Ranges of mainstream smoke formaldehyde yields from contemporary cigarette products

Oscar M Camacho, David Thorne, Kevin McAdam, Chuan Liu, and Martin Ward
British American Tobacco, Group Research & Development, Regents Park Road, Southampton, SO15 8TL, UK

Corresponding author : oscar_m_camacho@bat.com

INTRODUCTION

Cigarette smoke is a dynamic and complex aerosol containing over 5600 identified components, and possibly many thousands of further, unidentified constituents (Perfetti and Rodgman 2011). More than 150 of these constituents have toxic properties, and a number of researchers have attempted to distinguish which of these toxicants make the greatest contribution to the incidence of smoking related diseases (Fowles and Dybing, 2003).

Formaldehyde is an IARC Group 1 carcinogen present in cigarette smoke. Its yields from cigarettes are mandated for annual disclosure in Canada and Brazil, and is a constituent included on the FDA TPSAC draft initial list of Harmful or Potentially Harmful constituents of tobacco and cigarette smoke. Concerns over its potential contribution to smoking related diseases have also led to proposals for limits on the ratio of formaldehyde to nicotine yields as measured under Health Canada Intense (HCI) machine smoking conditions (Burns et al., 2008).

However, there is surprisingly little published data on formaldehyde yields from contemporary commercial cigarettes. Of the many thousands of cigarette brands on sale globally there are around 150 recent published ISO formaldehyde smoke yield values (and even fewer HCI smoke yields) covering a relatively narrow number of geographic sources.

This small dataset of machine measured formaldehyde yields limits our understanding of the normal ranges of yields that are found globally from commercial cigarette products. Similarly, there is relatively little information available on the differences in yields found from different product styles or different geographic locations.

The current work was conducted to substantially enhance the size of the available formaldehyde yield database, and to provide greater clarity on the ranges of formaldehyde yields commonly obtained from contemporary cigarette products.

DATA SOURCES

A database of formaldehyde smoke yields was assembled from a combination of published data from the scientific literature, regulatory reporting data and British American Tobacco (BAT) measured data. Non-commercial, reference product data and duplicate values from apparently different products were removed from the database. Data was divided into ISO yields and Health Canada Intense yields. FTC data was not combined with ISO data due to differences between the two smoking procedures and potential differences in yields. Massachusetts data was also not used due to the small numbers of products measured under this regime. Details of the data sources, numbers of products and references are provided in the Table.

The final database assembled from these sources comprises 700 products sourced from 69 geographical areas. The data collated into this database was drawn from three separate laboratories; significant inter-laboratory measurement differences are known to exist for some smoke analytes (Intorp et al 2010). Therefore the formaldehyde yield data was plotted against the ISO machine measured NFDPM (tar) yield in Figure 1, with different measurement laboratories distinguished by colour. No substantial systematic difference between laboratories was evident from this visual analysis.

Data Source	Data source	Measurement laboratory	Publication/ reporting year	Smoking Regime	Numbers of products	Ref.
Health Canada	Annual regulatory reporting	Labstat International	2004	ISO/HCI	60*/60	HC (2004)
Health Australia	One-off exercise	Labstat International	2002	ISO/HCI	15/15	Australia (2002)
UK TMA	One-off exercise	Arista Labs Europe	2003	ISO	25	Gregg et al (2003)
PM International Brands	Publication	Labstat International	2004	ISO/HCI	48/48	Counts et al (2004)
British American Tobacco	This poster	British American Tobacco	This poster	ISO	552	This poster

Figure 1: Formaldehyde yields as measured by three different laboratories

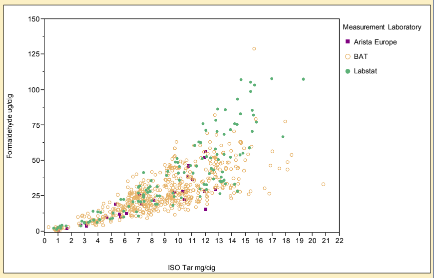
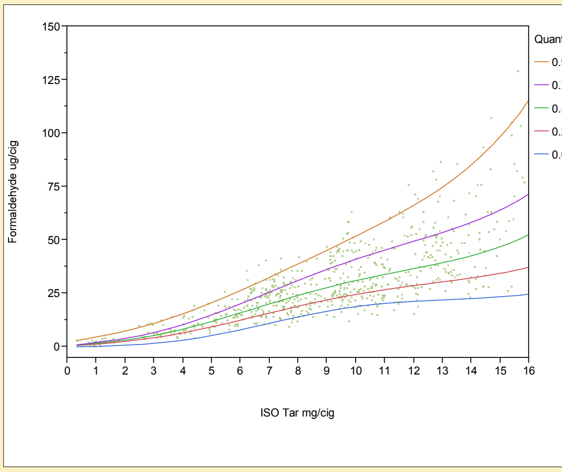


Figure 2: Estimated quantile regression for percentiles 5, 25, 50, 75 and 95 in comparison to the raw data.



DATA ANALYSIS METHOD

The ISO data were analysed by quantile regression using SAS® software. and global median, quartile, and major percentile values for formaldehyde yields at ISO tar from 1 to 15 mg were calculated.

The τth quantile of Y is defined as the inverse function $Q(\tau)=\inf\{y:F(y)\geq \tau\}$ where $0<\tau<1$. In words, it is the value of the response variable for which the probability distribution function in y, that is $F(y)=P(Y\leq y)$, is larger than or equal to the chosen τ.

Quantile regression produces a model of explanatory variables or covariates (in our case Tar) on the conditional percentiles of a response variable (formaldehyde).

The commonly known regression uses the Least-squares method to build a model defined by the conditional mean of the variable response Y given that the explanatory variable X has a particular value x. Similarly, quantile regression assesses the quantile function $Q(\tau \mid X=x)$. This conditional relationship can be written as the linear conditional function $Q(\tau \mid X=x) = x'\beta(\tau)$

which is resolved by minimising $\beta(\tau) = \arg\min \beta \in R^p \sum_{i=1}^{n} \rho_\tau(y_i - x'_i\beta)$ for $\tau \in (0,1)$

Quantile regression does not make assumptions about the distribution of the data. Therefore, data transformation is not required and does not improve the fitting of the regression. However, statistical inference is of benefit when data are normally distributed; hence a logarithmic transformation was carried out.

An empirical approach has been taken when choosing the most appropriate model. Inverse of Tar, root square of Tar, Tar, Tar^2 and Tar^3 were assessed.

For Formaldehyde the model chosen is: $\log Formaldehyde = \beta_0 + \beta_1 Tar + \beta_2 Tar^2 + \beta_3 Tar^3$

Six observations were identified as below LoQ in the formaldehyde data set. In order to create a more realistic data set these data were imputed by drawing random values from a continuous uniform distribution U(LoQ,0].

CONCLUSIONS

This work extends considerably the database of available formaldehyde smoke yields from contemporary cigarette products. In addition to a greater range of cigarette product styles, the database covers a significantly wider geographical scale than previously available.

There is significant variation in the formaldehyde yields measured globally; at a fixed ISO tar yield the variation in yields covers a range of 3-4 fold.

A statistical analysis approach has also been described that creates a mechanism and numerical framework which provides greater clarity on the range of formaldehyde yields commonly found with contemporary cigarette products.

The approach described in this poster has been extended to other toxicants

Poster presented at the CORESTA Smoke Science and Product Technology Joint Study Groups' Meeting, Graz, Austria. October 2011. Poster number STPOST 02

www.bat-science.com

References are available on request