Accepted refereed manuscript of: Dobson R & Semple S (2018) "How do you know those particles are from cigarettes?": An algorithm to help differentiate second-hand tobacco smoke from background sources of household fine particulate matter, *Environmental Research*, 166, pp. 344-347. DOI: https://doi.org/10.1016/j.envres.2018.06.019.

Report from the field

"How do you know those particles are from cigarettes?": An algorithm to help differentiate second-hand tobacco smoke from background sources of household fine particulate matter

Ruaraidh Dobson BSc (Hons) (corresponding author)

Institute for Social Marketing, University of Stirling

Tel: +44 7803 406 343

Email: r.p.dobson@stir.ac.uk

Sean Semple PhD

Institute for Social Marketing, University of Stirling

FUNDING STATEMENT

This research was carried out as part of a PhD study by one of the authors (RD) funded by a staff-waiver programme at the University of Aberdeen.

COMPETING INTERESTS STATEMENT

The authors have no competing interests to declare.

CONTRIBUTORSHIP STATEMENT

Ruaraidh Dobson contributed to experimental design and planning, carried out fieldwork and statistical analysis, drafted this paper and had final approval of the version to be published. Sean Semple contributed to experimental design and planning, redrafting and critical revision, and had final approval of the version to be published. Both authors are accountable for all aspects of the work.

KEYWORDS

Second-hand smoke; tobacco smoke exposure; air quality monitoring; particulate matter

ABSTRACT

Background

Second-hand smoke (SHS) at home is a target for public health interventions, such as air quality feedback interventions using low-cost particle monitors. However, these monitors also detect fine particles generated from non-SHS sources.

The Dylos DC1700 reports particle counts in the coarse and fine size ranges. As tobacco smoke produces far more fine particles than coarse ones, and tobacco is generally the greatest source of particulate pollution in a smoking home, the ratio of coarse to fine particles may provide a useful method to identify the presence of SHS in homes.

Methods

An algorithm was developed to differentiate smoking from smoke-free homes. Particle concentration data from 116 smoking homes and 25 non-smoking homes were used to test this algorithm.

Results

The algorithm correctly classified the smoking status of 135 of the 141 homes (96%), comparing favourably with a test of mean mass concentration.

Conclusions

Applying this algorithm to Dylos particle count measurements may help identify the presence of SHS in homes or other indoor environments. Future research should adapt it to detect individual smoking periods within a 24h or longer measurement period.

INTRODUCTION

Second-hand smoke (SHS) is a serious cause of poor indoor air quality in homes. Around 40% of children are regularly exposed worldwide,(GTSS Collaborative Group, 2006) putting them at risk of serious illness and impaired lung development(US Surgeon General, 2006).

For that reason interventions to promote smoke-free homes are of significant public health interest. Several interventions have been developed using air quality monitoring to inform parents of the impact of smoking on their indoor air quality, and the consequent effects on their children.(Dobson et al., 2017; Klepeis et al., 2013; Rosen et al., 2015; Wilson et al., 2013) A low-cost air quality monitor, the Dylos DC1700, has proved useful for monitoring $PM_{2.5}$ as a proxy for SHS in smokers' homes in these kinds of interventions.(Semple et al., 2015, 2013) The Dylos is a small, portable monitor which provides comparable accuracy at a considerably lower price than other widely used optical particle counters, such as the TSI Sidepak.In addition to being approximately one-tenth of the cost of the Sidepak instrument, the Dylos has several specific advantages in terms of low noise, simplicity of use and the ability to determine particle size distribution in terms of fine and coarse particulate (Semple et al., 2013)

$PM_{2.5}$ has been widely used as a proxy to quantify indoor concentrations of SHS in many settings including bars, homes and vehicles (Apelberg et al., 2013; Gorini et al., 2005) as reliable measurements can be taken easily and affordably over time using optical particle counters, in contrast to the high cost and complexity of more specific methods such as air nicotine measurement. Other activities in these settings can generate $PM_{2.5}$. These can include cooking emissions, combustion such as candle burning or the use of solid fuels for heating, and aerosols such as deodorants and hair sprays.(He et al., 2004) These sources can produce high concentrations of PM within a home which could be confused for SHS in interpretation.

Parents in previous intervention trials have been observed to deny and challenge messages about the risk of SHS, (Passey et al., 2016) and if feedback wrongly identifies non-SHS sources as being smoking activity this is likely to weaken the effectiveness of such approaches and make the participant question the validity of the measurement method. Developing reliable and accurate information on PM concentrations that are specifically linked to SHS is therefore important in the development of effective interventions.

The particle size distribution of tobacco smoke is known to skew towards fine and ultrafine particles. (Klepeis et al., 2003) The mean diameter of particles in tobacco smoke has been measured as 0.27µm (in the case of mainstream smoke) and 0.09µm (for sidestream smoke); smaller mean diameters than those associated with common household activities like frying, cleaning and the movement of people, and other sources (Abt et al., 2000) while still producing a sustained increase in particle mass concentration over time.(Semple and Latif, 2014)

The Dylos DC1700 provides data on both the fine and coarse fractions of particulate matter in the form of particle counts for particles larger than 0.5µm and particles larger than 2.5µm. It may therefore be possible to use this particle size information to distinguish between different sources of PM in a home, and potentially to classify homes as smoking or non-smoking.

This research uses particle concentration data measured in homes to develop and test a rule-based approach to determine whether tobacco was smoked in the home during the monitoring period. This information could be useful in providing air quality data to support behavioural interventions designed to encourage smokers to keep their homes smoke-free.

MATERIALS AND METHODS

Measuring mass concentrations and particle counts in homes

Previously reported methods (Semple et al., 2013) were used to assess $PM_{2.5}$ concentrations in homes. From previous work by our group (Semple, 2016), time resolved $PM_{2.5}$ data were already available from 116 smoking homes. Data from non-smoking homes were collected in the course of this research. Minute-by-minute particle counts reported by the Dylos DC1700 monitor were converted to estimated $PM_{2.5}$ concentrations using a previously developed equation (Equation 1).

$$PM_{2.5} = 0.65 + 4.16 \times 10^{-5}(Dylos\ total\ particle\ count - large\ particle\ count) + 1.57 \times 10^{-11}(Dylos\ total\ particle\ count - large\ particle\ count)^2$$

Equation 1 - Conversion of Dylos particle counts to approximate mass concentration (Semple et al., 2013)

Also, the large particle percentage, consisting of the particles larger than 2.5µm as a percentage of the total particles detected, was calculated for each minute for use in the algorithm.

Algorithm development

A four-step algorithm was developed to classify homes as smoking or non-smoking based on one day or more of Dylos-recorded data by excluding data points which were unlikely to be related to smoking. This algorithm was designed to use the ratio of large to small particles detected by the Dylos as a "signature" for the presence of SHS. Additional steps were intended to reduce noise in the data caused by brief fluctuations in levels of PM.

For each home:

1. Remove data where $PM_{2.5}$ concentration is below 5µg/m^3. This step is intended to account for low ambient concentrations of $PM_{2.5}$ which are not related to SHS. 5µg/m^3 was chosen as indoor $PM_{2.5}$ has been shown to correlate to 79% of

ambient $PM_{2.5}$ in similar conditions (Cyrys et al., 2004), while the average ambient $PM_{2.5}$ concentration in Scotland has been modelled at 6.6µg/m^3.(Sykes, 2016) Previous research on smoke-free homes has shown

2. For each minute of data, calculate the percentage of the total detected particles which are larger than 2.5µm in diameter. Remove data where the percentage of large particles is greater than a threshold (described throughout as the ‘Large Particle Threshold’ or LPT).
3. Remove data where a peak lasts for fewer than three minutes, to account for random fluctuations compared to the sustained impact of SHS on indoor air quality.(Semple and Latif, 2014)
4. Take the percentage of minutes in the log where data has not been removed in one of the steps above. This can be used as an “SHS score” to classify the home as smoking or non-smoking if the score is above a cut-off (determined experimentally).

Statistical analysis

Use of the algorithm relies on two factors: the LPT which best indicates smoking, and the best-performing cut-off value for the SHS score, over which a log can be classified as smoking. Receiver operating characteristic curves were used to determine these factors.

ROC curves are a common method for determining the efficacy of a diagnostic test. (Bewick et al., 2004) In an ROC curve, a test is carried out on a set of records, and its specificity and selectivity are plotted. This allows comparison between different tests using the area under the curve (AUC) of this plot – a mathematical representation of the overall effectiveness of the test. Tests which classify records more successfully than random have AUC values greater than 0.5, while a hypothetical perfect test would have a value of 1.0.

Variants of the algorithm using LPTs between 0.1% and 4.0% (stepped up in 0.1% increments) were applied to the full dataset of logs and the categorisation results plotted on an ROC curve using IBM SPSS v24.(IBM Corp., 2016) The LPT which resulted in the highest AUC was selected, and the curve analysed to find the SHS score cut-off which maximised selectivity and specificity. An ROC curve was also generated using the mean $PM_{2.5}$measured in each household as a predictor of smoking status. Custom Python 2.7 scripts were developed to apply the algorithm to Dylos data logs.

Smoke-free homes data collection

Participants working at three health charities in Scotland were recruited. Only people living in homes where smoking or e-cigarette use was not permitted were eligible to participate in the study. A target of 30 people was set as achievable with the time and resources available.

Participants were given a Dylos DC1700 monitor and an instruction sheet asking them to install and run the monitor for 48 hours in their main living space, elevated above floor level and away from doors and windows. This mirrored instructions given during previous studies of personal exposure to SHS.(Semple et al., 2012) Participants were also asked to keep a diary of events which could cause elevated PM in the home, including cooking and heating use.

After the monitoring period, the Dylos was returned to the research team and data was downloaded from it. A short report on air quality in the home was prepared for the participant and emailed to them, along with any relevant information on reducing air pollution in their home. The monitor's memory was then cleared prior to use with the next participant.

Smoking homes data

The pre-existing smoking homes dataset comprised minute-by-minute measurements from 116 homes, each spanning approximately 5 days, taken from the First Steps 2 Smoke-free

(FS2SF) study(Semple, 2016). Participants in that study self-reported that smoking took place regularly within the home. No data on other events which could affect air quality was available from these homes.

Ethics

Ethical approval for this study was given by the College Ethics Review Board of the College of Life Sciences and Medicine at the University of Aberdeen.

RESULTS

Estimated $PM_{2.5}$ concentrations in smoking and smoke-free homes

For the smoke-free home data collection part of the study 27 participants were recruited, with 25 of those completing the study. Homes were monitored for a mean of two days, eight hours and six minutes (ranging from one day, 20 hours and 45 minutes to three days, 13 hours and 32 minutes). Two participants withdrew or were unable to provide 24h of data.

When compared to the existing data from 116 smoking homes the 25 smoke-free homes had significantly lower concentrations of $PM_{2.5}$, with a geometric mean of 5.2µg/m^3 (geometric standard deviation (GSD) ±2.16), compared to 37.6µg/m^3 (GSD ±3.0) ($p<0.001$, natural logs compared with a one-tailed Student's t-test).

Mean large particle percentages were also significantly different, with a geometric mean in smoke-free homes of 7.49% compared to 3.56% in smoking homes ($p<0.001$, natural logs compared with Student's t-test).

Classifying homes algorithmically

Potential large particle percentages were selected based on previous research on the particle size distribution of SHS.(Klepeis et al., 2003) The algorithm was applied to the whole dataset using LPTs between 0.1% and 4.0%, incremented by 0.1%. The resulting output was plotted as an ROC curve to determine the LPT which maximised AUC. An LPT of 1.8% was most successful, with an AUC of 0.945.

Comparison with classification by $PM_{2.5}$ concentration

An ROC curve was plotted using the results of the algorithm along with the mean $PM_{2.5}$ concentrations of each home (Figure 1).

The AUC for selection using mean $PM_{2.5}$ was 0.937, while the algorithm classification attained 0.945. Both methods were highly successful in classifying homes, with the algorithm more successful (although this was not statistically significant).

Determining the SHS score classification cut-off

Coordinates of the curve values were examined to determine the best-performing SHS score. The value 1.455% maximised sensitivity (0.974) and specificity (0.88) and was therefore selected. Using this value, the algorithm classified only 3/116 smoking and 3/25 non-smoking homes incorrectly.

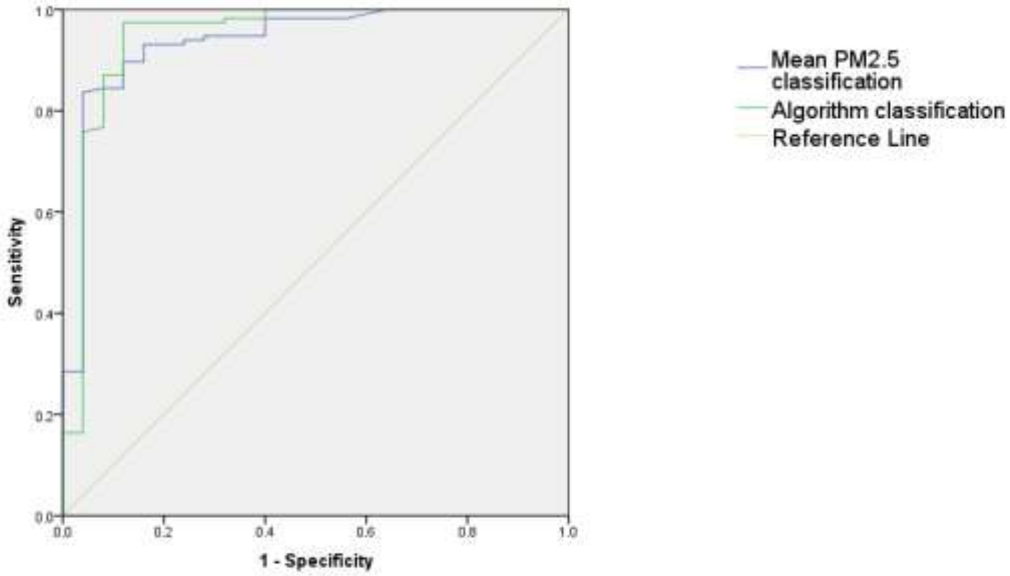

Figure 1 – Receiver operating characteristic curve comparing home classification using the algorithm with home classification using mean PM2.5 in a home. Mean $PM_{2.5}$ classification refers to the use of mean $PM_{2.5}$ concentrations in isolation as a score to classify homes as smoking or smoke-free. Algorithm classification refers to the use of steps 1-3 of the algorithm to produce an SHS score which was used as a classifier. Curves approaching the upper left corner of the chart represent effective classifiers. Sensitivity refers to the number of true positive smoking homes identified, while 1 – specificity refers to the number of false positives identified. Coordinates of the curve were analysed separately to decide on an SHS score cut-off point which would best indicate a smoking home.

DISCUSSION

It is possible to apply a simple algorithm to Dylos DC1700 particle number counts in order to predict with a high degree of certainty whether smoking occurred in a home during a multi-day monitoring period. While mean $PM_{2.5}$ concentration in the homes measured is clearly linked to smoking status, the algorithm was able to characterise homes independently from that factor, suggesting that the additional steps linked to large particle percentage and removing data where low concentrations of $PM_{2.5}$ are present are useful additions to the process of determining the presence of SHS in a home.

Previously, data from a similar monitor has been used to develop a logistic regression model to distinguish SHS from non-SHS sources of PM.(Dacunto et al., 2014) In this study, a large particle threshold of 1% was identified as indicative of SHS, similar to the threshold identified in this paper.

$PM_{2.5}$ is a well-recognised marker for SHS(Gorini et al., 2005) which has been used in a number of behavioural interventions.(Klepeis et al., 2013; Rosen et al., 2015; Wilson et al., 2013) PM sensors are well-developed, easily portable and inexpensive, allowing them to be used in a wide range of settings where it may be useful to measure SHS. This need not be limited to homes – for instance, this technique could be used to promote smoke-free public places laws.

Limitations

The limited number of measurements available from smoke-free homes made it impossible to determine whether the algorithmic approach was statistically superior to an approach based purely on the use of logistic regression analysis on the mean level of $PM_{2.5}$. A follow-up study in further smoke-free homes may determine this.

Although the classification rate established by the algorithmic approach was high, one in eight of the smoke-free homes tested was still mis-classified. These false positives could cause intervention participants who have succeeded in keeping their homes smoke-free to be told that they have not done so, potentially reducing their trust in the intervention.

Scotland has relatively low levels of outdoor particulate air pollution. The large particle threshold and score values developed in these circumstances may not hold true in countries with high levels of coarse particle pollution (including dust storms or other natural particulate pollution).(Ahmed et al., 1987) Further research should be carried out in these conditions.

In step 1 of the algorithm we have assumed that 79% of ambient PM will infiltrate, leading to an ambient $PM_{2.5}$ concentration indoors of around 5µg/m^3. Values below this concentration are therefore excluded from the result. Infiltration of ambient PM varies greatly depending on building ventilation and other factors, and so this assumption is unlikely to hold true generally. Furthermore, measurements in settings where ambient PM is significantly higher than 6.6μg/m^3 may cause few data points to be excluded at step 1, affecting the results of the algorithm. Further research is most likely necessary to test the algorithm in a variety of other settings, and to test the assumptions implicit in step 1. It may be beneficial to generate a specific "ambient $PM_{2.5}$" value for the time periods in which measurement takes place.

The results of this research can only be applied directly to the Dylos DC1700, with its two size bins. It may be possible to adapt the algorithm to other optical particle counters with multiple size bins, but research would be needed to measure their agreement with the size classifications made by the Dylos.

In general, optical particle counters are limited instruments compared to more labour- and time-intensive methods of detecting and quantifying PM, such as gravimetric methods. A wide range of factors can affect their results, including relative humidity, (Ruprecht et al., 2011) aerosol composition and the age of cigarette smoke in the air.(Dacunto et al., 2015) The particle number to mass concentration equation used in this study has been developed with reference to SHS aerosol only, so mass concentrations calculated by this method should be considered as estimates or approximations of exposure.

The effectiveness of the algorithm may be impeded in settings where there are other significant sources of $PM_{2.5}$, such as open flames. Similarly, high concentrations of PM_{10} in outdoor air could impede the effectiveness of the algorithm, raising the percentage of large particles measured by the monitor. This would be a particular concern in countries with high levels of outdoor air pollution.

Implications

Due to the well-known health implications of $PM_{2.5}$ in air and of SHS, particularly for children, interventions to reduce the number of homes in which smoking takes place are of importance in improving public health, with several recent studies describing the use of the Dylos DC1700 monitor.(Klepeis et al., 2013; Rosen et al., 2015; Semple et al., 2013)

While most people are well aware that SHS is harmful, many smokers blame other factors such as outdoor air pollution when presented with evidence of poor air quality in their homes. Researchers developing air quality feedback interventions for smoke-free homes or smoking cessation should consider incorporating this classification method to reinforce the specific danger of SHS.

Although this study took place solely in homes, the algorithm could be used to detect SHS in other indoor environments such as bars, casinos and other workplaces. This could be useful in assessing occupational exposure to SHS, and in providing evidence for advocacy for comprehensive smoke-free legislation.

REFERENCES

Abt, E., Suh, H.H., Allen, G., Koutrakis, P., 2000. Characterization of indoor particle sources: A study conducted in the metropolitan Boston area. Environ. Health Perspect. 108, 35–44.

Ahmed, A.S., Ali, A.A., Alhaider, M.A., Av, H.-A., 1987. Measurement of atmospheric particle size distribution during sand/duststorm in Riyadh, Saudi Arabia. Atmos. Environ. 21, 2723–2725. https://doi.org/10.1016/0004-6981(87)90205-8

Apelberg, B.J., Hepp, L.M., Avila-Tang, E., Gundel, L., Hammond, S.K., Hovell, M.F., Hyland, A., Klepeis, N.E., Madsen, C.C., Navas-Acien, A., Repace, J., Samet, J.M., Breysse, P.N., 2013. Environmental monitoring of secondhand smoke exposure. Tob. Control 22, 147–155. https://doi.org/10.1136/tobaccocontrol-2011-050301

Bewick, V., Cheek, L., Ball, J., 2004. Statistics review 13: Receiver operating characteristic curves. Crit. Care 8, 508. https://doi.org/10.1186/cc3000

Cyrys, J., Pitz, M., Bischof, W., Wichmann, H., Heinrich, J., 2004. Relationship between indoor and outdoor levels of fine particle mass, particle number concentrations and black smoke under different ventilation conditions. J. Expo. Anal. Environ. Epidemiol. 14, 275–283. https://doi.org/10.1038/sj.jea.7500317

Dacunto, P.J., Cheng, K.-C., Acevedo-Bolton, V., Jiang, R.-T., Klepeis, N.E., Repace, J.L., Ott, W.R., Hildemann, L.M., 2014. Identifying and quantifying secondhand smoke in source and receptor rooms: logistic regression and chemical mass balance approaches. Indoor Air 24, 59–70. https://doi.org/10.1111/ina.12049

Dacunto, P.J., Klepeis, N.E., Cheng, K.-C., Acevedo-Bolton, V., Jiang, R.-T., Repace, J.L., Ott, W.R., Hildemann, L.M., 2015. Determining $PM_{2.5}$ calibration curves for a low-cost particle monitor: common indoor residential aerosols. Environ. Sci. Process. Impacts 17, 1959–1966. https://doi.org/10.1039/C5EM00365B

Dobson, R., O’Donnell, R., de Bruin, M., Turner, S., Semple, S., 2017. Using air quality monitoring to reduce second-hand smoke exposure in homes: the AFRESH feasibility study. Tob. Prev. Cessat. 3. https://doi.org/10.18332/tpc/74645

Gorini, G., Gasparrini, A., Fondelli, M.C., Invernizzi, G., 2005. Second-Hand Smoke (SHS) Markers: Review of methods for monitoring exposure levels. Brussels.

GTSS Collaborative Group, 2006. A cross country comparison of exposure to secondhand smoke among youth. Tob. Control 15, ii4-ii19. https://doi.org/10.1136/tc.2006.015685

He, C., Morawska, L., Hitchins, J., Gilbert, D., 2004. Contribution from indoor sources to particle number and mass concentrations in residential houses. Atmos. Environ. 38, 3405–3415. https://doi.org/10.1016/j.atmosenv.2004.03.027

IBM Corp., 2016. IBM SPSS Statistics for Windows, version 24.0.

Klepeis, N.E., Apte, M.G., Gundel, L.A., Sextro, R.G., Nazaroff, W.W., 2003. Determining Size-Specific Emission Factors for Environmental Tobacco Smoke Particles. Aerosol Sci. Technol. 37, 780–790. https://doi.org/10.1080/02786820390220254

Klepeis, N.E., Hughes, S.C., Edwards, R.D., Allen, T., Johnson, M., Chowdhury, Z., Smith, K.R., Boman-Davis, M., Bellettiere, J., Hovell, M.F., 2013. Promoting smoke-free homes: a novel behavioral intervention using real-time audio-visual feedback on airborne particle levels. PLoS One 8, e73251. https://doi.org/10.1371/journal.pone.0073251

Passey, M.E., Longman, J.M., Robinson, J., Wiggers, J., Jones, L.L., 2016. Smoke-free homes: what are the barriers, motivators and enablers? A qualitative systematic review and thematic synthesis. BMJ Open 6, e010260. https://doi.org/10.1136/bmjopen-2015-010260

Rosen, L., Zucker, D., Hovell, M., Brown, N., Ram, A., Myers, V., 2015. Feasibility of Measuring Tobacco Smoke Air Pollution in Homes: Report from a Pilot Study. Int. J. Environ. Res. Public Health 12, 15129–42. https://doi.org/10.3390/ijerph121214970

Ruprecht, A.A., De Marco, C., Boffi, R., Mazza, R., Lopez, M.J., Moshammer, H., Dautzenberg, B., Clancy, L., Precioso, J., Invernizzi, G., 2011. Mass calibration and Relative Humidity compensation requirements for optical portable particulate matter monitors : the IMPASHS (Impact of smoke-free policies in EU Member States) WP2 preliminary results. Epidemiology 22.

Semple, S., 2016. First Steps to Smoke-free (FS2SF): using air-quality feedback to facilitate smoke-free homes, in: Scottish Smoking Cessation Conference.

Semple, S., Apsley, A., MacCalman, L., 2013. An inexpensive particle monitor for smoker behaviour modification in homes. Tob. Control 22, 295–298. https://doi.org/10.1136/tobaccocontrol-2011-050401

Semple, S., Garden, C., Coggins, M., Galea, K.S., Whelan, P., Cowie, H., Sánchez-Jiménez, A., Thorne, P.S., Hurley, J.F., Ayres, J.G., 2012. Contribution of solid fuel, gas combustion, or tobacco smoke to indoor air pollutant concentrations in Irish and Scottish homes. Indoor Air 22, 212–223. https://doi.org/10.1111/j.1600-0668.2011.00755.x

Semple, S., Ibrahim, A.E., Apsley, A., Steiner, M., Turner, S., 2015. Using a new, low-cost air quality sensor to quantify second-hand smoke (SHS) levels in homes. Tob. Control 24, 153–8. https://doi.org/10.1136/tobaccocontrol-2013-051188

Semple, S., Latif, N., 2014. How long does secondhand smoke remain in household air: analysis of PM2.5 data from smokers' homes. Nicotine Tob. Res. 16, 1365–70. https://doi.org/10.1093/ntr/ntu089

Sykes, D., 2016. PM2.5 and PM10 in Scotland. Glasgow.

US Surgeon General, 2006. The Health Consequences of Involuntary Exposure to Tobacco Smoke, The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General. Centers for Disease Control and Prevention (US).

Wilson, I., Semple, S., Mills, L.M., Ritchie, D., Shaw, A., O'Donnell, R., Bonella, P., Turner, S., Amos, A., 2013. REFRESH--reducing families' exposure to secondhand smoke in the home: a feasibility study. Tob. Control 22, e8. https://doi.org/10.1136/tobaccocontrol-2011-050212