Annals of Work Exposures and Health, 2018, 1–13
doi: 10.1093/annweh/wxy100
Original Article

Original Article

Expostats: A Bayesian Toolkit to Aid the Interpretation of Occupational Exposure Measurements

Jérôme Lavoué[1,2,*], Lawrence Joseph[3], Peter Knott[4], Hugh Davies[5], France Labrèche[1,6], Frédéric Clerc[7], Gautier Mater[7] and Tracy Kirkham[8]

[1]Department of Environmental and Occupational Health, School of Public Health, University of Montreal, 2375, chemin de la Côte Ste-Catherine, Montréal, Québec, H3T1A8, Canada; [2]University of Montreal hospital research center, 850 rue St-Denis, Montréal, Québec, H2X 0A9, Canada; [3]Division of clinical epidemiology, McGill University Health Centre, 2155 Guy Street, Montreal, Québec, H3H2R9, Canada; [4]GCG Health Safety Hygiene, 7/34 Navigator Place, Hendra, QLD 4011, Australia; [5]School of Population & Public Health, University of British Columbia, 2206 East Mall, Vancouver, British Columbia, V6T1Z3, Canada; [6]Institut de recherche Robert-Sauvé en santé et en sécurité du travail, 505 boul. De Maisonneuve Ouest, Montréal, Québec H3A3C2, Canada; [7]Institut National de Recherche et de Sécurité pour la prévention des accidents du travail et des maladies professionnelles (INRS), 65 boulevard Richard Lenoir, 75011 Paris, France; [8]Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, Ontario, M5T3M7, Canada

*Author to whom correspondence should be addressed. E-mail: jerome.lavoue@umontreal.ca

Submitted 12 July 2018; revised 1 November 2018; editorial decision 5 November 2018; revised version accepted 13 November 2018.

Abstract

Introduction: Interpretation of exposure measurements has evolved into a framework based on the lognormal distribution. Most available practical tools are based on traditional frequentist statistical procedures that do not satisfactorily account for censored data and are not amenable to simple probabilistic risk statements. Bayesian methods offer promising solutions to these challenges. Such methods have been proposed in the literature but are not widely and freely available to practitioners.
Methods: A set of computer applications were developed aimed at answering typical inferential questions that are important to occupational health practitioners: Is a group of workers compliant with an occupational exposure limit? Are some individuals within this group likely to experience substantially higher exposure than its average member? How does an intervention influence the distribution of exposures? These questions were addressed using Bayesian models, simultaneously accounting for left, right, and interval-censored data with multiple censoring points. The models are estimated using the JAGS Gibbs sampler called through the R statistical package.
Results: The Expostats toolkit is freely available from www.expostats.ca as four tools accessible through a Web application, an offline standalone application or algorithms. The tools include a variety

of calculations and graphical outputs useful according to current practices in analysis and interpretation of exposure measurements collected by occupational hygienists. Tool1 and its simplified version Tool1 Express focus on inferences from data from a similarly exposed group. Tool2 evaluates within- and between-worker components of variability, as well as the probability that an individual worker might be overexposed. Tool3 compares exposure data across groups, e.g. evaluates the effect of an intervention. Uncertainty management includes the calculation of credible intervals and produces probabilistic statements about the exposure metrics (e.g. probability that over 5% of exposures are above a limit).

Discussion: Expostats is the first freely available toolkit that leverages the flexibility of Bayesian analysis to perform an extensive list of calculations recommended in several international guidelines on the practice of occupational hygiene.

Keywords: exceedance fraction; 95th percentile; lognormal distribution; occupational exposure compliance; risk assessment; sampling strategies

Introduction

It has been long understood that levels of workplace exposure can vary considerably across both location and time. Statistical methods developed to address the challenge posed by such variability started to appear in the scientific literature in the 1960s (Kerr, 1962; Roach, 1966; Breslin *et al.*, 1967), and coalesced into guidelines published by several institutions that inform the practice of occupational hygiene in various countries (Leidel *et al.*, 1977; Hawkins *et al.*, 1991; BOHS, 1993; BOHS-NVvA, 2011; INRS, 2018). The European community recently updated recommendations (CEN, 2018). These methods are all based on the assumption that environmental variability is adequately modelled by a lognormal probability distribution. This model assumes that, for a given population (e.g. stainless steel welders in a parts manufacturing facility), the ensemble of exposure levels experienced by workers over a period of relatively stable work conditions follows a lognormal distribution. Despite this well-established theoretical framework, surprisingly few practical lognormal statistics tools have been available to support practitioners (Waters *et al.*, 2015).

Bayesian analysis has been proposed for use in the interpretation of workplace measurements, because it permits quantitative integration of expert judgement into the assessment (Ramachandran and Vincent, 1999; Hewett *et al.*, 2006; Sottas *et al.*, 2009; Banerjee *et al.*, 2014) as well as information gained from pre-existing measurements (Jones and Burstyn, 2017; Quick *et al.*, 2017). In addition, Bayesian analysis provides direct probabilistic statements about the questions and parameters of interest, more easily conveyed to non-statisticians than traditional hypothesis tests or confidence intervals (CI). Finally Bayesian analysis can account for non-detects in a theoretically optimal manner, a longstanding challenge in occupational hygiene (Wild *et al.*, 1996; Mcbride *et al.*, 2007; Ogden, 2010; McNally *et al.*, 2014; Huynh *et al.*, 2016). While these characteristics render this approach promising for the interpretation of occupational exposure data, it currently remains mostly unavailable to practitioners in our field.

A set of computer applications were developed, assembled into the Expostats toolkit (freely available from www.expostats.ca in the form of algorithms as well as online and standalone applications), aimed at supporting practitioners in calculating inferences using Bayesian models that provide: (i) estimates for all metrics currently recommended in national and international guidelines; (ii) inferences in the form of direct probabilistic statements amenable to easy risk communication; and (iii) inferences that account for left (<x), right (>x) and interval [(a–b)] censored data. This paper describes the Expostats toolkit's main features as well as a list of available analyses and Bayesian models used.

Methods

This section briefly presents: (i) a list of the lognormal exposure parameters recommended for decision making in industrial hygiene, which form the basis of the calculations included in the Expostats tools, (ii) the framework proposed in Expostats to manage uncertainty in exposure estimates, and (iii) the Bayesian models themselves.

Recommended exposure parameters for the interpretation of measurements

Because it is unrealistic to assess each worker, most risk assessment strategies rely on grouping workers performing similar tasks in the same environment into so-called homogenous (Mulhausen and Diamano, 1998) or similar exposure groups (SEG) (Ignacio and Bullock, 2008), in which only a fraction of the workers is sampled. The result of the assessment (e.g. estimation of exceedance fraction or 95th percentile) is then deemed valid for the whole group). While they could in theory be estimated for single workers' exposure distribution, the first two metrics described below are most often calculated for an SEG.

The term OEL is used here as a generic term to describe a threshold selected for air concentrations by the occupational health practitioner. It could be a regulatory OEL, a recommendation, or any arbitrary value judged relevant for the situation at hand. If no value is available, parameters of the lognormal distribution can still be estimated in Expostats using a value of 1.

Proportion of exposures exceeding a specific value (exceedance fraction)

This parameter is directly related to NIOSH's early proposal that less than 5% of exposures should exceed the occupational exposure limit (OEL) (Leidel *et al.*, 1975, 1977). Applied to shift-long exposures, the exposure distribution of interest would comprise all time-weighted-averaged (TWA) exposures occurring during a period of stable conditions. One would then collect a random sample from this exposure distribution and estimate the proportion of days expected to be associated with exposure over the OEL. This proportion is referred to as the exceedance fraction. The calculation of exceedance fraction is recommended by the Institut national de recherche et de sécurité (INRS) in France, the British and Dutch occupational hygiene societies (BOHS/NVvA) and the European committee for normalization (CEN), and forms the basis of the current French regulation (République Française, 2009; BOHS-NVvA, 2011; CEN, 2018; INRS, 2018). Comparing exceedance fraction to 5% is numerically equivalent to comparing the estimated 95th percentile to the OEL. The latter calculation is recommended in the current AIHA guidelines (Jahn *et al.*, 2015). Overexposure, i.e. a situation requiring intervention (or qualified as 'poorly controlled'), is hence defined as either exceedance fraction ≥ 5% or 95th percentile ≥ OEL.

Long-term arithmetic mean of the exposure distribution

Toxicokinetic models show that the arithmetic mean (AM) of the long-term distribution of exposure levels is a more appropriate risk metric for evaluating cumulative damage from exposure to chronic toxicants than estimates from the upper tail of the distribution (Rappaport, 1991). Within this framework, one would compare the AM of the exposure distribution with the OEL: overexposure is defined here as AM ≥ OEL. There has been some debate about the use of this parameter, as it is claimed that comparing AM to the OEL is less conservative than comparing exceedance fraction to 5% (Lyles and Kupper, 1996; Hewett, 1997; Tornero-Velez *et al.*, 1997). The current guidelines from the AIHA recommend this approach in the rare cases where the exposure limit has explicitly been defined as a long-term cumulative dose index ('LTA-OEL, Long-term average OEL') (Jahn *et al.*, 2015).

Probability of individual overexposure: probability that a random worker within the group would be overexposed despite an acceptable exposure distribution for the group as a whole

Following seminal work by Kromhout, Rappaport, and Symanski in the late 1990s (Kromhout *et al.*, 1993; Rappaport *et al.*, 1993; Symanski *et al.*, 2006), it was recognized that the practice of grouping workers into SEGs could result in underestimation of risk for some members of the group. Strategies to integrate between-worker variations in groups into decision making were then proposed (Rappaport *et al.*, 1995; Lyles et al., 1997a,b). They involved estimating the probability that a random worker within the group would be overexposed, an estimate that we call the probability of individual overexposure. The British-Dutch guidance defines overexposure for a worker as having their individual 95th percentile above the OEL (or equivalently, their individual exceedance fraction above 5%), whereas the earlier proposal by Rappaport *et al.* defined overexposure for a worker as having their individual AM above the OEL (Rappaport *et al.*, 1995; BOHS-NVvA, 2011).

Uncertainty management framework

Appraisal of uncertainty is an essential component of decision making in industrial hygiene (Waters *et al.*, 2015). For example, even if the point estimate of exceedance fraction for a group of workers is below 5%, how sure can we be that the true value is indeed <5%? This has traditionally been tackled through the use of confidence intervals and hypothesis tests. As an illustration, formulas to estimate a 95% upper confidence limit on the 95th percentile were proposed (Selvin *et al.*, 1987). If this value is lower than the OEL, the practical (though not formally correct) interpretation is that one can be 95% certain that the true underlying 95th percentile is indeed < OEL.

An alternative and more direct statement about uncertainty could be made. For example, calculating the probability that the true 95th percentile is below the OEL, which should be high (>95% in the above example), or, conversely, the probability that the true 95th percentile is above the OEL, which should be low (<5% in the example above). This seems both informative, and easy to convey to workers or employers as it provides a direct answer to the question 'what are the chances that exposure is too high?'.

The Expostats toolset implements Bayesian analysis that leads to direct statements about the degree of uncertainty in conclusions that can be drawn from data. It relies on three steps leading to a decision whether exposure is adequately controlled.

Step 1

Define overexposure, i.e. which characteristic of the exposure distribution corresponds to an unacceptable situation. Several criteria can be investigated:

- Exceedance fraction ≥ 5%
- 95th percentile (P95) ≥ OEL
- AM ≥ OEL

Step 2

Analyse the observed data and draw inferences from it. Instead of parameter point estimates with confidence intervals or hypothesis test, this is done through the estimation of the probability that the overexposure criterion is met, e.g. what is the probability that true 95th percentile is ≥ OEL given the data. This quantity is hereafter called the overexposure risk (see Table 1 for a glossary of terms used in this manuscript). This quantity can be used as a direct input for exposure management: is overexposure risk low enough that it is possible to consider exposure well controlled, or is it high enough that some action should be triggered (e.g. consider implementing exposure controls)?

Step 3 (optional)

Although overexposure risk provides complete information about uncertainty, risk managers often prefer results in the form of a dichotomy: does this situation require an intervention, yes or no? The last step implemented in Expostats permits to reach such a conclusion. It requires setting a threshold for overexposure risk: the situation can then be declared either adequately controlled if overexposure risk is lower than the selected value, or poorly controlled otherwise. That threshold is called the overexposure risk threshold and should be set before the analysis. It is used to separate 'low enough' as mentioned in step 2 from 'not low enough'. The widely accepted value for this threshold is 5%: the overexposure risk should be lower than 5% to declare a situation acceptable (Jahn *et al.*, 2015). To illustrate the correspondence between this and more traditional statements of uncertainty, we shall use the example of P95 ≥ OEL as the overexposure criterion. An overexposure risk below 5% is equivalent to 'the chances that the true 95th percentile is above the OEL are below 5%', which, in turn, means that

Table 1. Glossary of terms.

Exceedance fraction	Proportion of exposures levels in the population of interest that are above the exposure limit. Equivalently, probability for a single random exposure value to be above the OEL
95th percentile	The 95th percentile of a distribution is defined as the value below which lies 95% of the distribution
Overexposure	Characteristic of an exposure distribution that is unacceptable, i.e. which would trigger preventive action.
Exceedance threshold	Proportion of exposure levels over the OEL used as threshold to define overexposure (traditionally 5%)
Critical percentile	Percentile of the exposure distribution that will be compared to the OEL to evaluate overexposure (traditionally 95th percentile)
Overexposure risk	Probability that the criteria used to defined overexposure is met (e.g. 95th percentile ≥ 5%). Practically: probability of an unacceptable exposure situation
Overexposure risk threshold	Maximum allowable overexposure risk. This value, chosen a priori by the user, is used to create a dichotomy between 'adequately controlled' and 'poorly controlled' based on the overexposure risk. A traditional value used in the field of statistics would be 5%. The French OEL compliance definition is equivalent to an overexposure risk threshold of 30%
Probability of Individual overexposure	Probability that a random worker within a group would have their individual exposure distribution corresponding to overexposure (e.g. probability that a random worker within a group has his individual 95th percentile above the OEL).Can also be stated as: Proportion of workers with their individual exposure distribution corresponding to overexposure
Credible interval	While not formally equivalent, Bayesian credible intervals are usually interpreted in a similar way as the more traditional confidence intervals
R ratio	R ratio has been defined by Rappaport *et al.* as the ratio of the 97.5% percentile of the distribution of workers' individual AM divided by the 2.5th percentile of the same distribution

we are at least 95% certain that the 95th percentile is below the OEL, which, finally, is equivalent to 'the 95% upper confidence limit (or the upper credible limit if Bayesian analysis was used) on the 95th percentile is below the OEL' (a more traditional statement). The current British-Dutch and European guidelines, as well as the French regulation, recommend comparing the 70% upper confidence limit on the exceedance fraction to 5%, which is equivalent to a 30% overexposure risk threshold with the overexposure criterion: Exceedance fraction ≥ 5%.

The uncertainty analysis involving calculation of between- and within-worker variability has an added layer of complexity: such analyses yield estimates of the probability of individual overexposure, i.e. the probability that a random worker would have an unacceptable individual exposure distribution. A threshold for this quantity was proposed: i.e. probability of individual overexposure should be < 20% (BOHS-NVvA, 2011). However, as this probability is estimated, it is uncertain. Therefore, instead of only comparing the point estimate of the probability of individual overexposure to 20%, one can evaluate the chances that the true value is ≥ 20%, i.e. the chances that an intervention would be required. Expostats calculates this probability (one can use the thresholds of 5% or 30% mentioned above to reach a decision).

Bayesian models

Expostats relies on two different Bayesian models: Model 1 estimates parameters from a single lognormal distribution. It is used to analyse data for SEGs. Model 2 is a hierarchical model used in cases when there are repeated measurements on some workers. It allows estimating two components of variance (i.e. within- and between-worker variability) and applying recommendations from the British-Dutch guidance and from Rappaport *et al.* (1995).

Both models allow for three types of censored data: left-censored values correspond to non-detects or values reported as <Limit of quantification (LOQ), right-censored values correspond to values reported as >X, interval-censored values correspond to data reported as between two limits [e.g. (a–b)].

Bayesian analysis requires the specification of prior distributions across all parameters. For model 1, this implies a prior for the geometric mean (GM) and the geometric standard deviation (GSD). For model 2, it implies priors for the group GM and both the within- and between-worker GSDs. For variability, priors for model 1 and 2 were derived from McNally *et al.* (McNally *et al.*, 2014). For GM, a standard weakly informative uniform distribution was used, as described, e.g. in Huynh *et al.* or Banerjee *et al.* (Banerjee *et al.*, 2014; Huynh *et al.*, 2016).

The mathematical statements of the models and detailed information and discussion of the selected priors are presented Supplementary Appendix A in the Supplementary Material (available at *Annals of Occupational Hygiene* online).

Software implementation

Expostats is a web-based application available freely from www.expostats.ca as an online set of applications or standalone package. The tools were programmed through the combined use of (i) JAGS statistical package, serving as a Bayesian engine; (ii) the R statistical package and *rjags* extension, used to process data, make calls to JAGS, and create numerical and graphical results; and (iii) the SHINY application, which serves as an interface between R and users online. All algorithms underpinning Expostats are available upon request.

Expostats is currently available in both English and French, and it is constructed to allow translation to other languages.

Results

The Expostats toolkit contains four sub-applications. Tool1 (full and simplified versions), includes calculations useful for the analysis of exposure data from a SEG. Tool2 includes calculations in Tool1 plus accounts for within and between-worker variability, and evaluates the probability that some workers within the group are overexposed. Tool3 includes calculations in Tool1 and permits evaluation of the association of exposure levels with a single categorical variable.

The following sections provide an overview of each tool, along with example of numerical and graphical outputs.

SEG analysis (Tool1 and Tool1 express)

Tool1 would be the main tool for routine interpretation of industrial hygiene data where a set of measurements is available to estimate the exposure distribution for a SEG. For a given dataset and associated OEL, the first tab in Tool1 provides descriptive statistics, and a quantile-quantile and box and whisker plot to identify outliers and assess the adequacy of the lognormal model. The next three tabs present results based on the preferred exposure metric (exceedance fraction, exposure percentile, or AM, Table 2). In each case, users are provided with point estimates and credible intervals for GM, GSD, and the metric of interest. In

Table 2. Technical characteristics of the tools included in Expostats.

	Tool 1	Tool 2	Tool 3
Model			
Estimation of a single lognormal exposure distribution	✓	✓	✓
Estimation of a simple hierarchical model, where each worker has his own exposure distribution		✓	
Estimation of n lognormal exposure distributions (one for each category of a nominal variable)			✓
Distributional parameter estimates (point estimate and CrI) for a group | individual | category			
GM	✓| - | -	✓|✓| -	✓| - |✓
GSD	✓| - | -	✓|✓| -	✓| - |✓
Exceedance fraction of the OEL	✓| - | -	✓|✓| -	✓| - |✓
Percentile of the exposure distribution (i.e., critical percentile, default 95%))	✓| - | -	✓|✓| -	✓| - |✓
AM	✓| - | -	✓|✓| -	✓| - |✓
Components of variance estimated (point estimate and CrI)			
Parameters quantifying between-worker differences			
Within-worker GSD		✓	
Between-worker GSD		✓	
Within-worker correlation coefficient		✓	
R ratio		✓	
Parameters quantifying the possibility that some workers are overexposed			
Proportion of workers with their individual exceedance fraction above the exceedance threshold		✓	
Proportion of workers with their individual critical percentile > OEL		✓	
Proportion of workers with their individual AM > OEL		✓	
Decision on Exposure Acceptability			
Probability of overexposure for the following criteria for a **group | individual | category**			
1) Exceedance fraction ≥ exceedance threshold	✓| - | -	✓|✓| -	✓| - |✓
2) Critical percentile ≥ OEL	✓| - | -	✓|✓| -	✓| - |✓
3) AM ≥ OEL	✓| - | -	✓|✓| -	✓| - |✓
Probability of individual overexposure for the following criteria **accounting for between-worker variability**:			
1) Exceedance fraction ≥ exceedance threshold		✓	
2) Critical percentile ≥ OEL		✓	
3) AM ≥ OEL		✓	
Chances that the probability of individual overexposure is above a selected criteria (e.g. 20% for BOHS-NVvA)		✓	
Comparative analysis for two categories			
For any two categories of a selected variable:			
Ratio GM_1/GM_2 and CrI			✓
Ratio GSD_1/GSD_2 and CrI			✓
Ratio critical percentile_1/critical percentile_2 and CrI			✓
Ratio AM_1/AM_2 and CrI			✓
Exceedance fraction difference (exceedance fraction_1-exceedance fraction_2) and CrI			✓
For any of the above ratios or differences, the probability that they are above or below a pre-specified value is also calculated. E.g., probability that $GM_1/GM_2 > 2$			✓
Graphical evaluation of the data for a group | individual | category			
Descriptive plots			
Quantile-quantile (Q-Q) plot	✓| - | -	✓|✓| -	✓| - |✓
Box and whisker plot	✓| - | -	✓|✓| -	✓| - |✓
Plots assessing risk based on overexposure criteria (exceedance fraction, 95th percentile, long-term AM)			
Riskmeter - the probability that the exposure is too high (overexposure risk)	✓| - | -	✓|✓| -	✓| - |✓
Exceedance plot - the expected proportion of exposures that would exceed OEL	✓| - | -	✓| - | -	- | - | -

Table 2. Continued

	Tool 1	Tool 2	Tool 3
Sequential plot – Distribution of exposure levels assuming many measurements could have been taken	✓| - | -	✓|✓| -	✓| - |✓
Density plot - estimated underlying distribution of exposures	✓| - | -	✓| - | -	- | - | -
AIHA risk band plot –distribution of the uncertainty around the overexposure criteria across set categories	✓| - | -	✓|✓| -	✓| - |✓
Comparative box and whisker plots			
Exposure distribution of individual workers	- | - | -	- |✓| -	- | - | -
Exposure distribution by categories in a selected variable	- | - | -	- | - | -	- | - |✓
Exposure distribution by only two categories of a variable of interest	- | - | -	- | - | -	- | - |✓
Comparative plot of overexposure risk across categories			✓
Comparative AIHA risk band plot across categories			✓
Risk band plot of the distribution of uncertainty around probability of individual non-compliance		✓	
Customizable parameters			
Probability for the credible intervals (1–99.9%, default 90)	✓	✓	✓
Exceedance threshold (0.1–99.9, default 5%)	✓	✓	✓
Critical percentile (0.1–99.9, default 95%)	✓	✓	✓
Overexposure risk threshold (0.1–99%, default 30)	✓	✓	✓
Threshold for the probability of Individual overexposure (0.1–99%, default 20)		✓	
Coverage of the R ratio (1–99%, default 80)		✓	
Threshold for the within-worker correlation coefficient (0.01–0.99, default 0.2)		✓	
Expected change in GM, GSD, Critical percentile, Exceedance fraction (any value)			✓

addition to the point estimates and CrIs, Tool1 presents the overexposure risk and the final decision (adequately versus poorly controlled exposure situation) based on a default (albeit customizable) overexposure risk threshold of 30% in all three tabs.

Tool1 also provides graphs to help convey information useful for risk communication about the exposure distribution (Fig. 1). The exceedance plot shows the proportion of cases that would correspond to exposure>OEL imagining a large number of relevant exposure periods could be monitored (This could be many workers for many days, one worker for many days, many workers for many 15min periods, etc.…). The sequential plot shows the estimated distribution of exposure levels, also illustrating the case where many measurements could have been made. To communicate overexposure risk Tool1 presents a risk gauge, where a needle indicates the probability value within the related risk category (Fig. 2). Finally, the user is also shown the AIHA exposure band categories with their associated estimated probability, as initially presented by Hewett *et al.* (2006) and advocated by Banerjee *et al.* (2014). Probabilities are calculated based on the posterior distribution of the metric of interest (95th percentile, AM, or exceedance fraction). Tool1 was designed as comprehensive and highly customizable. A simplified version (Tool1 Express) was also created aiming to present succinct information that would be used by most practitioners.

Example

Let's assume the following dataset represents a random sample from the exposure distribution: 28.9, 19.4, <5.5, 89.0, 26.4, 56.1, with an OEL of 150, where the < symbol indicates a left-censored observation. The Bayesian calculations provide a point estimate and 90% credible interval for the exceedance fraction of 5.4% (0.4–25.7). Let's assume the assessor chooses an exceedance threshold of 10%: i.e. overexposure is defined as exceedance fraction ≥ 10%. Despite a point estimate (5.4%) below this threshold, the credible interval suggests the true underlying exceedance fraction could be as low as 0.4% and as high as 25.7%. A typical conclusion from this would be 'It is impossible to conclude with statistical significance that exceedance fraction is below 10%, or that it is above 10%'. Tool1 also indicates overexposure risk: 29%. The full interpretation is as follows: the point estimate for the exceedance fraction is 5.4%, with a 29% probability that the true underlying value is ≥10%. The assessor can then decide whether this probability is low enough to conclude that exposure is adequately controlled. The uncertainty management final conclusion, given the

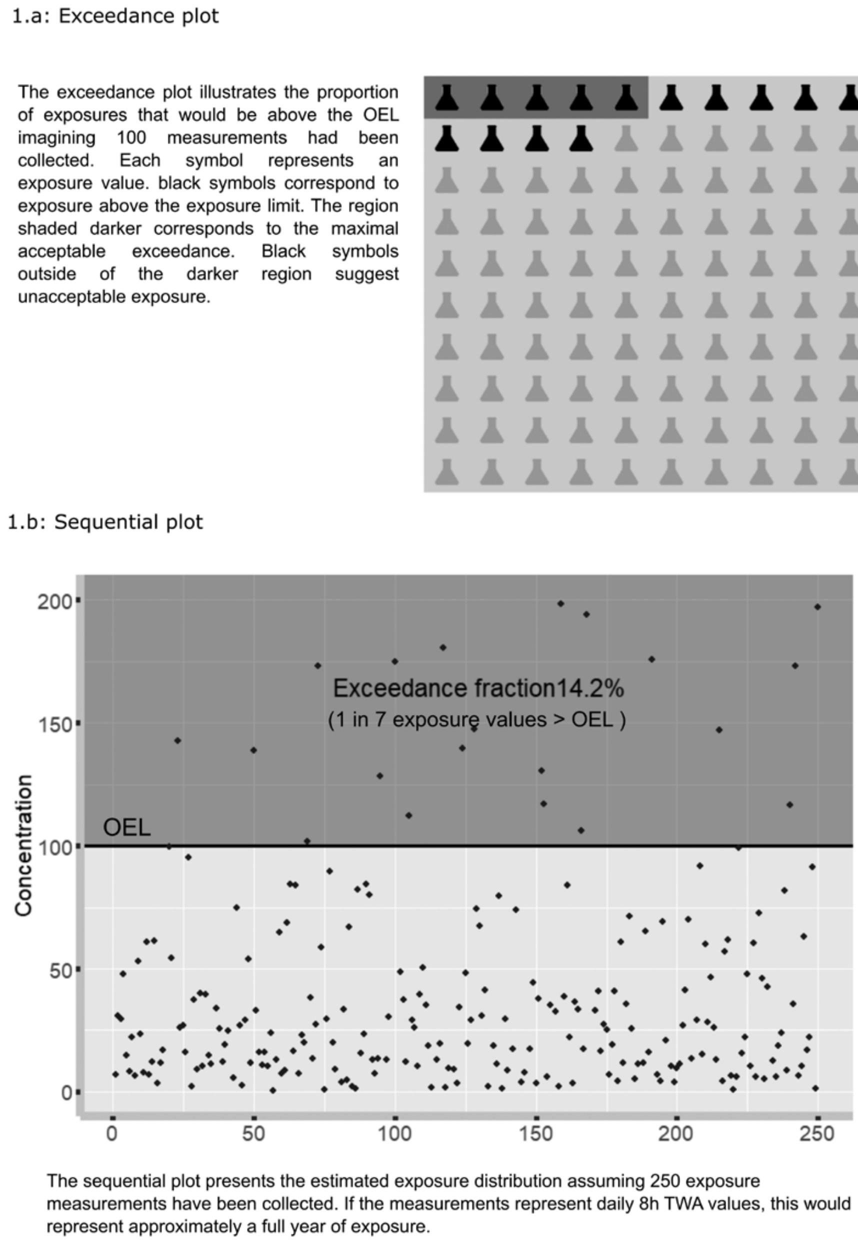

Figure 1. Selected graphs illustrating the exposure distribution in Tool1.

selected 30% overexposure risk threshold, declares that exposure is adequately controlled given overexposure risk is <30% in this situation.

Between-worker differences within a SEG (Tool2)

Tool2's main function is to estimate between and within-worker variance from a set of measurements with repeats for some workers within a SEG (Table 2). The Bayesian model used is similar in spirit to a traditional one-way random effect analysis of variance (ANOVA) model initially described in Kromhout *et al.* (1993) and Rappaport *et al.* (1993) and subsequently adopted in the British-Dutch and AIHA guidelines. Three outputs aim at quantifying the amount of variation between workers within the group. First, Tool2 provides point estimates

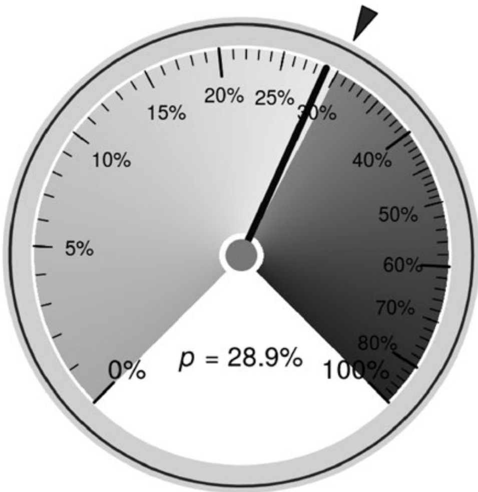

Figure 2. Illustration of the riskmeter plot.

and associated uncertainty for the between- and within-worker GSDs. Second, Tool2 estimates the within-worker correlation coefficient, which is the proportion of the total variance represented by the between-worker variance (the higher the between-worker variance, the higher the correlation within measurements from the same worker). Third, Tool2 provides the R ratio, a term first described as the fold range containing the middle 95% of the distribution of worker specific AMs (Rappaport *et al.*, 1993) (e.g. R=2 means that the ratio of the 97.5th to 2.5th percentiles of the distribution of worker specific AMs is 2). In simpler terms, R is approximately the ratio of the AM of the most exposed worker to the AM of the least exposed worker. Tool2 provides an estimate of R for any proportion of the distribution. It is noteworthy that R has the same value whether one is interested in the GM, or any percentile of the exposure distribution instead of the AM. In terms of risk, Tool2 also provides the probability that a random worker would have his own exposure distribution corresponding to overexposure according to several criteria (Table 2), as well as the associated uncertainty. Finally, Tool2 provides individual worker statistics based on the ANOVA model (i.e. the model assumes all worker specific distributions within the group share the same variability).

Example

As an example, selected results from the analysis of a dataset described in the British-Dutch guidance are presented regarding 30 cotton dust measurements comprised of six repeated observations for each of five workers (BOHS-NVvA, 2011). The between- and within-worker GSDs were, respectively, 1.3 (90% CrI 1.1–1.7) and 1.6 (90% CrI 1.5–1.8). Regarding criteria of homogeneity, the within-worker correlation coefficient point estimate was 0.19 (90% CrI 0.03–0.57). The R ratio (80% of the distribution, corresponding approximately to the ratio of the most to the least exposed worker in a group of 10) was 1.5 (90% CrI 1.2–2.5), showing moderate between-worker variability. In terms of risk of individual overexposure, the estimated probability of a random worker having his own 95th percentile above the OEL of 0.8 mg/m^3 was 74% (90% CrI 29–100), with 98% chances that the true underlying value is above 20%, the criterion set by the British-Dutch guidance. Expressed another way: the probability of individual overexposure is estimated at 74%, with 98% chances that the true value is above the criterion of 20%.

Figure 3 illustrates the differences between workers, showing actual measurements as well as the estimated underlying distributions.

Differences between groups (Tool3)

Tool3 was designed with the objective of providing an alternative to the traditional one-way ANOVA when one is interested in studying differences in exposure between groups (Table 2). The typical output of such analysis is a hypothesis test answering the question: are mean levels equal across groups? Such a question is of limited relevance. Hence, a true difference of 1% between means could be labelled as 'significant' in large sample sizes despite not being relevant in terms of prevention. Conversely, a large and important difference may be labelled as 'non-significant' by a hypothesis test owing to lack of statistical power from smaller sample sizes, an important error of omission. Moreover, a traditional ANOVA analysis on log-transformed exposure levels would not allow estimating differences between categories (including uncertainty in these differences) in terms of the relevant exposure metrics (e.g. exceedance fraction). To circumvent these issues, Tool3 allows estimating the probability that the difference between two groups is greater or smaller than a quantity judged relevant by the assessor (e.g. probability that the GM was reduced by at least a factor of 2, or that exceedance fraction decreased by 10%). These calculations are performed through the simultaneous application of the Bayesian model 1 to each group.

Example

We provide two examples for Tool3, using a dataset of formaldehyde measurements in the wood panel industry in Quebec, Canada, with a hypothetical OEL of 0.5 ppm (Lavoué *et al.*, 2005). The first example illustrates how Tool3 can help assess differences between groups: We compared the particle board (PB, n = 118

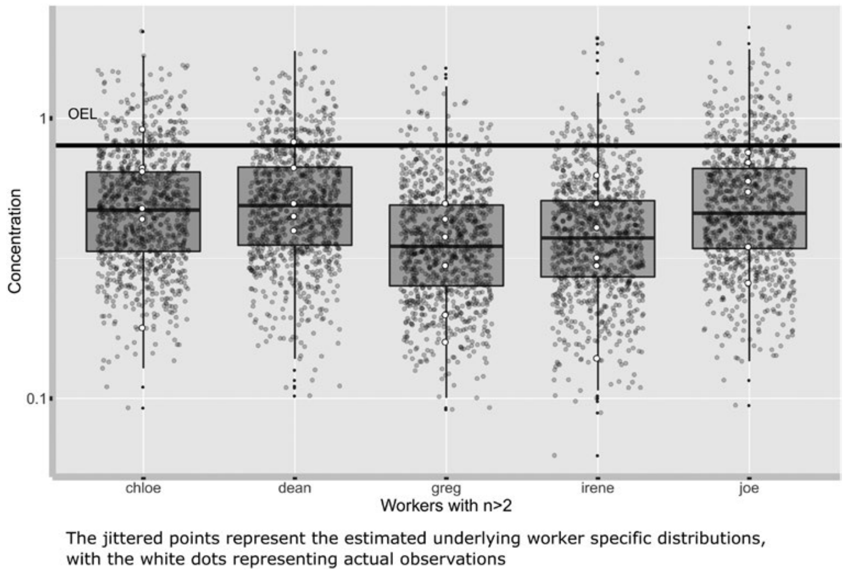

Figure 3. Boxplot illustrating worker differences in Tool2.

measurements) and oriented strand board (OSB, n = 97) categories of the process variable. The ratio of GMs by process (PB/OSB) is estimated to be 4.1 (90% CrI 3.4–4.9), with a 93% probability that the true underlying ratio is greater than 3.5, a hypothetical threshold of interest. The absolute difference in the exceedance fraction (PB-OSB) is estimated to be 13 (90% CrI 9.3 – 18). The ratio of GSDs (PB/OSB) is estimated to be 0.99 (90% CrI 0.85–1.10), showing very similar variability between the processes despite very different mean levels. The second example (Fig. 4) shows the capacity of Tool3 to graphically provide a comparative picture of overexposure risk across several groups. Comparing exposure levels across four different job titles in the wood panel plants using 95th percentile > OEL as the overexposure criterion, we find that the highest exposed job (job4) has a 99.98% overexposure risk, with the >OEL riskband filling the entire corresponding line in the comparative exposure band plot. In contrast to job4, job1 has only a 0.05% overexposure risk (as shown on the comparative overexposure risk plot). The corresponding risk band line provides more information, showing a 90% probability that the true 95th percentile is in the (10–50%)*OEL category, with the remaining 10% in the (1–10%)*OEL category.

Discussion

Following major methodological developments from the 1970s through to the end of the 1990s, the theory underpinning the interpretation of measurement data for decision making in industrial hygiene seems to have somewhat stabilized for the last two decades or so. Despite the availability of the theory in the scientific literature, and even of simplified syntheses in practical guidelines, few tools have been made available to implement the rather involved required methodology. Arguably the best known available tool is the IHSTAT (https://www.aiha.org/get-involved/VolunteerGroups/Pages/Exposure-Assessment-Strategies-Committee.aspx) spreadsheet application provided free of charge from the AIHA website. Other free and paid software appearing over the years, and still currently available, include SPEED (http://www.iras.uu.nl/speed/), ALTREX (http://www.inrs.fr/media.html?refINRS=outil13), BWStat (https://www.bsoh.be/?q=en/node/89), HYGINIST (http://www.tsac.nl/hyginist.html), IH Data analyst (https://www.easinc.co/ihda-software/), and ART (https://www.advancedreachtool.com/). While each, with various degrees of refinement and complexity, represents a significant contribution to helping practitioners perform lognormal analyses, the added contribution of the Expostats toolkit is mainly 3-fold.

First, Expostats contains the most comprehensive list of lognormal calculations that might be deemed relevant to the interpretation of industrial hygiene data based on current best practice. This includes analysis of data from a SEG, assessment of differences between workers, and comparison across several groups, all including estimation of any percentile of the distribution, exceedance fraction of the OEL and the AM. Indeed, to our knowledge Expostats currently represents the most

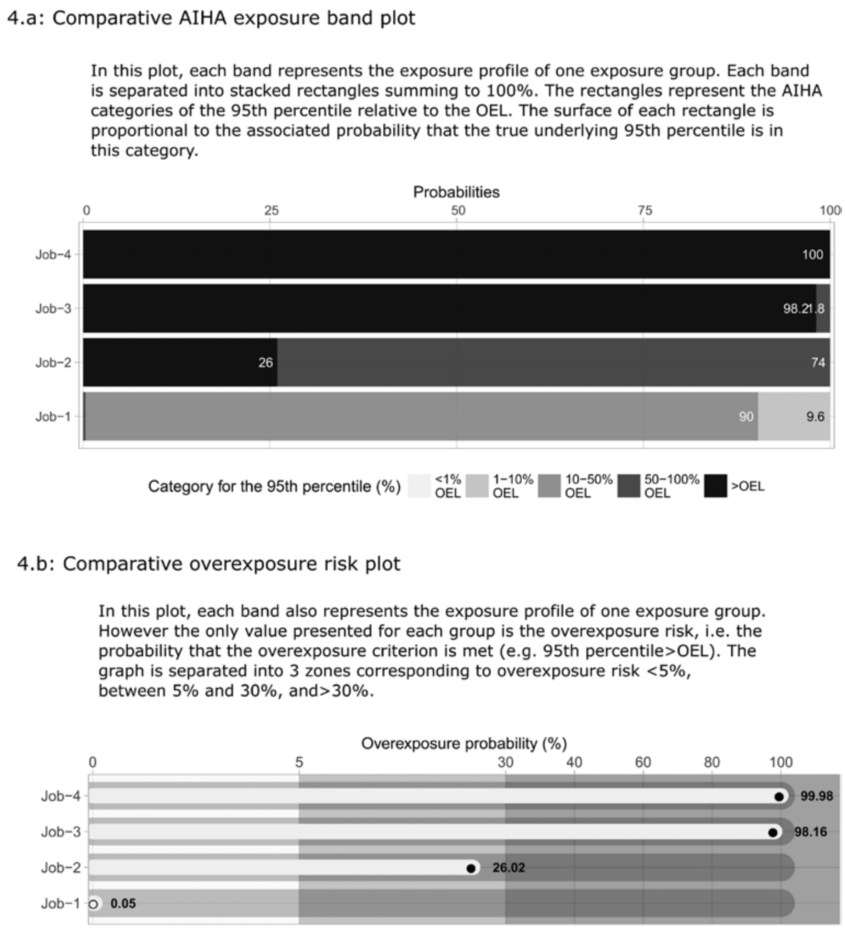

Figure 4. Comparative exposure band plot and overexposure risk plot in Tool3.

comprehensive lognormal data analysis toolset publicly available. In addition Expostats incorporates a high level of flexibility in parameter selection (Table 2), with several traditionally fixed parameters becoming customizable. As a consequence, the extensive number of numerical and graphical outputs might seem overwhelming for practitioners aiming to quickly obtain a diagnosis on a particular exposure situation. The creation of the express version of Tool1, focused on a restricted number of essentials, addresses this concern. The open source nature of Expostats, should also allow for the creation of applications tailored to any specific need/complexity level.

Second, the treatment of non-detects has long been a thorn in the side of exposure assessors, with editorials published in the field of industrial hygiene advocating for changes in practice and the development of more rigorous approaches (Helsel, 2010; Ogden, 2010). Significant progress has been reported recently (Krishnamoorthy *et al.*, 2009; Flynn, 2010; Ganser and Hewett, 2010), with several simulation studies comparing approaches (Hewett and Ganser, 2007; Huynh *et al.*, 2014, 2016). A recent comparison suggested Bayesian methods are optimally suited for this challenge, since they allow multiple censoring points, and accurately estimate the inherent uncertainty when data values are known only up to an interval (Huynh *et al.*, 2016). Expostats uses the same Bayesian approach as described early in Wild *et al.* (1996), and more recently in Huynh *et al.* (2014) and McNally *et al.* (2014), applied to all models, and extended from only left-censored data to interval- and right-censored data.

Third, objective management of uncertainty is central in all aspects of the Expostats tools. As advocated in a recent review by Waters *et al.* (2015), we propose a probabilistic framework for the interpretation of exposure measurements, where the conclusion relies on the probability that overexposure criteria are met rather than on point estimates and confidence intervals of exposure metrics. Consistent with the principles of risk characterization (U.S. EPA, 2000), Expostats uses simple probabilistic statements to improve the clarity of risk communication in the form 'there is an XX% chance that our overexposure criterion is met', to remove statistical jargon, use a language which is accessible to all stakeholders, but still describes the quantitative risk clearly and accurately.

Some limits of the present work should be acknowledged. First, the Bayesian models in Expostats do not make use of a powerful feature of Bayesian analysis, namely the use of informed prior information, where external knowledge is combined with the observed data to calculate the posterior distributions. This feature is used both in the ART and IH Data analyst tools. As discussed in detail in Supplementary Appendix A in the Supplementary Material (available at *Annals of Occupational Hygiene* online), several forms and sources of prior information can be used. While future iterations of the Expostat toolset may include this possibility, we opted to provide first a Bayesian alternative to current traditional frequentist tools, where information comes mainly from actual observations, with no prior information on the location of the lognormal distribution, and limited prior information about variability based on a historical database covering multiple workplaces. Second, Expostats is limited to the lognormal model. While this model is probably adequate in most workplace exposure situations, the more traditional normal distribution, other distributions (e.g. Poisson) or non-parametric procedures can be useful and are currently not included. Finally, it is worth mentioning that this paper does not represent a new set of recommendations for strategies in occupational risk assessment. Readers interested in discussions about the respective merits of the various metrics calculated by Expostats (e.g. exceedance versus AM) should consult the cited references. Expostats is merely a lognormal calculation toolset, albeit an advanced one, useful when the exposure assessor has decided to rely on probabilistic sampling and comparison with an OEL. The soundness of the conclusions drawn from it rely on the suitability of the lognormal model for the situation at hand, on the representativeness and quality of the samples collected, as well as the adequacy of priors and 'correctness' of the assumed risk model and OEL.

Conclusion

The Expostats toolset, in a way closer to knowledge translation than to pure novel scientific development, is meant to allow practitioners to utilize many theoretical developments formerly only available to academics. Interested readers are encouraged to participate in the ongoing evolution of this collaborative project by providing feedback and suggestions to the authors.

Supplementary Data

Supplementary data are available at *Annals of Work Exposures and Health* online.

Acknowledgements

Several colleagues supported the development of Expostats over the years, through providing feedback and suggestions for the design of the tools and calculations, including an international meeting held in Montreal in 2016. Occupational hygienists include Daniel Drolet, Robert Fraser, Pierrot Pépin, Michel Gérin, Denis Bégin, Paul Bozek, Martie Van Tongeren, Martine Chouvet; programmers include François Lemay, Patrick Belisle, Delphine Bosson-Rieutort, and Daniel Margulius. Many thanks to Igor Burstyn for his comments on the manuscript and methodological and editorial advice. This project was partially funded through grant number 2014-0073 from Institut de recherche Robert Sauvé en santé et en sécurité du travail.

Conflict Of Interest Declaration

Funding for this project was provided by J.L.'s personal research funds and partially by IRSST (grant 2014-0073). The authors declare no conflict of interest relating to the material presented in this Article. Its contents, including any opinions and/or conclusions expressed, are solely those of the authors.

References

Banerjee S, Ramachandran G, Vadali M *et al*. (2014) Bayesian hierarchical framework for occupational hygiene decision making. *Ann Occup Hyg*; **58**: 1079–93.

BOHS (1993) *British occupational hygiene society technical guide No. 11: sampling strategies for airborne contaminants in the workplace*. Leeds, UK: H and H Scientific Consultants Ltd.

BOHS-NVvA (2011) *Testing compliance with occupational exposure limits for airborne substances*. Available at: https://www.arbeidshygiene.nl/-uploads/files/insite/2011-12-bohs-nvva-sampling-strategy-guidance.pdf. Accessed 26 November 2018.

Breslin AJ, Glauberman H, George AC *et al*. (1967) The accuracy of dust exposure estimates obtained from conventional air sampling. *Am Ind Hyg Assoc J*; **28**: 56–61.

CEN (2018) *BS-EN 689:2018 workplace exposure. measurement of exposure by inhalation to chemical agents. Strategy for testing compliance with occupational exposure limit values*. Brussels, Belgium: British Standards Institution.

Flynn MR. (2010) Analysis of censored exposure data by constrained maximization of the Shapiro-Wilk W statistic. *Ann Occup Hyg*; **54**: 263–71.

Ganser GH, Hewett P. (2010) An accurate substitution method for analyzing censored data. *J Occup Environ Hyg*; **7**: 233–44.

Hawkins NC, Norwood SK, Rock JC (1991) *A strategy for occupationnal exposure assessment*. Fairfax, VA: American Industrial Hygiene Association.

Helsel D. (2010) Much ado about next to nothing: incorporating nondetects in science. *Ann Occup Hyg*; **54**: 257–62.

Hewett P. (1997) Mean testing: I. advantages and disadvantages. *Appl Occup Environ Hyg* **12**: 339–46.

Hewett P, Ganser GH. (2007) A comparison of several methods for analyzing censored data. *Ann Occup Hyg*; **51**: 611–32.

Hewett P, Logan P, Mulhausen J *et al*. (2006) Rating exposure control using Bayesian decision analysis. *J Occup Environ Hyg*; **3**: 568–81.

Huynh T, Quick H, Ramachandran G *et al*. (2016) A Comparison of the β-substitution method and a Bayesian method for analyzing left-censored data. *Ann Occup Hyg*; **60**: 56–73.

Huynh T, Ramachandran G, Banerjee S *et al*. (2014) Comparison of methods for analyzing left-censored occupational exposure data. *Ann Occup Hyg*; **58**: 1126–42.

Ignacio JS, Bullock WH. (2008) *A strategy for assessing and managing occupational exposures*. 3rd edn. Fairfax, VA: AIHA Press.

INRS (2018) *Statistical interpretation of measurement results (in French)*. Paris: Institut national de recherche et de sécurité.

Jahn SD, Bullock C, Ignacio JS. (2015) *A strategy for assessing and managing occupational exposures*. 4th edn. Fairfax, VA: AIHA Press.

Jones RM, Burstyn I. (2017) Bayesian analysis of occupational exposure data with conjugate priors. *Ann Work Expo Health*; **61**: 504–14.

Kerr GW. (1962) Use of statistical methodology in environmental monitoring. *Am Ind Hyg Assoc J*; **23**: 75–82.

Krishnamoorthy K, Mallick A, Mathew T. (2009) Model-based imputation approach for data analysis in the presence of non-detects. *Ann Occup Hyg*; **53**: 249–63.

Kromhout H, Symanski E, Rappaport SM. (1993) A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. *Ann Occup Hyg*; **37**: 253–70.

Lavoué J, Beaudry C, Goyer N *et al*. (2005) Investigation of determinants of past and current exposures to formaldehyde in the reconstituted wood panel industry in Quebec. *Ann Occup Hyg*; **49**: 587–602.

Leidel N, Busch K, Crouse WE. (1975) *NIOSH Technical Information: Exposure measurement action level and occupational environmental variability (NIOSH 76–131)*. Cincinnati, OH: US Department of Health, Education, and Welfare.

Leidel NA, Busch KA, Lynch CF (1977) *NIOSH occupational exposure sampling strategy manual*. Cincinnati, OH: US Department of Health, Education, and Welfare.

Lyles RH, Kupper LL. (1996) On strategies for comparing occupational exposure data to limits. *Am Ind Hyg Assoc J*; **57**: 6–15.

Lyles RH, Kupper LL, Rappaport SM. (1997a) A lognormal distribution-based exposure assessment method for unbalanced data. *Ann Occup Hyg*; **41**: 63–76.

Lyles RH, Kupper LL, Rappaport SM. (1997b) Assessing regulatory compliance of occupational exposures via the balanced one-way random effects ANOVA Model. *J Agric Biol Environ Stat*; **2**: 64–86.

Mcbride S, Williams R, Creason J. (2007) Bayesian hierarchical modeling of personal exposure to particulate matter. *Atmos Environ*; **41**: 6143–55.

McNally K, Warren N, Fransman W *et al*. (2014) Advanced REACH Tool: a Bayesian model for occupational exposure assessment. *Ann Occup Hyg*; **58**: 551–65.

Mulhausen JR, Diamano J. (1998) *A strategy for assessing and managing occupational exposures*. Fairfax, VA: AIHA Press.

Ogden TL. (2010) Handling results below the level of detection. *Ann Occup Hyg*; **54**: 255–6.

Quick H, Huynh T, Ramachandran G. (2017) A method for constructing Informative priors for Bayesian modeling of occupational hygiene data. *Ann Work Expo Health*; **61**: 67–75.

Ramachandran G, Vincent JH. (1999) A Bayesian approach to retrospective exposure assessment. *Appl Occup Environ Hyg*; **14**: 547–57.

Rappaport SM. (1991) Assessment of long-term exposures to toxic substances in air. *Ann Occup Hyg*; **35**: 61–121.

Rappaport SM, Kromhout H, Symanski E. (1993) Variation of exposure between workers in homogeneous exposure groups. *Am Ind Hyg Assoc J*; **54**: 654–62.

Rappaport SM, Lyles RH, Kupper LL. (1995) An exposure-assessments strategy accounting for within- and between-worker sources of variability. *Ann Occup Hyg*; **39**: 469–95.

République Française (2009) Arrêté du 15 décembre 2009 relatif aux contrôles techniques des valeurs limites d'exposition professionnelle sur les lieux de travail et aux conditions d'accréditation des organismes chargés des contrôles. *J Off la république française Texte* 35 sur 156.

Roach SA. (1966) A more rational basis for air sampling programs. *Am Ind Hyg Assoc J*; **27**: 1–12.

Selvin S, Rappaport S, Spear R *et al*. (1987) A note on the assessment of exposure using one-sided tolerance limits. *Am Ind Hyg Assoc J*; **48**: 89–93.

Sottas PE, Lavoué J, Bruzzi R *et al*. (2009) An empirical hierarchical Bayesian unification of occupational exposure assessment methods. *Stat Med*; **28**: 75–93.

Symanski E, Maberti S, Chan W. (2006) A meta-analytic approach for characterizing the within-worker and between-worker sources of variation in occupational exposure. *Ann Occup Hyg*; **50**: 343–57.

Tornero-Velez R, Symanski E, Kromhout H *et al*. (1997) Compliance versus risk in assessing occupational exposures. *Risk Anal*; **17**: 279–92.

U.S. EPA (2000) *Risk characterization: science policy council handbook. EPA-100-B-00-002*. Washington, DC: Science Policy Council, U.S. Environmental Protection Agency.

Waters M, McKernan L, Maier A *et al*. (2015) Exposure estimation and interpretation of occupational risk: enhanced information for the occupational risk manager. *J Occup Environ Hyg*; **12**(Suppl 1): S99–111.

Wild P, Hordan R, Leplay A, *et al*. (1996) Confidence intervals for probabilities of exceeding threshold limits with censored log-normal data. *Environmetrics* **7**: 247–59.