Annals of Work Exposures and Health, 2019, Vol. 63, No. 2, 230–241
doi: 10.1093/annweh/wxy098
Advance Access publication 9 December 2018
Original Article

Original Article

Evaluation of Exposure Assessment Tools under REACH: Part II—Higher Tier Tools

Eun Gyung Lee[1,*], Judith Lamb[2,8], Nenad Savic[3], Ioannis Basinas[2], Bojan Gasic[4], Christian Jung[5], Michael L. Kashon[6], Jongwoon Kim[7,8], Martin Tischer[5], Martie van Tongeren[2,9], David Vernez[3] and Martin Harper[1,10]

[1]Exposure Assessment Branch (EAB), Health Effects Laboratory Division (HELD), National Institute for Occupational Safety and Health (NIOSH), 1095 Willowdale Road, Morgantown, WV 26505, USA; [2]Institute of Occupational Medicine (IOM), Avenue North, Heriot Watt Research Park, Riccarton, Edinburgh EH14 4AP, UK; [3]Institute for Work and Health (IST), University of Lausanne and Geneva, 1066 Epalinges-Lausanne, Switzerland; [4]Swiss State Secretariat for Economic Affairs (SECO), Bern, Switzerland; [5]Federal Institute for Occupational Safety and Health (BAuA), Dortmund, Germany; [6]Biostatistics and Epidemiology Branch, Health Effects Laboratory Division (HELD), National Institute for Occupational Safety and Health (NIOSH), 1095 Willowdale Road, Morgantown, WV 26505, USA; [7]Korea Institute of Science and Technology (KIST) Europe, Campus E 7.1, D-66123 Saarbrücken, Germany; [8]Present address: Korea Research Institute of Chemical Technology (KRICT), 141 Gajeongro, Yuseong, Daejeon 34114, Korea. [9]Present address: Centre for Occupational and Environmental Health and Centre for Epidemiology, Faculty of Biology, Medicine, and Health, The University of Manchester, Manchester, UK. [10]Present address: Zefon International Inc., 5350 SW 1st Lane, Ocala, FL 34474, USA; Department of Environmental Engineering Science, University of Florida, Gainesville, FL 32611, USA.

*Author to whom correspondence should be addressed. Tel: +1-304-285-6041; e-mail: dtq5@cdc.gov

Submitted 3 August 2017; revised 3 October 2018; editorial decision 24 October 2018; revised version accepted 7 November 2018.

Abstract

Stoffenmanager®v4.5 and Advanced REACH Tool (ART) v1.5, two higher tier exposure assessment tools for use under REACH, were evaluated by determining accuracy and robustness. A total of 282 exposure measurements from 51 exposure situations (ESs) were collected and categorized by exposure category. In this study, only the results of liquids with vapor pressure (VP) > 10 Pa category having a sufficient number of exposure measurements (n = 251 with 42 ESs) were utilized. In addition, the results were presented by handling/activity description and input parameters for the same exposure category. It should be noted that the performance results of Stoffenmanager and ART in this study cannot be directly compared for some ESs because ART allows a combination of up to four subtasks (and nonexposed periods) to be included, whereas the database for Stoffenmanager, separately developed under the permission of the legal owner of Stoffenmanager, permits the use of only one task to predict exposure estimates. Thus, it would be most appropriate to compare full-shift measurements against ART predictions (full shift including nonexposed periods) and task-based measurements against task-based Stoffenmanager predictions. For liquids with VP > 10 Pa category, Stoffenmanager®v4.5 appeared to be reasonably accurate and robust when predicting exposures [percentage of measurements exceeding the tool's 90th percentile estimate (%M >T) was 15%]. Areas that could potentially be improved include ESs involving the task of handling of liquids on large surfaces or large work pieces, allocation of high and medium VP inputs, and absence of local exhaust ventilation input. Although the ART's median predictions appeared to be reasonably accurate for liquids with VP > 10 Pa, the %M >T for the 90th percentile estimates was 41%, indicating that variance in exposure levels is underestimated by ART. The %M >T using the estimates of the upper value of 90% confidence interval (CI) of the 90th percentile estimate (UCI90) was considerably reduced to 18% for liquids with VP > 10 Pa. On the basis of this observation, users might be to consider using the upper limit value of 90% CI of the 90th percentile estimate for predicting reasonable worst case situations. Nevertheless, for some activities and input parameters, ART still shows areas to be improved. Hence, it is suggested that ART developers review the assumptions in relation to exposure variability within the tool, toward improving the tool performance in estimating percentile exposure levels. In addition, for both tools, only some handling/activity descriptions and input parameters were considered. Thus, further validation studies are still necessary.

Keywords: exposure assessment tools; inhalation tools; REACH; REACH higher tools; tier tools; validation

Introduction

Tier 1 exposure assessment tools described in the European Chemicals Agency (ECHA) (2016) R14 guidance are designed to be simple and easy to use and to be conservative by overestimating a potential exposure for a defined exposure scenario. If exposures estimated using the first tier tools exceed the derived no-effect level of a substance, it is recommended to use higher tier tools that are developed to generate exposure levels with greater accuracy and less uncertainty. Higher tier tools for inhalation exposure include Stoffenmanager® and the Advanced REACH Tool (ART). The background information of both tools using the same concept of source–receptor approach by considering near- and far-field regions is well described in the ECHA R14 guidance and previous publications (Marquart *et al.*, 2008; Tielemans *et al.*, 2008; Cherrie *et al.*, 2011; Fransman *et al.*, 2011; Schinkel *et al.*, 2011; Tielemans *et al.*, 2011; van Tongeren *et al.*, 2011; Schinkel *et al.*, 2013; McNally *et al.*, 2014; Schinkel *et al.*, 2014). In addition, a conceptual evaluation for Stoffenmanager has been performed by Hesse *et al.* (2015).

Currently, few studies are available regarding the external validity of the tools with a summary of the validations for the higher tier tools provided in Supplementary Table 1 (available at *Annals of Work Exposures and Health* online). A sufficient number of exposure measurements and adequate contextual information on exposure determinants are required to properly evaluate tools for use under REACH. Some though design to collect exposure measurements as specifically to evaluate the higher tier tools, they were restrained to only a limited number of exposure measurements despite acquisition of information on exposure determinants (Hofstetter *et al.*, 2013; Landberg *et al.*, 2015). Others dealt with relatively large numbers of exposure measurements using pre-existing data extracted from databases (Schinkel *et al.*, 2010; Vink *et al.*, 2010; Koppisch *et al.*, 2012; Savic *et al.*, 2017; Spinazze *et al.*, 2017; van Tongeren *et al.*, 2017). Unfortunately, because such pre-existing data were not collected for the purpose of the tools' evaluation, crucial contextual information was lacking. As a consequence, assumptions for unknown input parameters were unavoidable, leading to uncertain validation results.

As described in Part I (Lee *et al.*, 2019), a project entitled 'Evaluation of Tier 1 Exposure Assessment Models (ETEAM) used under REACH,' was recently completed using pre-existing exposure data from Europe and the USA. The validation results for European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) Targeted Risk Assessment (TRA; version 2 and 3), Metals' Estimation and Assessment of Substance Exposure (MEASE) (version 1.02.01), easy-to-use workplace control scheme for hazardous substances (EMKG)-EXPO-TOOL, and Stoffenmanager® (version 4.5) were reported by van Tongeren *et al.* (2017). During the ETEAM project, a concern about using the assumptions for the unknown input parameters was raised at the advisory board meeting. Therefore, as an extended study on the ETEAM project, this study was conducted to evaluate the REACH exposure tools using exposure measurement results and all required contextual information on exposure determinants. We evaluated Tier 1 tools and higher tier tools using the same exposure measurement data, and the validation results are presented in a series of articles. The validation results of the Tier 1 tools are reported as Part I (Lee *et al.*, 2019). This Part II paper summarizes the results of the validation of the higher tier tools [Stoffenmanager v4.5 (hereinafter referred to as Stoffenmanager) and ART v1.5] in terms of accuracy and robustness.

Methods

Field surveys and development of exposure situation

National Institute for Occupational Safety and Health (NIOSH) collected personal exposure measurements at 18 workplaces in the USA, covering paint, aircraft, and wind mill industries; dry-cleaning shops; hospital and denture labs; and print shops among others. As a result, 282 exposure measurements (number of measurements applicable for Stoffenmanager and ART) were collected and analyzed by a NIOSH contract laboratory or an in-house lab according to the relevant NIOSH or Occupational Safety and Health Administration (OSHA) sampling and analytical method. The collected measurements in this study included no data below the limit of detection and were clustered into the same exposure categories as those defined by Lee *et al.*, 2019. Those exposure categories were defined as follows: (i) aqueous solutions, (ii) liquids with a vapor pressure (VP) ≤ 10 Pa at a room temperature, (iii) liquids with a VP > 10 Pa at a room temperature, (iv) powders, and (v) solid objects. The exposure category 'metal processing' was excluded as being out of scope for Stoffenmanager and ART. During the collection of exposure measurements, we also collected contextual information on exposure determinants required for tools' input parameters. Detailed information about tasks and number of samples for each task are listed in Supplementary Table 2 (available at *Annals of Work Exposures and Health* online).

After the field samplings, a NIOSH senior industrial hygienist who participated in the sampling campaigns developed 51 exposure situations (ESs) as table formats. Each ES included a narrative for the general description of the situation/activity, environmental conditions and exposure pattern, product information, etc. No exposure measurements were included. All ESs were then transferred to a Microsoft Access database, developed as part of the ETEAM project. This Access database is user friendly with various add-ons including the description of ES and options of tool selection among all Tier 1 tools and Stoffenmanager. Additional information about the field measurements and development of ESs are in Lee *et al.*, 2019.

Translation of contextual information into the tools' input parameters

The Microsoft Access database was sent to six assessors (E.L., J.L., N.S., B.G., J.K., and M.T.), and the translation of contextual information into the tool input parameters was carried out independently. For Stoffenmanager, assessors were asked to use the Microsoft Access database, developed under the permission of the legal owner of Stoffenmanager. For ART, due to the complexity of the tool mechanism, each assessor was asked to create an account in ART (https://www.advancedreachtool.com/) and to code the input parameters directly into the web-based tool. The ART allows the inputs and outputs to be exported into a Microsoft Excel report, and all reports from all assessors and ESs were collected together with a summary of the range of ART-generated percentiles of the predicted exposure and associated confidence intervals (CIs).

Among 51 ESs, 24 ESs (23 ESs from liquids with VP > 10 Pa category and 1 ES from powders category) had multiple subtasks (from 2 to 4 subtasks) with various control methods. Although it is possible to consider multiple subtasks within Stoffenmanager at www.stoffenmanager.nl, this does not apply to the Stoffenmanager algorithms that were programmed in Microsoft Access database we used. In this study, therefore, we combined several subtasks into one task (defined based on a job task determined by company) and applied the lowest exposure control method in the same manner as we did for the evaluation of Tier 1 tools (Lee *et al.*, 2019). As ART is capable of accounting for a maximum of four subtasks, subdivision of task was carried out where appropriate.

All the collected input parameters from the assessors were discussed at a face-to-face meeting in July 2015 to achieve a consensus on final inputs. The meeting was led by the NIOSH staff who participated in the sampling campaigns and developed ESs. Pictures and video clips, taken by the company's permission, were used for further explanation. The assessors were blind to the collected exposure measurements.

Generation of the tool estimates

With the agreed input parameters for each tool, the Stoffenmanager semiquantitative scores were calculated using the published algorithms incorporated into the Microsoft Access database. The score from the tool was then converted to a quantitative exposure estimate using the equations by Schinkel *et al*. (2010). For Stoffenmanager, the 50th, 75th, and 90th percentile estimates were reported. For ART, estimates were generated using the web-based tool and reported full-shift exposures of 50th, 75th, and 90th percentiles along with the upper value of 90% CI of each percentile estimate (i.e. UCI50, UCI75, and UCI90).

Data analyses

Some of the collected exposure measurements represented only specific tasks when the exposure was to a chemical for the whole sampling time. Some others included both exposed and nonexposed times because we were not able to change or hold a sampling media every time a worker conducted different subtasks. Thus, it was necessary to adjust the time-weighted average (TWA) exposures into the comparable measurements to the tools' estimates. Both Stoffenmanager and ART are considered to be task-based tools. For Stoffenmanager, we converted the TWA exposures to the task-based exposure data by dividing the exposed time over the sampling duration. Stoffenmanager generates exposure estimates as a distribution from which the user can select specific percentiles. For example, if a user selects a 90th percentile estimation, it means that the 90% of the exposure measurements in the exposure distribution of the measured data were below the selected 90th percentile estimation.

ART also provides exposure estimates in the form of a distribution but either for full-shift (recommended for REACH evaluation) or long-term exposures by considering both exposed and nonexposed times. A user can then select percentiles of the exposure distribution (i.e. 50th, 75th, 90th, 95th, or 99th) and a CI for each percentile (i.e. interquartile, 80%, 90%, or 95%) to be estimated. The former expresses exposure variability whereas the latter indicates the uncertainty around the percentile estimate. For the ART evaluation, the full-shift exposures were compared with the equivalent ART estimates. For the full-shift exposure measurements, we converted the TWA exposures (based on the sampling time during the field surveys) to the corresponding full-shift exposures by assuming zero exposure for the uncollected time (if a sampling time was less than 8 h).

To determine an accuracy and robustness of each tool, all individual exposure measurement results were assumed to be independent. For both tools, the 50th, 75th, and 90th percentile estimates were selected for the purpose of the evaluation. In addition, the upper value of 90% CI of each percentile estimate was obtained for ART. To determine the accuracy of the tools, we compared each measurement value with the median (50%) tool estimate of Stoffenmanager and ART using the following equations (Hornung, 1991):

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

Relative bias was then calculated as follows (Schinkel *et al*., 2010):

$$\text{Relative bias} = \left(e^{bias} - 1\right) \times 100\%,$$

where $\hat{y}_i$ = predicted (median) exposure level for the i^{th} set of exposure factors in the validation set (log transformed), y_i = measured exposure for the i^{th} set of exposure factors (log transformed), and n_0 = number of measurements in the validation set. The bias indicates a distance of the tool estimate from the true value. A positive bias implies overestimation by a tool compared to an exposure measurement; a negative bias implies underestimation. The smaller value of the relative bias indicates the more accurate results for the exposure estimation. Pearson correlation coefficients (r_p) were calculated to determine the relationship between exposure data and tool estimates (both log-transformed). In addition, a linear relationship between the exposure measurement results and the corresponding estimates was determined by performing linear regression analysis. All data analyses were performed using the Statistical Analysis Software (SAS) v. 9.4. In order to determine each tool's robustness, percentages of exposure measurement data exceeding the corresponding tool's estimates (%M > T) were calculated for each percentile estimate and the upper value of 90% CI of each percentile estimate (ART only).

It should be noted that the performance results of Stoffenmanager in this study cannot be compared directly against those from ART (or vice versa) because ART allows a combination up to four subtasks (and nonexposed periods) to be included whereas the version of Stoffenmanager that we used permits only one task to predict exposure estimates. Among 51 ESs, 24 ESs (~47%) included multiple subtasks with different control methods. For example, ES 19 (batch-making task) had four subtasks with control methods of no presence of local exhaust ventilation (LEV) and a fully enclosed system. For

ESs having multiple subtasks, we selected the least control method (i.e. no LEV present for ES 19) for Stoffenmanager to generate a conservative estimate, whereas all control methods for each subtask were applied for the ART.

The results were presented by exposure category, handling/activity description, and tool input parameters. In situations where multiple activity descriptions were available for the ART, the activity description showing the longest exposure time was selected to represent the activity for a task. When an ES had two subtasks having the same longest exposure time, a combined name of two activities was reported for the presentation purpose; e.g. HC&A represents a combination of handling activity of contaminated objects (HC) and activity with relatively undisturbed surfaces with no aerosol formation (A).

Results

Although the data analysis was performed for all exposure categories, only the results of liquids with VP > 10 Pa are summarized later, because these had a sufficient number of exposure measurements. Nonetheless, despite the results of the other exposure categories being inconclusive due to small sample sizes, these findings are still valuable and thus they are reported in Supplementary Tables 3 and 4 and Supplementary Figures 1 and 2 (available at *Annals of Work Exposures and Health* online) to provide additional information (without drawing any conclusions).

Description of workplace measurement data

Table 1 summarizes personal exposure measurement results for liquids with VP > 10 Pa. This category includes a wide range of exposure data from 0.07 mg m^{-3} to 6653 mg m^{-3} for task-based exposure levels and from 0.01 mg m^{-3} to 1455 mg m^{-3} for full-shift exposure levels, covering workplaces from small laboratories to heavy industries. A total of 42 ESs with 251 exposure measurements were included. In comparison, other exposure categories having a few ESs (2 or 3) and small sample sizes (≤11) showed narrower, lower ranges of exposure measurement results (Supplementary Table 2, available at *Annals of Work Exposures and Health* online). When the exposure measurement data of liquids with VP > 10 Pa were grouped by handling/activity description, the activity of handling of liquids using low pressure, low speed, or on a medium-sized surface (LPLS) had the most number of exposure measurements (n = 124) for Stoffenmanager (Table 2 and Fig. 1). For ART, the activity of spreading of liquid products (SLP) had the most number of measurements (n = 109; Table 3 and Fig. 2). By input parameter, medium VP and the

Table 1. Summary of the exposure measurement results and tools' estimates (liquids with VP > 10 Pa).

Tool	ES No.	n	Exposure measurement results			Tool's estimates								
						Summary of median estimates			r_p	Bias	Relative Bias (%)	%M > T		
			AM (mg m^{-3})	GM (mg m^{-3})	Range (mg m^{-3})	AM (mg m^{-3})	GM (mg m^{-3})	Range (mg m^{-3})				>50th (>UCI50)	>75th (>UCI75)	>90th (>UCI90)
Stoffenmanager[a]	42	251	214	24.5	0.07–6653	50.9	31.6	3.38–180	0.43*	0.3	29	46 (NA)	22 (NA)	15 (NA)
ART[b]	42	251	60.9	7.88	0.01–1455	21.4	3.73	<0.01–100	0.61*	−0.7	−53	65 (34)	53 (24)	41 (18)

Note that the performance between Stoffenmanager and ART is not comparable due to the use of different ESs for some ESs having multiple subtasks. ES No. = number of exposure situations (ESs) developed by NIOSH; n = number of personal exposure; AM = arithmetic mean; GM = geometric mean; r_p = Pearson correlation coefficient (log-transformed data); %M > T = percentage of exposure measurements exceeding the tool estimates; UCI50, UCI75, and UCI90 = the upper value of 90% confidence interval of the 50th, 75th, and 90th percentile estimates, respectively; NA = not available.

[a]Summary of exposure measurement results was based on task-based exposure levels calculated from the collected measurements.

[b]Summary of exposure measurement results was based on full-shift exposure levels.

*P-value < 0.05.

Table 2. Summary of the performance by handling description ($n > 10$) and by input parameter for liquids with VP > 10 Pa (Stoffenmanager).

Category			ES No.	n	r_p	Bias	Relative bias (%)	%M > T		
								>50th	>75th	>90th
Activity	LPHS		4	56	0.22	1.4	319	16	4	0
Description	LS		7	23	0.20	−3.0	−95	96	91	83
	SS		4	30	−0.11	0.8	116	30	10	0
	LPLS		16	124	0.76*	0.2	28	52	22	15
Input parameter	Vapor	High	6	37	0.60*	−0.5	−37	62	38	19
	pressure[a]	Medium	31	190	−0.09	0.1	13	47	20	16
		Low	5	24	−0.42*	2.4	1046	13	8	0
	LEV	Yes	7	50	0.74*	1.7	431	24	4	0
		No	35	201	0.23*	−0.1	−9	51	26	18

ES No = number of exposure situations (ESs) for which data were available; n = number of exposure measurements; r_p = Pearson correlation coefficient (log-transformed data); LPHS = handling of liquids (using low pressure, but high speed) without creating a mist or spray/haze; LS = handling of liquids on large surfaces or large work pieces; SS = handling of liquids on small surfaces or incidental handling of liquid; LPLS = handling of liquids using low pressure, low speed, or on medium-sized surfaces.

[a]Low vapor pressure: <500 Pa at room temperature; medium vapor pressure: 500 ≤ VP ≤ 10 000 Pa; high vapor pressure: VP > 10 000 Pa.

*P-value < 0.05.

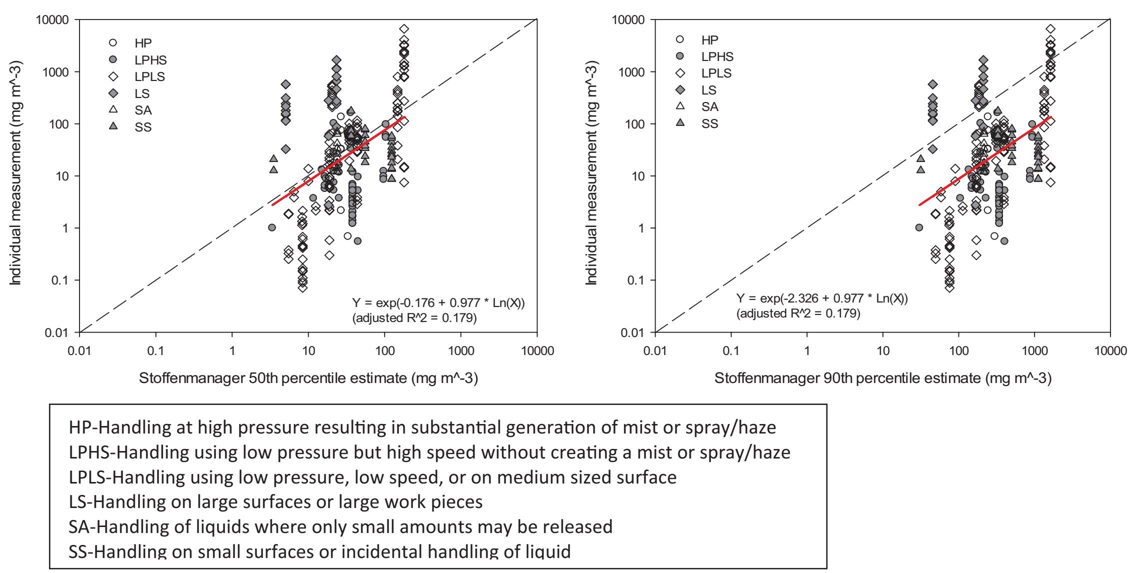

Figure 1. Comparison of exposure measurement results with the tool's 50th and 90th percentile estimates (Stoffenmanager; liquids with VP > 10 Pa). The solid line indicates a regression line and the dashed line indicates 1:1 line. Note that HP and SA activities were not included in Table 2 due to small sample sizes.

absence of LEV included the most number of exposure measurements.

Comparison of exposure measurement data with Stoffenmanager estimates

The results of Stoffenmanager exposure estimates and of the comparison between task-based exposure data and the corresponding tool's estimates are presented in Table 1 for liquids with VP > 10 Pa. The tool's median (50th percentile) estimates ranged from 3.38 to 180 mg m^{-3}, which is a narrower range compared to the range of exposure measurements, with a higher geometric mean (GM; 31.6 mg m^{-3}) than that of exposure measurement results (24.5 mg m^{-3}). Correlation

Table 3. Summary of the ART performance by activity description ($n > 10$) and by input parameter for exposures to liquids with VP > 10 Pa.

Category			ES No.	*n*	r_p	Bias	Relative bias (%)	%M > T		
								>50th (>UCI50)	>75th (>UCI75)	>90th (>UCI90)
Activity description	HCO		3	25	0.59*	−0.3	−26	72 (24)	52 (4)	36 (0)
	FL		12	62	0.89*	−1.4	−76	92 (44)	85 (18)	65 (6)
	HC&A		7	22	0.38	−3.4	−97	100 (86)	91 (82)	86 (73)
	SLP		15	109	−0.03	−0.6	−47	58 (29)	40 (28)	31 (22)
	NAF		1	15	**	2.2	785	0 (0)	0 (0)	0 (0)
Input parameter	Vapor Pressure[a]	High	6	37	0.53*	−0.05	−5	57 (16)	35 (3)	22 (0)
		Medium	31	190	0.14	−0.7	−51	62 (63)	51 (29)	39 (22)
		Low	5	24	0.88*	−2.1	−87	100 (34)	96 (27)	88 (13)

ES No. = number of exposure situations (ESs) for which data were available; n = number of exposure measurements; r_p = Pearson correlation coefficient (log-transformed data); UCI50, UCI75, and UCI90 = the upper value of 90% confidence interval of the 50th, 75th, and 90th percentile estimates, respectively; HCO = handling of contaminated objects or paste, FL = falling liquids; HC&A = combined two activities, handling of contaminated objects or paste and activities with relatively undisturbed surfaces (no aerosol formation); SLP = spreading of liquid products; NAF = activities with relatively undisturbed surfaces (no aerosol formation).

[a]Low vapor pressure: < 500 Pa at room temperature; medium vapor pressure: 500 ≤ VP ≤ 10 000 Pa; high vapor pressure: VP > 10 000 Pa.

*P-value < 0.05.

**Not calculated because of single ART estimation.

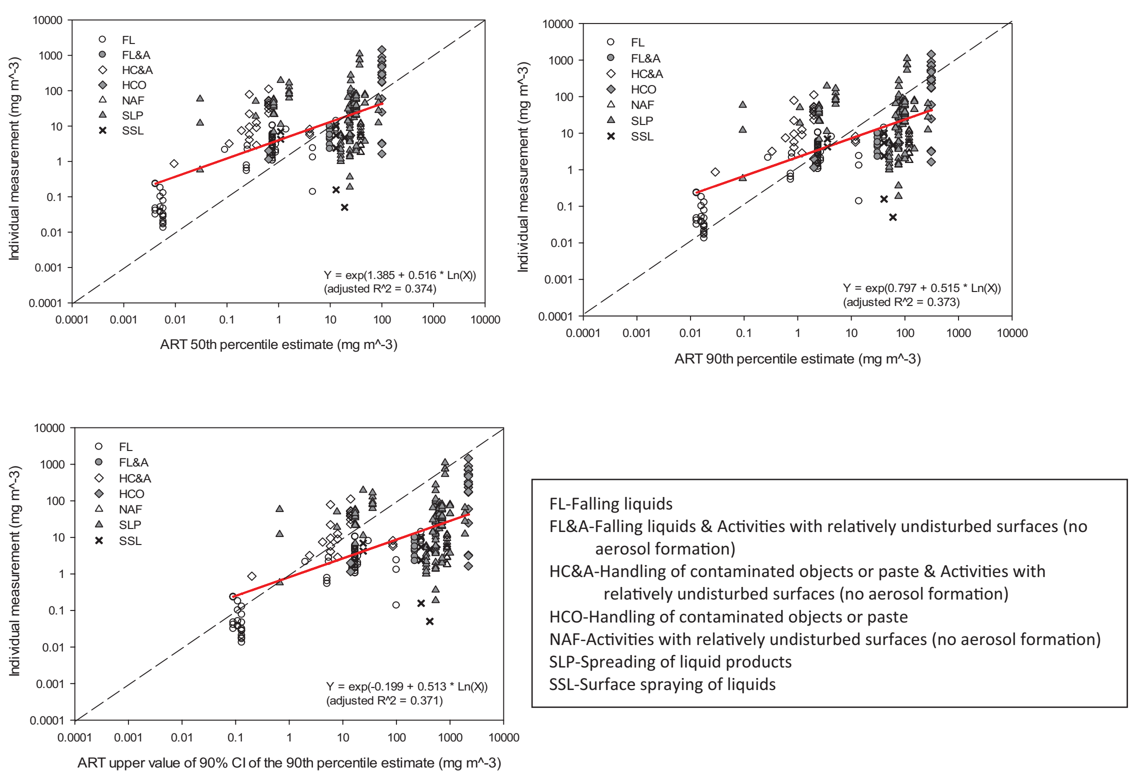

Figure 2. Comparison of exposure measurement results with the tool's estimates of 50th percentile, 90th percentile, and the upper value of 90% CI of the 90th percentile (ART; liquids with VP > 10 Pa). The solid line indicates a regression line and the dashed line indicates 1:1 line. Note that FL&A and SSL activities were not included in Table 3 due to small sample sizes.

between the log-transformed Stoffenmanager estimates and log-transformed exposure measurement results appeared to be moderate [Pearson correlation coefficient (r_p) = 0.43]. The tool showed a positive relative bias of 29%, indicating that on average the tool overestimated exposures when compared with the measured data. The percentage of exposure measurement results exceeding the tool's percentile estimates (%M > T) was 46%, 22%, and 15% for the 50th, 75th, and 90th percentile estimates, respectively. As shown in Fig. 1, the regression results indicate significant linear relationships between the exposure measurement data and estimates (P-value < 0.0001). The adjusted R^2 was less than 0.18 indicating high variability

Although Table 1 does not include the results for handling solid objects (3 ESs with n = 11), it should be noted that the tool predicted zero exposures for all 3 ESs (ES 8—wire extraction, ES 10—packing and shipping, and ES 12—bar feeding) because the tool assumed zero exposure emission for the selection of dust amount leading to zero exposure level (Supplementary Table 3, available at *Annals of Work Exposures and Health* online).

Table 2 shows a summary of the performance of the tool by handling description and by input parameter. Overall, Stoffenmanager underestimated exposures for the task of handling of liquids on large surfaces or large work pieces (LS) with a high %M > T (83%) for the 90th percentile estimates and a negative bias (−3.0). For the other tasks, the positive biases suggest that overall the tool overestimates exposures. The results of relative bias indicate that the tool was most accurate for the LPLS activity and least accurate for LPHS among four tasks considered in this study. The correlation was high for the LPLS (handling of liquids using low pressure, low speed, or on medium-sized surfaces) and either negative or low for the other tasks.

The %M > T for the 90th percentile was zero for the allocation of low VP (<500 Pa), whereas the %M > T for the high VP (> 10 000 Pa) and medium VP (500 ≤ VP ≤ 10 000 Pa) allocation was greater than 15% (Table 2). The tool's performance was least accurate for the low VP option and most accurate for the medium VP option. Negative correlations were observed for the allocation of medium and low VP. For the LEV input parameter, the tool's accuracy was better when LEV was absent compared to when LEV was present, but a negative bias in the former setting indicates that the tool is likely to underestimate exposure levels. The LEV presence showed a better robustness than LEV absence [%M > T = 0% (LEV present) versus 18% (LEV absent) for the 90th percentile estimates].

Comparison of exposure measurement data with ART estimates

Table 1 and Fig. 2 present the results of ART exposure estimates and of the comparison between exposure data and the corresponding tool's estimates. Compared to the range of exposure measurement results (0.01–1455 mg m^{-3}), the range of estimates was considerably lower (<0.01–100 mg m^{-3}); the AM and GM of estimates were also considerably lower than those of exposure measurement data. The correlation between the estimates and measurement data was moderate. The results of bias indicated that overall the tool underestimated exposures and the accuracy of the tool was −53%. When individual exposure data were compared with the tool estimates, the percent of measurement results exceeding the 50th, 75th, and 90th percentile estimates (%M > T) for liquids with VP > 10 Pa was 65%, 53%, and 41%, respectively, suggesting that the exposure variability may be underestimated (Table 1). In particular, ART estimates for this category appeared to underestimate at the lower exposure levels shown in Fig. 2. On the other hand, the %M > T for the upper value of 90% CI of each percentile estimate (i.e. UCI50, UCI75, and UCI90) was considerably decreased compared to that of percentile estimate alone (e.g. 65%M > T using the 50th percentile versus. 34%M > T using the UCI50). Statistically significant linear relationship with high variability (adjusted R^2 < 0.4) was observed when the ART 50th and 90th percentile estimates and the upper value of 90% CI of the 90th percentile estimates (P-values < 0.0001) were compared with the exposure measurement results (Fig. 2).

As shown in Table 3, overall, ART underestimated exposures for all selected tasks except activities with relatively undisturbed surfaces (no aerosol formation) (NAF), and the results of the comparison between individual measurement data and the tool's 90th percentile estimates also indicated underestimation of exposures (all %M > Ts greater than 30%). A moderate or high correlation was observed for all tasks except SLP and NAF. The relative bias for NAF was the highest among the tasks, but note that this was based on only one ES. The %M > Ts for the UCI50, UCI75, and UCI90 were considerably reduced when compared to those for the associated percentile estimates for handling of contaminated objects or paste (HCO) and falling liquids (FL) activities, whereas no substantial changes of %M > T were observed for the HC&A and SLP tasks.

The ART's performance for the high VP option was best among the VP input parameter options in terms of accuracy (Table 3). However, the results of negative biases demonstrated that the tool underestimated exposure levels. The %M > T for the upper value of 90% CI of the 90th percentile estimates was greater than 10% for the medium and low VP but less than 10% for the high VP. Note that an impact of the LEV input was not calculated for the ART because of the presence of multiple options. That is, ESs having multiple control strategies could not be simply allocated to either LEV presence or absence.

Discussion

Description of workplace measurement data

When the ESs were stratified by handling/activity descriptions, for some ESs, only a few exposure measurements (n < 3) were available depending on the size of the company and number of employees; only four handling tasks (LPHS, LS, SS, and LPLS) for Stoffenmanager and five activity tasks (HCO, FL, HC&A, SLP, and NAF) for ART included more than 10 exposure measurements for use in the comparisons. Therefore, it is necessary to collect additional exposure measurements for handling/activity descriptions and the other exposure categories (aqueous solutions, liquids with VP ≤ 10 Pa, powders, and solid objects) having limited sample sizes. Even for exposures to liquids with VP > 10 Pa, this study covered limited chemical agents; for example, most chemicals tested in the study had VP between 500 and 10 000 Pa (n = 190) and thus, it is still necessary to collect more exposure measurements for chemicals having VP less than 500 Pa and greater than 10 000 Pa. Nevertheless, the current study forms one of the most valuable evaluations for the tools because almost all contextual information required for the tools' input parameters was retrieved during the collection of exposure measurements.

Comparison of exposure measurement data with Stoffenmanager estimates

For exposures to liquids with VP > 10 Pa having a sufficient number of exposure measurements, a positive bias of 0.3 and a relative bias of 29% demonstrated, on average, an overestimation of the tool with good accuracy (Table 1). When individual exposure measurement results were compared with the tool's 90th percentile estimates, the proportion of M > T was 15%. An almost identical %M > T for the 90th percentile estimates was observed from the previous study (14%, n = 1326) by van Tongeren *et al.* (2017), whereas other prior studies reported either higher or lower %M > T than the 15%M > T in this study [i.e. 22%, n = 27, by Spinaze *et al.* (2017); 7%, n = 72, by Schinkel *et al.* (2010)]. Vink *et al.* (2010) evaluated Stoffenmanager v4.0 using 1-methoxypropan-2-ol measurements (n = 745 extracted from German MEGA database) collected during professional spraying paint works and reported that the tool's 75th percentile estimates were lower than the exposure measurement data for 3 out of 4 tasks. Landberg *et al.* (2015) compared tool estimates for exposure during eight tasks involving high- and low-volatility liquid chemicals and reported ~27%M > T (comparing median exposure measurements against the 90th percentile estimates), whereas this study showed a lower percentage exceedance. It is known that different percentiles of the exposure distribution represent variability (Tielemans *et al.*, 2011). The %M > Ts with 50th, 75th, and 90th percentile estimates were 46%, 22%, and 15%, respectively (Table 1), addressing that the tool's variability among different percentile estimates appeared to match very closely to the tool's predictions of the percentiles (e.g. about half of the exposure measurements were below than the 50th percentile estimate). Although the %M > T for the 90th percentile estimates was 15%, greater than 10%, combining the results with other percentile estimates indicates a good performance of the tool in terms of robustness. Overall, the tool appeared accurate and robust enough (based on the results of various parameters including bias, relative bias, correlation coefficient, and %M > T for the 50th, 75th, and 90th percentile estimates) to predict exposures for this category.

Although the results are presented in the online Supplementary Material (not evaluated in this article), for the category of solid objects, workers handled finished solid bars and wires for all three ESs (ESs 8, 10, and 12), and the tool assumed no exposure emission based on the physical form (i.e. zero score for intrinsic emission), generating modeled exposure levels of zero. In practice, we measured exposure ranges from <0.01 to 0.33 mg m^{-3} (GM 0.03 mg m^{-3}) for a task-based exposure, although the exposure ranges were low. The tool did not seem to apply a correct score for intrinsic emission for handling solid objects. The findings of this study clearly suggest that the tool's mechanism for predicting exposure during handling of solid objects requires review.

When considering the impact of handling description, a review of the algorithm of the LS task that resulted in underestimation of exposures (based on a negative bias of 3.0 and 83%M > T for the 90th percentile estimates) is indicated. Analysis of the impact of input parameters for liquids with VP > 10 Pa showed that the allocation of high or medium VP and the absence of LEV resulted in higher %M > Ts compared to the other allocation for each input parameter. van Tongeren *et al.* (2017) showed similar %M > Ts for the 90th percentile estimates for the medium (19%, n = 887) and low (5%, n = 131) VP inputs, whereas the %M > T for the high VP was lower (3%, n = 308) compared to this study (19%; Table 2). Interestingly, the allocation of LEV between the van Tongeren *et al.* and this study resulted in conflicting results [%M > T for the 90th percentile estimates of LEV presence versus absence were 0% versus 18% in this study and 30% versus 3% in the van Tongeren *et al.* (2017) study]. In this light, it may be useful for the tool developer to review the assigned scores for the options of vapor pressure and LEV inputs.

In summary, Stoffenmanager seems accurate and robust based on the results of various parameters, but there are areas for improvement of the tool. It should be noted that the tool tended to overestimate the real exposures because we adopted the least control method to predict estimates when multiple control methods were present. The Stoffenmanager algorithm we used did not allow subtasks and consequently a future study is planned to conduct another validation study using the multiple subtasks and compare the performance between Stoffenmanager and ART.

Comparison of exposure measurement data with ART estimates

For exposures to liquids with VP > 10 Pa, a negative bias and %M > Ts for the 50th, 75th, and 90th percentile estimates suggest that the tool underestimates exposure, in particular at the lower exposure levels. These findings are in agreement with Spinazze *et al.* (2017) who reported ~25% of exposure concentrations exceeding the tool's 90th percentile predictions (n = 28; exposure data extracted from previous studies) for organic solvents. However, Hofstetter *et al.* (2013) reported that the ART 50th percentile estimate was about 2.9 times higher than the exposure measurements of toluene in laboratory-based spray painting tasks. In this study, although no spraying tasks were involved, the findings by Hofstetter *et al.* contradict the present results that showed 50th percentile estimates to be considerably lower than the exposure measurements. The %M > Ts for the 50th, 75th, and 90th percentile estimates were 65%, 53%, and 41%, respectively, for this category (Table 1), showing that the %M > T does not appear to reduce sufficiently to match with the increasing percentiles (i.e. underestimation of exposure variability). Savic *et al.* (2017) also observed a similar conclusion when the ART 90th percentile estimates were compared with exposure data from a Swiss database; they suggested use of the upper 90% CI of the 90th percentile estimate. In this study, the %M > Ts using the UCI50, UCI75, and UCI90 estimates were considerably reduced compared to those using the associated percentile estimates [e.g. from 41% (90th percentiles) to 18% (UCI90)]. As explained in the Methods section, a CI of a percentile estimate explains the uncertainty around the percentile estimate, and this may be one way to adopt the tool's estimate. That is, users might be to consider using the upper limit value of 90% CI of the 90th percentile estimate in order to predict reasonable worst case situations.

All activities in Table 3 but NAF showed %M > Ts (for the 90th percentile estimates) greater than 30%, suggesting a need for the tool developers' further attention for these activities. For the NAF activity, results are inconclusive because only one ES was considered; however, the tool generally overestimated exposure. In addition, the tool's performance based on the choice of VP input parameter was determined to be poor, showing negative biases and greater than 20%M > T (for the 90th percentile estimates) for all options (high, medium, and low VP). When the comparison was made using the upper value of 90% CI of the each percentile estimate, activities including HCO and FL and the high and low VP input showed %M > Ts reduced to match with the increasing percentiles. The %M > Ts for the UCI50, UCI75, and UCI90 were not greatly reduced for HC&A and SLP activities and for the medium VP input, still requiring a review of these activities and input parameters.

On the basis of the findings of this study, it is evident that ART underestimates exposures in situations considered in the article and tool developers should focus on two potential underlying sources: (i) calibration of the ART algorithm and (ii) reviewing the assumptions on underlying exposure variability. Unfortunately, calibration errors (e.g. bias) cannot be determined in this study due to insufficient exposure measurements per ES. The ART estimate is derived by a log-normal mixed effects model with random effects representing three variabilities of between worker, within worker, and between company (Tielemans *et al.*, 2011; McNally *et al.*, 2014). The variabilities of between worker and within worker were adopted from geometric standard deviation values reported by Kromhout *et al.* (1993) and the between-company variability was taken from Symanski *et al.* (2006); Kromhout *et al.* included about 14 000 exposure measurements from 165 homogeneous exposure groups collected from 1974 to 1989 and Symanski *et al.* included a total of 49 807 measurements from 571 groups collected between 1993 and August 2005. Both studies included exposure measurement results with sampling duration over a full shift (≥4 h). Results from this study suggest that the assumed exposure variance within ART, which was based on old data sets, is possibly underestimated. The tool developers may wish to consider review of these variance components.

The current study evaluated only exposure estimates based on the mechanistic tool in ART. Combining this estimation with the tool's internal database or the users' own measurements using Bayesian statistics provided in ART can reduce its uncertainty (McNally *et al*., 2014). This would be a more practical approach for the tool users.

Conclusions

In this evaluation study, sufficient sample sizes were available only for exposures to liquids with VP > 10 Pa. For these exposures, it was determined that the accuracy of the ART's median predictions is similar to that of Stoffenmanager. However, the high proportions of exposure measurement results exceeding the ART percentile estimates are concerning and need to be addressed by the model developers by reviewing the algorithms, calibration, and/or assumed exposure variability. Both tools were developed based on the same concept of source–receptor approach by considering near- and far-field regions. However, the underlying method predicting exposure estimates is different; the Stoffenmanager estimate is based on log-normal, mixed effect regression models, with random between- and within-company components of variance, whereas the ART estimate is predicted by a log-normal mixed effects model (Tielemans *et al*., 2008) with random effects representing three variabilities of between worker, within worker, and between company (Tielemans *et al*., 2011; McNally *et al*., 2014). In addition, ART offers a broader range of input parameters compared to Stoffenmanager even for a single input variable (e.g. control strategy). Where all of the required contextual information for ART input parameters is unavailable, exposure assessors may therefore prefer to use Stoffenmanager. Nevertheless, Stoffenmanager still itself needs to be improved, particularly concerning the LS task and its input parameters for VP and LEV for this category. On the basis of the findings of this study, another option for ART might be to consider using the upper value of 90% CI of the 90th percentile estimate for predicting reasonable worst case situations, although some areas of activity and input parameter still need to be improved.

Although the results were not thoroughly summarized in this study, for the other exposure categories, the current study was limited to low ranges of exposure levels and insufficient numbers of exposure measurements from a small number of ESs (see the online Supplementary Material for the information). Further validation studies by covering a broader range of ESs with large exposure data sets will be required to provide insight on the performance for these tools.

In addition, the presence of potential errors caused from various uncertainties, such as in the interpretation of the input parameters, exposure measurement data, and inherent sources of variability cannot be totally excluded from our study. However, the uncertainty from exposure measurement results is likely rather negligible because all personal exposure measurements were sampled and analyzed according to well-established methodologies (e.g. NIOSH or OSHA methods). Similarly, the obtained consensuses that were used to run the tools minimize the uncertainty from the interpretation of the tool input parameters. Yet as expert judgment was still involved in the decision making, this process could become a source of error, which could not be determined in this study. Finally, the inherent uncertainty of the estimation process (e.g. scores of intrinsic emission, handling of materials, local controls, and general ventilation) for each tool could be a major source of error but cannot be determined from this study.

Supplementary Data

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

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention (CDC). Mention of any company or product does not constitute endorsement by NIOSH/CDC.

Funding

This project was funded internally by the National Institute for Occupational Safety and Health (Project CAN number: 939ZXEY).

Acknowledgements

We are sincerely thankful to Dr. Araceli Sanchez Jimenez (Institute of Occupational Medicine, Edinburgh, UK) and Dr. Jens Tørsløv (DHI groups, Hørsholm, Denmark) for reviewing this article prior to journal submission. We also would like to offer special thanks to the employees who participated in our study and the employers who allowed us to access their facilities. We acknowledge that two coauthors, M. van Tongeren and M. Tischer, were involved in the development of the Advanced Reach Tool (ART). We are also grateful for the permission granted by Cosanta BV (the legal owner of Stoffenmanager®) to use the Microsoft Access database in the current study, developed as part of the Eteam project.

References

Cherrie JW, Maccalman L, Fransman W *et al.* (2011) Revisiting the effect of room size and general ventilation on the relationship between near- and far-field air concentrations. *Ann Occup Hyg*; **55**: 1006–15.

European Chemicals Agency (ECHA). Chapter R.14: Occupational exposure assessment. *Guidance on Information requirements and chemical safety assessment*. ISBN: 978-92-9495-081-9. 2016, Helsinki, Finland: European Chemicals Agency. Available at https://echa.europa.eu/documents/10162/13632/information_requirements_r14_en.pdf. Accessed 20 June 2017.

Fransman W, Cherrie J, van Tongeren M *et al.* (2011) Development of a mechanistic 49 model for the Advanced REACH Tool (ART), *Ann Occup Hyg*; **55**: 957–979.

Hesse S, Hahn S, Schroeder K *et al.* (2015) Evaluation of Tier 1 Exposure Assessment Models under REACH (eteam) Project: Substudy report on gathering of background information and conceptual evaluation. Dortmund/Berlin/Dresden: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Available at http://www.baua.de/de/Publikationen/Fachbeitraege/F2303-D5.pdf?__blob=publicationFile&v=3

Hofstetter E, Spencer JW, Hiteshew K *et al.* (2013) Evaluation of recommended REACH exposure modeling tools and near-field, far-field model in assessing occupational exposure to toluene from spray paint. *Ann Occup Hyg*; **57**: 210–20.

Hornung RW. (1991) Statistical evaluation of exposure assessment strategies. *Appl Occup Environ Hyg*; **6**: 516–520.

Koppisch D, Schinkel J, Gabriel S *et al.* (2012) Use of the MEGA exposure database for the validation of the Stoffenmanager model. *Ann Occup Hyg*; **56**: 426–39.

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.

Landberg HE, Berg P, Andersson L *et al.* (2015) Comparison and evaluation of multiple users' usage of the exposure and risk tool: Stoffenmanager 5.1. *Ann Occup Hyg*; **59**: 821–35.

Lee E, Lamb J, Savic N *et al.* (2019) Evaluation of exposure assessment tools under REACH: Part I – Tier 1 tools. *Ann Work Expo Health*; **63**: 218–229.

Marquart H, Heussen H, Le Feber M *et al.* (2008) 'Stoffenmanager', a web-based control banding tool using an exposure process model. *Ann Occup Hyg*; **52**: 429–41.

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.

Savic N, Gasic B, Schinkel J *et al.* (2017) Comparing the Advanced REACH Tool's (ART) estimates with Switzerland's occupational exposure data. *Ann Work Expo Health*; **61**: 954–64.

Schinkel J, Fransman W, Heussen H *et al.* (2010) Cross-validation and refinement of the Stoffenmanager as a first tier exposure assessment tool for REACH. *Occup Environ Med*; **67**: 125–32.

Schinkel J, Fransman W, McDonnell PE *et al.* (2014) Reliability of the Advanced REACH Tool (ART). *Ann Occup Hyg*; **58**: 450–68.

Schinkel J, Ritchie P, Goede H *et al.* (2013) The Advanced REACH Tool (ART): incorporation of an exposure measurement database. *Ann Occup Hyg*; **57**: 717–27.

Schinkel J, Warren N, Fransman W *et al.* (2011) Advanced REACH Tool (ART): calibration of the mechanistic model. *J Environ Monit*; **13**: 1374–82.

Spinazzé A, Lunghini F, Campagnolo D *et al.* (2017) Accuracy evaluation of three modelling tools for occupational exposure assessment. *Ann Work Expo Health*; **61**: 284–98.

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.

Tielemans E, Noy D, Schinkel J *et al.* (2008) Stoffenmanager exposure model: development of a quantitative algorithm. *Ann Occup Hyg*; **52**: 443–54.

Tielemans E, Warren N, Fransman W *et al.* (2011) Advanced REACH Tool (ART): overview of version 1.0 and research needs. *Ann Occup Hyg*; **55**: 949–56.

van Tongeren M, Fransman W, Spankie S *et al.* (2011) Advanced REACH Tool: development and application of the substance emission potential modifying factor. *Ann Occup Hyg*; **55**: 980–8.

van Tongeren M, Lamb J, Cherrie JW *et al.* (2017) Validation of lower tier exposure tools used for REACH: comparison of tools estimates with available exposure measurements. *Ann Work Expo Health*; **61**: 921–38.

Vink SR, Mikkers J, Bouwman T *et al.* (2010) Use of read-across and tiered exposure assessment in risk assessment under REACH–a case study on a phase-in substance. *Regul Toxicol Pharmacol*; **58**: 64–71.