



Cross-validation and refinement of the Stoffenmanager as a first tier exposure assessment tool for REACH

Jody Schinkel, Wouter Fransman, Henri Heussen, et al.

Occup Environ Med 2010 67: 125-132 originally published online September 22, 2009

doi: 10.1136/oem.2008.045500

Updated information and services can be found at:

<http://oem.bmj.com/content/67/2/125.full.html>

	<i>These include:</i>
References	This article cites 32 articles, 18 of which can be accessed free at: http://oem.bmj.com/content/67/2/125.full.html#ref-list-1
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To order reprints of this article go to:

<http://oem.bmj.com/cgi/reprintform>

To subscribe to *Occupational and Environmental Medicine* go to:

<http://oem.bmj.com/subscriptions>

Cross-validation and refinement of the Stoffenmanager as a first tier exposure assessment tool for REACH

Jody Schinkel,¹ Wouter Fransman,¹ Henri Heussen,² Hans Kromhout,³
Hans Marquart,¹ Erik Tielemans¹

¹Business Unit Quality & Safety, TNO Quality of Life, The Netherlands ²Expert Centre for Chemical Risk Assessment, Arbo Unie, The Netherlands ³Environmental Epidemiology Division, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

Correspondence to

Jody Schinkel, TNO Quality of Life, PO Box 360, 3700 AJ Zeist, The Netherlands;
jody.schinkel@tno.nl

Accepted 5 June 2009
Published Online First
22 September 2009

ABSTRACT

Objectives For regulatory risk assessment under REACH a tiered approach is proposed in which the first tier models should provide a conservative exposure estimate that can discriminate between scenarios which are of concern and those which are not. The Stoffenmanager is mentioned as a first tier approach in the REACH guidance. In an attempt to investigate the validity of the Stoffenmanager algorithms, a cross-validation study was performed.

Methods Exposure estimates using the Stoffenmanager algorithms were compared with exposure measurement results (n=254). Correlations between observed and predicted exposures, bias and precision were calculated. Stratified analyses were performed for the scenarios "handling of powders and granules" (n=82), "handling solids resulting in comminuting" (n=60), "handling of low-volatile liquids" (n=40) and "handling of volatile liquids" (n=72).

Results The relative bias of the four algorithms ranged between -9% and -77% with a precision of approximately 1.7. The 90th percentile estimate of one out of four algorithms was not conservative enough. Based on these statistics and analyses of residual plots the underlying algorithm was adapted. Subsequently, the calibration and the cross-validation dataset were merged into one dataset (n=952) used for calibrating the adapted Stoffenmanager algorithms. This new calibration resulted in new exposure algorithms for the four scenarios.

Conclusions The Stoffenmanager is capable of discriminating among exposure levels mainly between scenarios in different companies. The 90th percentile estimates of the Stoffenmanager are verified to be sufficiently conservative. Therefore, the Stoffenmanager could be a useful tier 1 exposure assessment tool for REACH.

INTRODUCTION

Exposure assessment is preferably based on a sufficient number of exposure measurements.^{1 2} Since exposure measurements are often considered to be expensive and time consuming or not possible, (quantitative) exposure modelling is sometimes used as an alternative. However, modelling is complex and this is clearly reflected in the range of different approaches. Several (statistical) models have been developed to predict inhalation exposure under different circumstances. Some deterministic models have been developed for use with defined tasks such as embalming³ or laboratory work.⁴ Alternatively, exposure models for regulatory risk

What this paper adds

- ▶ Generic models used for predicting inhalation exposures in risk assessment have been validated only to a limited extent.
- ▶ The validity and reproducibility of the predictions of any model should be evaluated before results can be used safely and effectively in the risk assessment process.
- ▶ This validation study shows that the Stoffenmanager can be a useful quantitative tier 1 exposure assessment tool for REACH.
- ▶ A reasonable worst case estimate (eg, 90th percentile estimate) of the Stoffenmanager is a good measure for compliance testing.

assessment assess exposure for a broad range of exposure scenarios. The same holds for control banding systems providing advice on control in the workplace.⁵

Examples of these so-called generic models and control banding systems for inhalation exposure are the Estimation and Assessment of Substance Exposure model,⁶ COSHH essentials,^{7 8} ECETOC TRA,⁹ Stoffenmanager^{10 11} and the "Easy-to-use workplace control scheme for hazardous substances" (EMGK) (<http://www.reach-helpdesk.de/en/Exposure/Exposure.html>). Under REACH (the European chemicals legislation), the manufacturers and importers of dangerous substances are required to develop exposure scenarios that describe how risks relating to the substances can be controlled for workers, the general population and the environment (http://ec.europa.eu/environment/chemicals/reach/pdf/2007_02_reach_in_brief.pdf). Because many thousands of substances need to be assessed, a tiered approach is recommended in which relatively simple exposure models are used to efficiently assess worker exposure. Several of the models mentioned above have been summarised as first tier approaches in the REACH guidance.¹² Under REACH a tiered approach is proposed in which the first tier should provide a conservative (ie, protective) system that can discriminate between substances in scenarios which are of concern for workers' health and those which are not.¹³ An exposure scenario in this case includes a set of conditions that describes how the substance is used or manufactured and which risk management measures are implemented.

Methodology

The validity and reproducibility of the predictions of any model should be evaluated before results can be used safely and effectively in the risk assessment process. The lack of proper evaluation of new exposure models before introduction into the occupational hygiene profession has been criticised.¹⁴ Most of the above mentioned generic models have been validated only to some extent and often long after their introduction.^{15–19} The most comprehensive validation study focused on Estimation and Assessment of Substance Exposure, indicating that the model needed complete revision if it was to provide more reliable estimates of exposure.¹⁸ In general, the validation studies do not cover the broad range of scenarios necessary to provide a complete picture of the validity domain. In addition, it is unclear whether different models provide comparable results for the same scenario. This lack of insight does not allow a sound weighting of available tools, which will be clearly necessary in the scope of regulatory risk assessments.

In The Netherlands, the Stoffenmanager was initially designed as a prioritisation tool to make qualitative risk assessments and to provide advice on control in the workplace.¹⁰ To derive a quantitative exposure algorithm, relative exposure rankings of the Stoffenmanager have been compared with a comprehensive set of exposure measurements using mixed effect regression models.¹¹ With this calibrated model the scores can be converted into quantitative exposure estimates in concentration units.

During the collation of exposure measurement data used for this calibration of the Stoffenmanager, a web-based exposure database was developed. This database is referred to as the SToffenmanager Exposure And Modelling database (STEAMbase). The tight link between the Stoffenmanager and STEAMbase enables a cycle of periodical model validation and refinements. Updating of model calibration with new exposure data is necessary as workplace scenarios and exposure levels will change over time.^{20–21} Therefore, in our opinion a regular cycle of model development, model validation and model refinement is necessary so that models can be used in a responsible way.

The first loop through the cycle has been completed and described by Marquart *et al.*¹⁰ and Tielemans *et al.*¹¹ The focus of the present paper is to validate the predictions from the Stoffenmanager algorithms in the scope of first tier models for REACH. This paper describes the conclusions from this cross-validation, the new calibration of the algorithms after model improvements based on this validation, and the adapted algorithms. The ability of the Stoffenmanager to discriminate between exposure levels across and within premises was also discussed.

MATERIALS AND METHODS

Exposure algorithm

The exposure algorithm has been extensively described by Marquart *et al.*¹⁰ It is based on a source-receptor approach developed by Cherrie *et al.*²² and incorporates modifying factors related to source emission and dispersion of contaminants. Most parameters are divided into classes with scores on a logarithmic scale. The total personal exposure score (C_t) is the sum of exposure levels due to near field sources, far field sources and diffusive background sources, adjusted for possible separation of the worker from the source (η_{imm}). A source is considered to be in the near field if it is located within 1 m of the head of the worker. A far field source is more than 1 m away from the worker's head but still influences the exposure of the worker because the source

is within the same work area. Equation 1 represents the Stoffenmanager algorithm.

$$C_t = \left((E \cdot a) + \left(E \cdot H \cdot \eta_{lc} \cdot \eta_{gv-nf} \right) + \left(E \cdot H \cdot \eta_{lc} \cdot \eta_{gv-ff} \right) \right) \cdot \eta_{imm} \quad (1)$$

E =intrinsic emission of the product; for solids E is the dustiness (in classes) of the product; for liquids E is $(P_i/30\,000) \cdot F_i$ (with P_i representing vapour pressure (Pascal) and F_i the fraction of substance i in a mixture);

a =potential of diffusive sources not captured by questions regarding the far field sources;

H =score for type of handling;

η_{lc} =effect of local control measures;

η_{gv-nf} =effect of general ventilation in combination with room size for the near field (NF);

η_{gv-ff} =effect of general ventilation in combination with room size for the far field (FF);

η_{imm} =effect of separation of the worker from the source.

The semi-quantitative exposure score C_t has been calibrated with measured exposure data using mixed effects models to derive a quantitative exposure estimate.¹¹ The regression equations resulting from this calibration enable prediction of geometric mean exposures and any percentile prediction of the distribution for a given Stoffenmanager score (C_t).

STEAMbase

STEAMbase is an inhalation exposure database and its structure follows the principles and parameters of the Stoffenmanager algorithm (equation 1). Therefore, relevant Stoffenmanager exposure determinants like dustiness for solid materials, vapour pressure for liquid, type of handling, local control measures, room ventilation, room size, good housekeeping and separation of the worker from the source were stored per performed task. In addition, required core information for the exposure measurements²³ was also stored in STEAMbase (eg, sample technique, sample strategy, analytical method).

STEAMbase contains both task-based and shift-based measurements collected for the purpose of the Stoffenmanager calibration and evaluation projects. The collation of data is an ongoing process. Measurements have been collated from three sources: (1) measurement surveys specially conducted for the calibration and evaluation projects, (2) occupational exposure datasets from TNO studies funded by the Dutch government in the past, and (3) exposure data collected using a network of industry and occupational health services participating in the Dutch "VAST program". The VAST program was established by the Dutch Ministry of Social Affairs and Employment to assist small and medium enterprises (SMEs) in reinforcing the policy on working conditions and hazardous substances (<http://vast.szw.nl>).

Assignment of Stoffenmanager scores

In the time between the selection of data for the calibration of the Stoffenmanager¹¹ and the date we started this cross-validation study, additional data were stored in STEAMbase. Measurement data used for the calibration of the Stoffenmanager algorithms were not selected for the cross-validation dataset. All other data available in STEAMbase matching the following criteria were selected:

- Contextual information for all tasks performed during the measurement is available for all relevant determinants of the Stoffenmanager algorithm.

- ▶ Core information on sample strategy (sampling duration, random sampling, etc) and sample technique (sampling device, analytical method, etc) was documented.
- ▶ Measurement of personal inhalation exposure.

A score was assigned based on the contextual information and following the tables presented by Marquart *et al*¹⁰ for each Stoffenmanager parameter. Scores were assigned by the first author (JS) and reviewed by most of the co-authors (WF, HH, HM, ET). The parameter for which the assignment of a score most often led to disagreement by the co-authors was the handling score. In a meeting between the authors the handling scores were discussed until there was a consensus classification. This consensus classification was used to derive final estimates. When multiple tasks were performed during a measurement, a time-weighted Stoffenmanager score (C_i) was calculated.

Data processing and cross-validation

The data were analysed using SAS statistical software (v 9.1.3). Measured concentrations and estimated Stoffenmanager scores were found to approximate a log-normal rather than a normal distribution, so descriptive statistics are presented both as arithmetic and geometric mean levels with geometric standard deviation (GSD) and range. In situations where measured values were below the limit of detection (LOD), 0.5 times the LOD was substituted for measured values ($n=16$).²⁴

Spearman correlation coefficients were calculated to study the relation between Stoffenmanager scores and measured exposure concentrations. The level of conservatism of the algorithms is tested by calculating the proportion of measurements exceeding the 90th percentile exposure estimates of the predicted exposure distribution using the models derived during the calibration of the Stoffenmanager.¹¹ A proportion of 10% or less exceeding the 90th percentile estimate was considered to indicate sufficient conservatism. Bias and precision were calculated as described previously (equations 2 and 3).²⁵ Bias was defined as the mean difference between predicted estimates and measured exposure on a logarithmic scale.

$$\text{Bias} = \frac{\sum_{i=1}^{n_0} (\hat{y}_i - y_i)}{n_0} \quad (2)$$

$$\text{Precision} = \sqrt{\frac{\sum_{i=1}^{n_0} [(\hat{y}_i - y_i) - \text{bias}]^2}{n_0 - 1}} \quad (3)$$

\hat{y}_i = ln predicted exposure level for the i-th set of exposure factors in the validation set

y_i = ln measured exposure for the i-th set of exposure factors

n_0 = number of measurements in the validation set

A positive bias indicates tendency of the model to overestimate exposure, a negative bias indicates underestimation by the model. The relative bias was defined as the difference between predicted and measured exposure divided by the measured exposure (equation 4). The bias presents the systematic error of the model and shows the capability of the model to predict the "true" value. The variability around the bias is given by the precision which is the standard deviation of mean difference between measured values and predicted estimates on a logarithmic scale.

$$\text{Relative bias} = (e^{\text{bias}} - 1) * 100\% \quad (4)$$

Stratified analyses were made for different scenarios. The solid scenario is divided into scenarios describing handling of powders and granules (like scooping, dumping and filling) and scenarios handling solid materials resulting in comminuting (like sanding wood or grinding stone). The mechanisms for emission of dust are different for these two scenarios. Separation of intrinsic emission and type of handling is clear for powders and granules. In contrast, when handling products resulting in comminuting it is difficult to assign the intrinsic emission of the product because this is not independent of the performed task. For example, in processing wood or stone, the machine type affects the dustiness of the formed by-product (eg, sawing wood produces a smaller wood fraction than planing wood, but in both situations the product is solid). Therefore, the parameter for intrinsic emission was combined with the score for type of handling for this scenario.

Also, two different mechanisms of source emission are apparent for scenarios describing handling of liquid products. Evaporation is an important mechanism for volatile substances, whereas the likelihood of aerosol release determines emission in low-volatile products. This separation between both types of exposure processes was illustrated in the figures presented by Tielemans *et al*.¹¹ Therefore, separate analyses were also conducted to explore the differences between volatile (vapour pressure >10 Pascal) and low-volatile substances (vapour pressure ≤10 Pascal).

Model refinement and new calibration

The cross-validation of the Stoffenmanager algorithms based on bias and precision statistics and the analyses of residual plots provided important insights for refinement of the Stoffenmanager model. A broader range of classes was defined for the handling parameter and the intrinsic emission parameter was combined with the score for type of handling for the scenario handling solids resulting in comminuting. Subsequently, the previous calibration dataset¹¹ and data from the present study were merged to update the calibration using the adapted mechanistic model. The mixed-effect regression models described earlier¹¹ were used to further explore the relation between the new Stoffenmanager scores and exposure measurements. The natural log of exposure data was used as dependent variable and the log of the Stoffenmanager score as independent variable, with a random premise component of variance. A compound symmetric covariance structure was assumed in the presented models. Calibrations were conducted separately for the four scenarios.

RESULTS

Cross-validation for solid scenarios

To evaluate the model for solid scenarios a dataset of 142 exposure measurements was selected from STEAMbase following the criteria described above. Within this dataset 12 different industries were included with a median sampling time of 3 h and measured exposure levels ranging from 0.03 to 646 mg/m³ representing a total GSD of 7.4 (table 1). The lowest exposure was found during wet abrasion of a wooden sports floor, while the highest exposure levels occurred during bulk transshipment of products. The exposure data that were included in this evaluation phase covered a broad range of activities.

The Spearman correlation (r_s) between Stoffenmanager scores and measurements appeared to be moderate ($r_s=0.47$, $n=142$,

Methodology

Table 1 Descriptive statistics of available measured exposure data for solid and liquid scenarios

Type of industry	Substance	N	K	Median sampling time (min)	AM (mg/m ³)	GM (mg/m ³)	GSD	Minimum (mg/m ³)	Maximum (mg/m ³)
Overall solid scenarios		142	23	181	28.6	2.78	7.42	0.03	646.4
Carbon black industry	Inhalable dust	10	1	115	6.09	2.34	4.69	0.20	30.8
Cardboard industry	Inhalable dust	2	1	395	0.35	0.32	1.91	0.20	0.50
Cement industry	Inhalable dust	12	1	121	9.54	5.75	2.52	1.80	49.5
Chemical production industry	Inhalable dust	1	1	10	79.9	79.9		79.9	79.9
Metal industry	Inhalable dust	3	2	348	10.4	5.99	3.50	2.82	25.4
Paving	Inhalable dust	20	1	227	0.32	0.28	1.68	0.13	0.63
Pottery	Inhalable dust	4	1	323	1.37	1.17	1.84	0.71	2.80
Waste processing	Inhalable dust	4	1	330	13.1	5.51	5.16	0.87	39.1
Railway maintenance	Inhalable dust	10	1	105	4.19	2.55	2.93	0.61	16.0
Refinishing	Inhalable dust	8	1	58	11.0	1.41	19.7	0.03	39.0
Transshipment	Inhalable dust	38	3	173	90.7	12.4	9.19	0.25	646.4
Woodworking industry	Inhalable dust	30	9	307	4.22	1.84	2.79	0.30	60.5
Overall liquid scenarios		112	36	215	43.3	1.91	49.9	0.002	299.8
Pest control/disinfection	Biocides (alkyl dimethylbenzyl ammonium chloride)	21	10	305	0.05	0.02	4.28	0.004	0.34
Agriculture	Pesticides (captan)	9	1	7	0.02	0.01	2.40	0.002	0.06
Car body repair shops	Isocyanates (total HDI oligomers)	7	3	6	0.82	0.07	21.3	0.004	2.87
	Organic solvents (total volatile organic compounds)	1	1	19	135.9	135.9		135.9	135.9
Metal industry	Organic solvents (total volatile organic compounds)	6	1	74	123.8	87.1	2.58	31.8	299.8
Orthopaedic shoe manufacturing	Organic solvents (total volatile organic compounds)	7	6	510	89.2	77.5	1.79	37.5	152.3
Shoe repair industry	Organic solvents (total volatile organic compounds)	24	4	510	43.1	34.0	2.06	6.37	121.0
Silkscreen printing industry	Organic solvents (total volatile organic compounds)	1	1	452	16.3	16.3		16.3	16.3
Chemical production industry	Organic solvents (total volatile organic compounds)	15	2	62	37.7	6.79	7.28	0.43	210.0
Furniture industry	Organic solvents (total volatile organic compounds)	13	4	145	118.0	67.2	4.06	2.70	252.7
Transport	Organic solvents (total volatile organic compounds)	1	1	40	189.6	189.6		189.6	189.6
Cleaning industry	Formaldehyde	5	1	354	0.08	0.04	5.24	0.01	0.15
Textile industry	Sodium hydroxide	2	1	214	0.01	0.01	1.47	0.01	0.02

AM, arithmetic mean; GM, geometric mean; GSD, geometric standard deviation; K, number of premises; N, number of samples.

$p < 0.001$) and seemed to be better for handling resulting in comminuting ($r_s = 0.69$, $n = 60$, $p < 0.001$) than for handling of powders and granules ($r_s = 0.41$, $n = 82$, $p < 0.001$) (table 2). A relative bias of -59% for the overall model for solid scenarios indicates that the model, on average, underestimates exposure to inhalable dust in this dataset (table 2). The proportion of measurements exceeding the estimated 90th percentile, was 19% for the overall solid model.

Stratified analyses were performed for the two types of handling solids. For the group handling powders and granules the relative bias was -77% . So the measured exposures were on average almost a factor of two higher than predicted. Moreover, within this group 29% of the measurements had a measured

exposure above the predicted 90th percentile based on the Stoffenmanager scores. Conversely, for handling resulting in comminuting the relative bias was much smaller (-9%) and only 5% of the measurements were higher than the 90th percentile prediction. The precision of the bias for the overall inhalable dust model and the group handling powders and granules was comparable (precision = 1.7).

When modelling the data an assumption of log-normal distribution and homoscedasticity was made. If these basic assumptions for modelling hold, the residual plot should obtain a horizontal band with no hint of any systematic trends.²⁶ However, figure 1 illustrates a tendency of greater deviance with higher measurement results.

Table 2 Cross-validation of the Stoffenmanager algorithms

Group	N	r_s	Bias	Precision	Relative bias (%)	% >90th percentile
Overall solids	142	0.47	-0.90	1.81	-59	19
Handling of powders and granules	82	0.41	-1.49	1.66	-77	29
Handling resulting in comminuting	60	0.69	-0.10	1.71	-9	5
Overall liquids	112	0.85	-0.42	1.70	-34	10
Low-volatile substances	40	0.20	-0.97	1.75	-62	15
Volatile substances	72	0.63	-0.11	1.61	-11	7

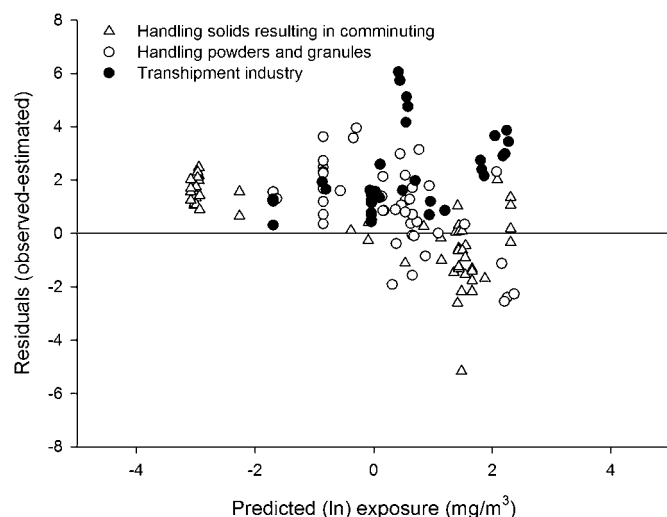


Figure 1 Plot of residuals of measured exposure and estimated exposure as a function of the predicted exposure for handling of solids.

Cross-validation for liquid scenarios

The dataset for liquid scenarios contained 112 exposure measurements from 12 different industries. As expected, the lowest exposures were found during scenarios with low-volatile substances, while the highest exposures occurred during spraying volatile substances. The measured exposure levels ranged from 0.002 to 300 mg/m³, representing a range of six orders of magnitude. This indicates a broad range of activities with very different exposures (table 1). The median sampling time was 3.5 h.

The correlation coefficient between Stoffenmanager scores and measurements for liquid scenarios appeared to be good ($r_s=0.85$, $n=112$, $p<0.001$). The measurements indicated that the Stoffenmanager was negatively biased (−34%) for liquid scenarios (table 2), indicating predicted exposure was half as low as the measured exposure. Ten per cent of the measurements within this cross-validation dataset were above the predicted 90th percentile. The precision ≈ 1.7 was comparable with the precision of the bias found within the solid scenarios (table 2).

The groups of low-volatile and volatile substances were also analysed separately. For low-volatile substances poor correlation was seen ($r_s=0.20$, $n=40$, $p=0.21$) and we found a moderate correlation for volatile substances ($r_s=0.63$, $n=72$, $p<0.001$). For the low-volatile and volatile substances the relative bias was −62% and −11%, respectively. Moreover, the percentage of measurement results above the 90th percentile prediction was 15% for low-volatile and 7% for volatile substances. The precision of the bias of ≈ 1.7 was almost the same for the different models (table 2).

The residual plot for liquid scenarios (figure 2) shows overestimation of the low exposures and underestimation of the higher exposures.

Model refinement

The range in exposure to inhalable dust found in this paper and the paper of Tielemans *et al*¹¹ was 0.0004–300 mg/m³. These seven orders of magnitude are not reflected in the range of Stoffenmanager scores. Figure 1 shows that the exposure is particularly underestimated in the transshipment industry. Furthermore, Tielemans *et al*¹¹ presented a figure which showed that for handling very small amounts of product (eg, weighing drugs in pharmacies) exposures were overestimated. This provides a clear

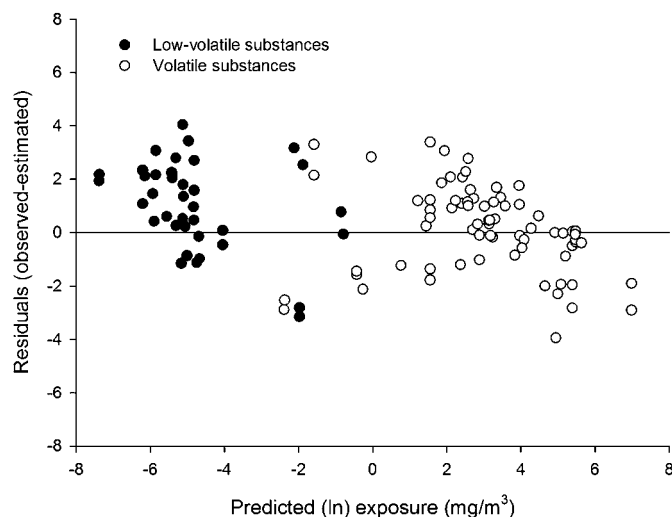


Figure 2 Plot of residuals of measured exposure and estimated exposure as a function of the predicted exposure for handling of liquids.

indication that the range of relative scores in the handling parameter is not large enough to cover the broad scope of activities. Therefore, the range was increased from two (ie, 0.1–10) to three (ie, 0.03–30) orders of magnitude by adding new categories to the handling score table as presented by Marquart *et al*.¹⁰ A lower handling score (0.03) was added for handling very low quantities of products (<10 g) and a higher score (30) for activities where tons of product are handled in an uncontrolled manner (see online appendix A). In addition, a new table for handling scores was developed for handling resulting in comminuting where product and handling type are combined (see online appendix B).

Furthermore, a small change was made in the table describing scores for handling liquids. A score of 0.03 was added to the table covering, for example, pipetting very small amounts of liquid. A score of 30 was considered not to be necessary for the liquid scenarios. The new tables for handling scores are presented in online appendix C.

Calibration of new models

Because the mechanistic model of the Stoffenmanager was adapted and more measurements were available for calibration, the relation between the Stoffenmanager model and measured exposure levels was investigated once more by using mixed effects models with a random premise effect. The data used in the previous calibration ($n=698$)¹¹ and the dataset collected for the cross-validation of these algorithms described in this paper ($n=254$) were merged. The new database ($n=952$) is used to calibrate the new models for handling powders and granules ($n=408$), handling resulting in comminuting ($n=112$), handling low-volatile substances ($n=256$) and handling volatile substances ($n=176$). The four new models are presented in table 3. The slope of the regression line shows a positive linear relation between the log-transformed Stoffenmanager scores and the log-transformed exposure measurement results for all scenarios. These regression equations enable the estimation of geometric mean exposures for a given Stoffenmanager score (C_i) following equation 5.

$$\hat{Y} = e^{\beta_0 + \beta_1 \ln(C_i)} \quad (5)$$

\hat{Y} = exposure prediction
 β_0 = intercept
 β_1 = slope of the regression line
 C_i = Stoffenmanager score

Methodology

Table 3 Results of the mixed effect regression models for the new calibration, with premise included as random effect (model A) and model A with Stoffenmanager score added as fixed effect (model B)

Scenario	K	n	Model	β_0	β_1	s^2_{bp}	s^2_{wp}	% s^2_{bp} explained	% s^2_{wp} explained
Handling powders and granules	76	408	A	0.70 (0.25)*		4.01	1.81		
			B	1.60 (0.18)**	0.69 (0.05)**	1.74	1.27	57	30
Handling resulting in comminuting	29	112	A	1.25 (0.31)**		2.29	1.27		
			B	0.38 (0.27) (0.23)**	0.78 (0.12)**	1.18	1.04	48	18
Handling low-volatile substances	86	256	A	-3.67 (0.22)**		2.21	3.82		
			B	3.77 (0.92)**	0.66 (0.08)**	1.06	3.38	52	12
Handling volatile substances	48	176	A	3.47 (0.30)**		3.72	1.34		
			B	5.32 (0.32)**	0.53 (0.07)**	1.76	1.19	53	11

K, number of premises; n, number of measurements.

β_0 =intercept; β_1 =regression coefficient; s^2_{bc} =between premise variance component; % s^2_{bc} explained=proportion between premise variance explained by the model; s^2_{wc} =within premise variance component; % s^2_{wc} explained=proportion within premise variance explained by the model.

* $p<0.01$; ** $p<0.001$.

The calibration provides insight in the variation around the geometric mean exposure. This variation can be used to predict the exposure for any percentile of interest with the following formulas (equations 6 and 7):

$$RWCEstimate = \hat{Y} \cdot M \quad (6)$$

$$M = e^{Z \cdot \sqrt{s^2_{bc} + s^2_{wc}}} \quad (7)$$

Z=Z value for normal distribution

s^2_{bc} =between premise variance

s^2_{wc} =within premise variance

The proportion of total variance in exposure explained by the Stoffenmanager algorithms ranged between 26% and 48% for the four models. The Stoffenmanager explained 48–57% of the variance between premises across industries and 11–30% of the within premise variance (table 3). This indicates that the Stoffenmanager algorithms could discriminate between different exposure situations mostly found between premises, and is less accurate in discriminating among relatively similar exposure situations generally found within the same premises.

To illustrate that the Stoffenmanager explained less variance in more similar exposure situations, the data were grouped according to industries. Unfortunately, these groups were very small. For each scenario table 4 illustrates calibrations of the models for the two industries where most measurements were conducted. Indeed, less variance was explained by the Stoffenmanager if within an industry approximately the same exposure situations occurred. For example, almost no variance is explained in the car body repair shops and in the pest control industry although measured exposure levels within these industries vary substantially.

DISCUSSION

This paper presents the results of the cross-validation of the Stoffenmanager exposure model using independent datasets. This evaluation resulted in an adapted Stoffenmanager model and a new calibration based on more exposure measurements, covering a broader range of work environments. The cross-validation method as used in this paper is a valuable technique for evaluating predictive models developed from exposure measurements.²⁵

Table 4 Results of the mixed effect regression models per industry, with premise included as random effect (model A) and model A with Stoffenmanager score added as fixed effect (model B)

Scenario Industry	K	n	Model	β_0	β_1	s^2_{bp}	s^2_{wp}	% s^2_{bp} explained	% s^2_{wp} explained
Handling powders and granules Pharmacies	9	78	A	-2.9 (0.21)**		0.04	2.90		
			B	0.05 (0.57)	0.78 (0.14)**	0.12	2.03	0	30
Bakeries	17	56	A	0.26 (0.17)		0.10	1.25		
			B	0.87 (0.19)**	0.61 (0.13)**	0.09	0.86	0	31
Handling resulting in comminuting Construction industry	11	29	A	2.74 (0.43)**		1.72	0.52		
			B	2.20 (0.52)**	0.33 (0.22)	1.14	0.60	34	0
Woodworking industry	14	51	A	0.45 (0.20)		0.34	0.71		
			B	-0.13 (0.24)	0.57 (0.16)**	0.28	0.57	18	20
Handling low-volatile substances Car body repair shops (isocyanates)	19	97	A	-2.58 (0.27)**		0	6.89		
			B	5.00 (1.71)*	0.75 (0.17)**	0	5.75	0	17
Pest controllers	20	59	A	-4.16 (0.46)**		3.63	0.79		
			B	0.63 (2.43)	0.38 (0.19)	3.03	0.80	17	0
Handling volatile substances Metal industry	14	62	A	3.98 (0.38)**		1.64	1.54		
			B	5.71 (0.45)**	0.48 (0.10)**	0.61	1.35	63	12
Orthopaedic branch	11	33	A	4.81 (0.33)** (0.33)****		0.98	0.52		
			B	5.51 (0.50)**	0.28 (0.16)	0.76	0.52	22	0

K, number of premises; n, number of measurements.

β_0 =intercept; β_1 =regression coefficient; s^2_{bp} =between premise variance component; % s^2_{bp} explained=proportion between premise variance explained by the model; s^2_{wp} =within premise variance component; % s^2_{wp} explained=proportion within premise variance explained by the model.

* $p<0.01$; ** $p<0.001$.

Few comparably detailed evaluations of predictive models exist, and those models are industry specific and not as generic as the Stoffenmanager algorithms. Models developed for specific scenarios are expected in advance to be more accurate compared to the performance of generic models. There is generally a trade-off between accuracy of the model and broadness of application domain. Indeed, the relative bias found in other studies was slightly smaller than found in our study and roughly varied between -50% and 50% .^{27–30} Friesen and colleagues³¹ found very low relative bias for a model predicting historical dust- and wood-exposure in sawmills (-12% and -1%). Also, because of the generic purposes of the Stoffenmanager, less precise predictions were expected when investigating the predictive capabilities of the algorithms. The precision of the (relative) bias of our algorithms was similar,^{28, 29} or less precise,^{27, 31} compared with other studies. Currently, most generic models are validated only to a limited extent. This prohibits proper comparison of the Stoffenmanager with similar generic exposure models.

The proportion of exposure measurements exceeding the 90th percentile estimate of the algorithms should be 10% or less in order to be useful as a tier 1 screening tool. For the solid scenarios the proportion exceeding the 90th percentile estimates was 19%. Stratified analysis showed that this proportion was 29% (18% in the transshipment industry) in the group handling of powders and granules and 5% in the group handling resulting in comminuting. Evaluating the algorithm for liquid scenarios, the proportion above the 90th percentile estimates was 10% for the overall model and 15% and 7% for the low-volatile and volatile scenarios, respectively. However, within the low-volatile scenarios the estimated exposure was in some cases lower than the LOD for that substance. An analysis excluding measurements below the LOD showed that only 6% of the measurements had a result above the estimated 90th percentile. Overall, three out of four algorithms were sufficiently conservative and only one algorithm (handling of powders and granules) appeared to be not protective enough. This algorithm was therefore adapted, introducing a broader range of relative scores in the handling parameter.

Previously, three types of uncertainty were summarised¹¹ resulting in scattering of exposure measurements given a certain Stoffenmanager score. The first type was uncertainty in input parameters, such as product information from safety data sheets or contextual information about circumstances described in the reports.²⁵ Secondly, there is uncertainty in measurement data. This could be due to analytical error across laboratories or differences in aerosol sampling instruments.^{25, 32} However, measurement error is generally believed to be minor as compared to true exposure variability.³³ Thirdly, a more fundamental reason for discrepancies is model uncertainty. Although the underlying concepts incorporate the critical determinants of exposure,^{16, 34} the model could not deal with all the determinants influencing exposure (eg, personal behaviour, ventilation rate). Besides, the relative scores of the model parameters are largely based on expert judgement and may not represent their “real effect” on exposure. Ideally, much more effort should be devoted to scientific underpinning of the scores. Currently, additional work in this area is being conducted and results are expected soon.¹³

The dataset contains both shift- and task-based measurements. Also, for the four groups of scenarios separate analyses were conducted only using task measurements (results not shown). We expected smaller parameter uncertainty in task measurements than in multiple task measurements as collection

of contextual information is more complex in the latter case. These task-based models showed better performance for the group handling powders and granules. For the three other groups, the models showed only very small improvements (results not shown). Because the models did not change substantially, while the number of measurements underpinning the models did decrease substantially, it was decided to calibrate algorithms using both task and multiple task measurements in the dataset.

After cross-validation of the algorithms, model refinements were made and new algorithms were derived from the total dataset. Despite the fact that this dataset consists of almost 1000 measurements across different industries and premises, we should note that the dataset will not contain the diversity of scenarios needed for REACH to cover all the combinations of exposure determinants. Therefore, the validity of the Stoffenmanager algorithms is limited to the scenarios evaluated in this study. Furthermore, the models could only be used for estimating inhalable dust, liquid aerosols and vapour exposures. Exposures to respirable dust, gases or fibres for example, were not part of this study. Hence, these types of exposures are beyond the validity domain of the Stoffenmanager.

The Stoffenmanager algorithms were derived from a statistical model based on a deterministic model and therefore incorporate a limitation. This limitation is provoked by the limitations of a deterministic model. A true determinant may be missed or an incorrect value given to a certain determinant.³⁵ The explained total variance by the Stoffenmanager model was between 26% and 48%. The unexplained variance was comparable with a GSD of 5.7–8.2. Explained variance was much larger for between-premise components of variance as compared to within-premise components of variance, indicating that the Stoffenmanager can primarily discriminate between scenarios in different premises. This could be due to the fact that exposure determinants vary more between premises than within premises. In most cases the Stoffenmanager algorithms are not sophisticated enough to discriminate between more subtle exposures differences among workers within a scenario. Therefore, it is concluded that the Stoffenmanager algorithms could predict average exposures at scenario level. To ensure a conservative estimate for an individual worker, the unexplained variance within scenarios must be taken into account. This could, for example, be done by using the conservative 90th percentile of a predicted exposure distribution as a measure to assess compliance testing.

The evaluation of the Stoffenmanager algorithms has resulted in model improvements and in four new algorithms based on more exposure data. The new models for the solid scenarios differ only slightly from the first models presented by Tielemans *et al*,¹¹ whereas 142 measurements were added to the dataset. For the liquid scenarios 112 additional measurements were added and the model refinements resulted in a separate model for volatile and low-volatile liquids. However, the usefulness of this cross-validation depends heavily upon how representative the measured data were.²⁹ Following our philosophy of model development, these algorithms should be validated again with an independent dataset in the future.

Based on this cross-validation study and subsequent adaptations, we conclude that the Stoffenmanager can be used as a tier 1 exposure assessment model for regulatory risk assessment. It is hoped that more validity studies will follow in the near future in order to provide insight into the validity domain of the various models allowed for exposure assessment under REACH.

Methodology

Funding The Dutch Ministry of Social Affairs and Employment.

Competing interests None.

Provenance and peer review Not commissioned; not externally peer reviewed.

REFERENCES

- Rappaport SM.** Assessment of long-term exposures to toxic substances in air. *Ann Occup Hyg* 1991;**35**:61–121.
- Kromhout H, et al.** Design of measurement strategies for workplace exposure. *Occup Environ Med* 2002;**59**:349–54.
- Hornung RW, Herrick RF, Stewart PA, et al.** An experimental design approach to retrospective exposure assessment. *Am Ind Hyg Assoc J* 1996;**57**:251–6.
- Keil C, Murphy R.** An application of exposure modeling in exposure assessments for a university chemistry teaching laboratory. *J Occup Environ Hyg* 2006;**3**:99–106.
- Zalk DM, Nelson DI.** History and evolution of control banding: a review. *J Occup Environ Hyg* 2008;**5**:330–46.
- Tickner J, Friar J, Creely KS, et al.** The development of the EASE model. *Ann Occup Hyg* 2005;**49**:103–10.
- Russell RM, Maidment SC, Brooke I, et al.** An introduction to a UK scheme to help small firms control health risks from chemicals. *Ann Occup Hyg* 1998;**42**:367–76.
- Maidment SC.** Occupational hygiene considerations in the development of a structured approach to select chemical control strategies. *Ann Occup Hyg* 1998;**42**:391–400.
- ECETOC.** Targeted risk assessment, technical report no. 93. Brussels, Belgium: European Centre for Ecotoxicology and Toxicology of Chemical, 2004.
- Marquart H, Heussen H, Le Feber M, et al.** 'Stoffenmanager', a web-based control banding tool using an exposure process model. *Ann Occup Hyg* 2008;**52**:429–41.
- Tielemans E, Noy D, Schinkel J, et al.** Stoffenmanager exposure model: development of a quantitative algorithm. *Ann Occup Hyg* 2008;**52**:443–54.
- European Chemicals Agency.** Occupational exposure estimation. In: *Guidance on information requirements and chemical safety assessment*. Helsinki: ECHA, 2008: chapter R. 14.
- Tielemans E, Warren N, Schneider T, et al.** Tools for regulatory assessment of occupational exposure: development and challenges. *J Expo Sci Environ Epidemiol* 2007;**17**:S72–80.
- Kromhout H.** Author's reply. *Occup Environ Med* 2002;**59**:788–9.
- Jones RM, Nicas M.** Evaluation of COSHH essentials for vapor degreasing and bag filling operations. *Ann Occup Hyg* 2006;**50**:137–47.
- Creely KS, Tickner J, Soutar AJ, et al.** Evaluation and further development of EASE model 2.0. *Ann Occup Hyg* 2005;**49**:135–45.
- Bredendiek-Kämper S.** Do EASE scenarios fit workplace reality? A validation study of the EASE model. Estimation and assessment of substance exposure. *Appl Occup Environ Hyg* 2001;**16**:182–7.
- Cherrie JW, Hughson GW.** The validity of the EASE expert system for inhalation exposures. *Ann Occup Hyg* 2005;**49**:125–34.
- Tischer M, Bredendiek-Kämper S, Poppek U.** Evaluation of the HSE COSHH Essentials exposure predictive model on the basis of BAUA field studies and existing substances exposure data. *Ann Occup Hyg* 2003;**47**:557–69.
- Kromhout H, Vermeulen R.** Long-term trends in occupational exposure: are they real? What causes them? What shall we do with them? *Ann Occup Hyg* 2000;**44**:325–7.
- Creely KS, Cowie H, van Tongeren MJ, et al.** Trends in inhalation exposure—a review of the data in the published scientific literature. *Ann Occup Hyg* 2007;**51**:665–78.
- Cherrie JW, Schneider T, Spankie S, et al.** A new method for structured, subjective assessments of past concentrations. *Occup Hyg* 1996;**3**:75–83.
- Rajan B, Alesbury R, Carton B, et al.** European proposal for core information for the storage and exchange of workplace exposure measurements on chemical agents. *Appl Occup Environ Hyg* 1997;**12**:31–9.
- Hornung RW, Reed L.** Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg* 1990;**5**:46–51.
- Hornung RW.** Statistical evaluation of exposure assessment strategies. *Appl Occup Environ Hyg* 1991;**6**:516–20.
- Kleinbaum DG, Kupper LL, Muller KE, et al.** *Applied regression analysis and other multivariate methods*. 3rd edn. Pacific Grove, CA: Duxbury Press, 1998.
- Astrakianakis G, Seixas NS, Camp JE, et al.** Modeling, estimation and validation of cotton dust and endotoxin exposures in Chinese textile operations. *Ann Occup Hyg* 2006;**50**:573–82.
- Burstyn I, Boffetta P, Burr GA, et al.** Validity of empirical models of exposure in asphalt paving. *Occup Environ Med* 2002;**59**:620–4.
- Hornung RW, Greife AL, Stayner LT, et al.** Statistical model for prediction of retrospective exposure to ethylene oxide in an occupational mortality study. *Am J Ind Med* 1994;**25**:825–36.
- Liu Y, Woodin MA, Hauser R, et al.** Estimation of personal exposures to particulate matter and metals in boiler overhaul work. *J Occup Environ Med* 2005;**47**:68–78.
- Friesen MC, Davies HW, Teschke K, et al.** Predicting historical dust and wood dust exposure in sawmills: model development and validation. *J Occup Environ Hyg* 2005;**2**:650–8.
- Kenny LC, Aitken R, Chalmers C, et al.** A collaborative European study of personal inhalable aerosol sampler performance. *Ann Occup Hyg* 1997;**41**:135–53.
- Nicas M, Simmons BP, Spear RC, et al.** Environmental versus analytical variability in exposure measurements. *Am Ind Hyg Assoc J* 1991;**52**:553–7.
- Tielemans E, Schneider T, Goede H, et al.** Conceptual model for assessment of inhalation exposure: defining modifying factors. *Ann Occup Hyg* 2008;**52**:577–86.
- Stewart P, Stenzel M.** Exposure assessment in the occupational setting. *Appl Occup Environ Hyg* 2000;**15**:435–4.