass Spectrometry (TD-GC-MS). A replication is the complete re-creation of a sample by following the complete injection and dilution process. A Trace GC oven and DSQII mass spectrometer (Thermo Fisher Scientific®) is coupled with a TD-100xr (Markes®) thermal desorber to analyze Tenax® TA packed cartridges. Sampling bags containing the atmospheres are connected to the cartridges and emptied with a GilAir Plus Personal (Sensidyne®) sampling pump. To ensure good separation, a VOC-specialized capillary column is used (Rxi®-624Sil MS, 0.25mm ID 1.4um 60m, Restek®). The samples are analyzed using the following GC oven program: hold 50°C for 5min, 10°C/min to 180°C, 6°C/min to 205°C, hold 205°C for 5min.

As the method of reference, GC-MS had to undergo a calibration using the chosen compounds. Calibration of the GC-MS was made with 3 different concentrations (equivalent to 0.01, 0.5, 1.0 ppmv in a filled sampling bag). Three complete repetitions were made for each GC-MS calibration point, with freshly made samples. Standards were made by spiking Tenax®-packed sorption cartridges with pure analyte diluted in analytical-grade methanol. Concentration of synthesized atmospheres was evaluated by sorption on Tenax® and GC-MS analysis. Chromatogram's peaks areas were compared with calibration data, and there never was more than 20% for the coefficient of variation in concentration between replications (intra-day or inter-day). The atmosphere synthesis process was therefore deemed reproducible for mixtures of compounds in low concentration.

#### 2.4 Design of Experiment and calibration

A two-level complete factorial design is used to plan the necessary synthetic gas mixes. The models use five parameters, each with two possible levels: relative humidity (20-90%) and final gaseous concentration of each of the four VOCs: 1-propanol (0-1.59ppmv), 2-pentanone (0-1.41ppmv), Heptanal (0-1.51ppmv), n-Decane (0-1.53ppmv). These concentrations are the result of an injection of 0.4µL of propanol, 0.5µL of Pentanone, 0.7µL of Heptanal, or 1.0µL of Decane in a gas bag containing 2.4L of dry analytical air. For median points, the middle point of the two levels is used. Once injected, the bag is then dynamically diluted with MFCs, taking 30mL/min from the bag combined with 1L/min of CO<sub>2</sub>-enriched and humidified analytical air. As creating and processing samples took time, the experience plan was split into 4 blocks with 8 points and 3 median points each (11 samples per block). A median point was processed at the beginning, the middle and the end of the block. A single block was completed in a day. The fourth block was replicated a second time in order to be adsorbed on sorption cartridges and analysed using GC-MS.

The experience plan itself was generated and analysed using the R language (FrF2 and DoE.base packages). The DoE approach enables the discovery of cross-influences between parameters while keeping the number of test runs (samples) to a manageable amount, which wouldn't be the case with a more traditional approach. However, a calibration of the e-nose using variables and combinations of variables was operated to compare the results of the experience plan with traditional calibration.

For the calibration, three replications of 4 levels of concentration for Pentanone, Heptanal, Humidity, Pentanone with Heptanal, n-Decane and Heptanal with 20% RH at 40°C were performed. A fourth replication was made to be adsorbed on sorption cartridges for GC-MS control analysis. The injected volume of liquid VOC were identical for both calibration and DoE, only the dilution rate varied. For calibration, the following rates were used 15mL/min diluted in 1L/min (output: around 0.75ppm), 30mL/min in 1L/min (1.50ppm), 30mL/min in 0.5L/min (3.0ppm), 30mL/min in 0.3mL/min (5.0ppm). For DoE, only the 1.50ppm dilution rate was used – since it is a 2-points experience plan. Concentrations above 1ppm were compensated for adsorption on cartridges by lowering the volume of sample adsorbed, in regard to the GC-MS calibration range.

## 2.5 Data treatment

Obtained sensor raw conductance data undergoes pre-treatment using a code written in R language. The pre-treatment automatically filters out the signal noise without losing information. This smoothing is done by using a mobile average that adapts its strength according to each sensor's signal. This is done by minimising kurtosis loss while minimising point-to-point steps in absolute value, in this way we avoid the loss of useful outliers while keeping background noise to a minimum.

Then, the signal is analysed: the start and end of each "peak" (rise in conductance induced by gas samples) is identified by looking at the second derivative. The code then extracts useful variables for each peak while taking baseline into account: area under curve, peak maximum, slope (rising and descending) as well as the mean temperature and humidity. Everything is then compiled into a straightforward data frame.

This compiled information was then treated by a custom R-based code for DoE analysis (FrF2 and DoE.base packages). The code extracts a regression line for every variable and every sensor, and also computes second order effects.

## 3. Results and discussion

### 3.1 Calibration versus Experimental Design

Regression lines were calculated using peak maximum (see 2.5) values for each combination of variables and sensor, for both calibration data and DoE data. It is to be underlined that DoE data leave only two points to obtain a regression line (without replications), while calibration data has four (with three replications). The results given from regression line equations for pentanone, heptanal, decane and humidity are summarised in Table 1. Regression slopes express the sensor's sensitivity to the tested variable. For clarity and compactness' sake, we compared sensor's calculated slopes across compounds and attributed a qualitative appreciation of its sensitivity. Across all tested commercial sensors, sensitivities vary to some extent. Interestingly, some sensors seem to react to the presence of Decane (G3530, MP901...), and humidity appears to reduce the sensitivity to Heptanal for most sensors. The mixture of Pentanone and Heptanal has shown that the response of most sensors to the mixture was inferior to the sum of the individual response of each compounds (the response of MP901 makes exception with a response roughly equal to the sum). Sensitivity to humidity varies greatly between sensors, some of them (G8530) displaying close to no reaction to water alone.

As a side note, GC-MS quantification cannot be operated with Propanol for this experiment due to technical considerations. Therefore, propanol was not used for calibration and was not quantified for DoE results.

*Table 1 : Sensitivity of the sensors to some of the tested variables, using data from calibration experiments. "0" indicates lack of sensitivity, "+" to "+++" indicates low to great sensitivity (sensitivity is evaluated by comparing a same sensor's slopes when exposed to the different atmospheres)*

Sensor		Variable						
		T2603	G3530	G1430	G2530	G8530	MP901	BME680
Humidity		0	++	+++	++	0	0	+++
Pentanone		++	+++	++	++	+++	+++	+
Heptanal	90% HR	+++	++	++	++	++	++	+
Heptanal	20% HR	+++	+++	+++	+++	++	+++	+++
Decane		0	++	+	+	0	+	0
Pentanone + Heptanal		+++	+++	+++	+++	++	+++	++

However, these results are not comparable with those of the DoE, displaying 4% to 3000% difference. Some DoE regression slopes are even negative, which is unexpected and raises suspicion on the dataset. Since there are only two points for each slope in the DoE approach, any mistake during sample processing drastically changes the outcome. This underlines the differences of the two methods and the impossibility to use DoE in place of calibration due to impracticality and incertitude.

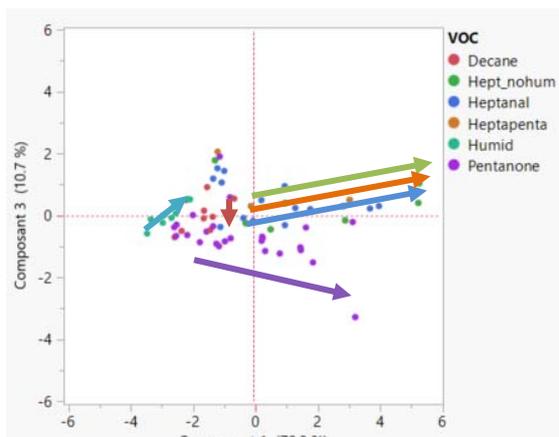


Figure 1a: PCA on calibration data. Arrows point toward increasing concentration/RH, with respect to legend's colours.

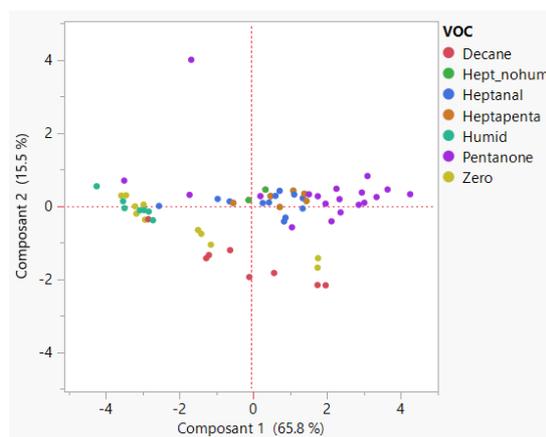


Figure 1b: PCA on normalized calibration data. Normalization removes the effect of concentration. "Zero" are 90% RH blanks.

The hypothesis that the difference between DoE and calibration resides in the lack of reproducibility of the samples or sensors' readings can be excluded for two reasons: firstly, there was little variation across the four days and twelve median points (between 6 and 20% coefficient of variation depending on the sensor model, with values very close to those found in the GC-MS reproducibility analysis for both inter and intra-day), indicating a good reproducibility of both sample synthesis and sensor readings in DoE. Secondly, similar samples used in DoE and calibration give similar sensor readings according to a statistical t-test (2% significance level). We can therefore conclude that samples did not vary significantly in composition and sample synthesis errors can be excluded.

However, for some samples, values differ surprisingly greatly, so much in fact that simple uncontrolled variations in concentration are not an acceptable explanation. One likely explanation would be that samples were perturbed in their processing order during the DoE experiment, resulting in readings differing completely from those of calibration. This would explain the negative slope obtained with some regression lines on DoE data, which would not be possible otherwise. As most variables are evaluated with only two points during DoE, any error in processing translates into erroneous results. In that way, Design of Experiments has no error tolerance. One of the main advantages of DoE for the purpose envisioned here is that it saves time. However, time saving aspects have to be put into perspective with the quality of the data obtained. The calibration approach provides three repetitions of 4 different concentrations and can cover one variable per day with the used setup. Using our setup, the DoE approach provides one repetition of 2 different concentrations and can cover all variables in four days. As we want second order interactions as well as first order interactions for 5 variables (RH and 4 VOCs), the number of needed regressions lines is 15. Therefore, the calibration approach amounts to 15 days. On the other hand, the DoE approach amounts to 3 times 4 days to equal the number of replications of the calibration approach. DoE approach is therefore shorter than calibration by 3 days only if we want a similar level of statistical power. This experiment is also interesting for other projects, notably in environmental applications: while monitoring odours and gases, it is commonplace to encounter chemical species of interest to which an electronic nose needs to be sensitive. Calibration allows for feature selection and array optimization prior to field use. General recommendations regarding humidity and VOC sampling are also applicable for other studies including VOCs, as these compounds are being studied extensively as pollutants of indoor and outdoor air.

### 3.2 Principal Component Analysis

When using data from a gas sensor array, using Principal Component Analysis (PCA) brings new information on the performance of the array as a whole. It gives insight on the separability of the samples and the expected quality of the discrimination. The following analysis is therefore completely unrelated to what was previously presented but is based on the calibration data nonetheless. Figure 1a shows the PCA for components 1 and 3, which had the clearer separation between groups (component 2 explained 11% of the variance but performed less). Using the values of each sensor, our setup has shown that it could differentiate most of the used mixtures and differentiate between varying concentrations. As expected, there is some overlap between groups with similarities (Heptanal, Heptanal with less humidity, Heptanal with Pentanone (50% each)). The influence of concentration on the PCA can be removed by normalising sensor data (Gardner & Bartlett, 1999), which often helps separating groups. Figure 1b shows this, blanks ("Zero" in the legend) and

samples containing humidity alone being further apart from VOC samples. Decane samples are well separated, but somewhat close to some “Zero” samples. Since Zero samples are blank samples, and Decane has a tendency to build up within sample bags despite numerous washing procedures with dry clean air, it has been confirmed by GC-MS analysis that these blank samples were in fact contaminated with Decane. Heptanal shows a rather good separation from Pentanone, with the Heptanal/Pentanone mix logically appearing between them. Heptanal with less humidity appears confounded with Heptanal with 90% RH, which indicates that the sensor array can identify Heptanal regardless of the humidity level.

#### 4. Conclusions

This paper answers the needs of a project aimed at detecting early-stage lung cancer with a breath-analysing gas sensor array. Commercial sensors composing the array need to be characterized due to project needs. We are therefore away from any form of breath discrimination testing. Breath is a complex mixture of gases that varies in time and between individuals, and therefore requires a whole different approach than simple calibration or experimental design. The signature identifying a mixture is made up of every sensor’s response, and individual sensitivities are not sufficient. It is however important for sensors to react differently to various compounds, and knowing how each of them behave is useful for array improvements. As an array, the sensors are able to distinguish VOCs between them and from humid air and show encouraging discrimination results on PCA plots. It is also apparent on normalized PCA plots that Heptanal could be identified as such regardless of the level of humidity in the sample, which is good as it seems that humidity tends to reduce sensors’ sensitivity to Heptanal. We exposed a sensor array to synthetic gas mixes containing varying contents in humidity and four VOCs according to a factorial experimental design, and compared the obtained regression lines to the ones from calibration data. The goal was to evaluate the general behaviour of sensors and see if it was possible to substitute calibration with DoE analysis. Using DoE would have been time-saving and helped gaining more information from a single experiment. Unfortunately, we cannot conclude that calibration and experimental design are similar on the basis of the obtained data. DoE data cannot be used as a substitute for calibration. Replications of the experiment would be needed to firmly confirm these conclusions. Most MOS sensors are, in general, insensitive to long-chain alkanes, as it was observed for almost every sensor tested here. However, two of the tested sensors have shown a positive response to increasing concentrations of decane, which is found odd as a previous experiment exposed no such behaviour. PCA of normalized data also shows good separation of Decane from the other VOCs. A complementary experiment will be held to confirm this finding.

#### Acknowledgments

This paper was redacted under the framework of PATHACOV project, funded under the Interreg France-Wallonie-Vlaanderen program, with the support of the European Regional Development Fund.

#### References

- Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., Gavin, A., Visser, O., & Bray, F. (2018). Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer*, 103, 356-387. <https://doi.org/10.1016/j.ejca.2018.07.005>
- Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J. W. W., Comber, H., Forman, D., & Bray, F. (2015). Reprint of: Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *European Journal of Cancer*, 51(9), 1201-1202. <https://doi.org/10.1016/j.ejca.2015.05.004>
- Gardner, J. W., & Bartlett, P. N. (1999). *Electronic noses, principles and applications*.
- National Lung Screening Trial Research Team, Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., Gareen, I. F., Gatsonis, C., Marcus, P. M., & Sicks, J. D. (2011). Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine*, 365(5), 395-409. <https://doi.org/10.1056/NEJMoa1102873>
- PATHACOV - Diagnostic de pathologies humaines par analyse de COV dans l'air expiré. (2020). [Project website]. <https://pathacov-project.com/>
- Plummer, M., de Martel, C., Vignat, J., Ferlay, J., Bray, F., & Franceschi, S. (2016). Global burden of cancers attributable to infections in 2012: A synthetic analysis. *The Lancet Global Health*, 4(9), e609-e616. [https://doi.org/10.1016/S2214-109X\(16\)30143-7](https://doi.org/10.1016/S2214-109X(16)30143-7)
- Righettoni, M., Amann, A., & Pratsinis, S. E. (2015). Breath analysis by nanostructured metal oxides as chemoresistive gas sensors. *Materials Today*, 18(3), 163-171. <https://doi.org/10.1016/j.mattod.2014.08.017>
- Silvestri, G. A., Pastis, N. J., Tanner, N. T., & Jett, J. R. (2016). Clinical Aspects of Lung Cancer. In Murray and Nadel's Textbook of Respiratory Medicine (p. 940-964.e22). Elsevier. <https://doi.org/10.1016/B978-1-4557-3383-5.00053-1>