Annals of Work Exposures and Health, 2023, **67**, 402–413
https://doi.org/10.1093/annweh/wxac091
Advance access publication 3 January 2023
Original Article

The Chartered Society for Worker Health Protection

Evaluation of Stoffenmanager® and ART for Estimating Occupational Inhalation Exposures to Volatile Liquids

Eun Gyung Lee*

Field Studies Branch, Respiratory Health Division, National Institute for Occupational Safety and Health, 1095 Willowdale Road, Morgantown, WV, USA

*Tel: 1-304-285-6146; e-mail: dtq5@cdc.gov

Abstract

In practice, workers often handle the same chemical(s) of interest under different control measures (e.g. local ventilation, enclosed system) during a full shift. Stoffenmanager® allows users to predict either task-based or full-shift exposures. However, most previous studies evaluated the tool by comparing task-based exposures with measured exposures. Also, limited evaluation studies of the Advanced REACH Tool (ART) with the Bayesian approach (ART+B) are available, requiring additional evaluation studies. The performance of Stoffenmanager® and ART with and without the Bayesian approach was evaluated with measured full-shift exposures to volatile liquids in terms of accuracy, precision, and conservatism. Forty-two exposure situation scenarios (including 251 exposures), developed based on job tasks and chemicals handled during tasks from workplaces, were used to generate full-shift estimates. The estimates were then compared with measured exposures using various comparison methods. Overall, Stoffenmanager® appeared to be the most accurate among the testing tools, while ART+B was the most precise. The percentage of measured exposures exceeding the tools' 90th percentile estimates (%M>T) demonstrated that Stoffenmanager® (16%M>T) and ART+B (13%M>T) were more conservative than ART (41%M>T). When the 90% upper confidence limit of the 90th percentile estimate was considered, the level of conservatism changed from low (41%M>T) to medium (17%M>T) for ART and from medium (13%M>T) to high (0.8%M>T) for ART+B. The findings of this study indicate that no single tool would work for all ESs. Thus, it is recommended that users select a tool based on the performance results of three components (i.e. accuracy, precision, and conservatism), not depending on one or two components. The strength of this study is that the required tools' input parameters were obtained during the sample collection to minimize assumptions for many input parameters. In addition, unlike other previous studies, multiple subtasks, which happen often in workplaces, were incorporated in this study. Nevertheless, the present study did not cover all activities listed in the tools and was limited to volatile liquids, suggesting further studies cover other exposure categories (e.g. solid, metal) and diverse activities.

Keywords: Advanced REACH Tool (ART) with and without Bayesian approach; exposure assessment tools; inhalation exposure tools; REACH; Stoffenmanager®

What's important about this paper?

Stoffenmanager® and the Advanced REACH Tool (ART) are used to estimate workplace exposures via inhalation. This study explored the performance of these methods, including ART with Bayesian analysis, in 42 exposure scenarios, and found that accuracy, precision, and conservatism of the estimated exposures varied among the scenarios and tools. Collection of method input parameters during exposure monitoring or job observation is important, including for work subtasks, to minimize assumptions.

Received: September 6, 2022. Accepted: December 2, 2022.
Published by Oxford University Press on behalf of The British Occupational Hygiene Society 2023.

Introduction

The European Chemicals Agency (ECHA) suggested two-tiered approaches—Tier 1 and higher tier—to predict occupational exposure estimates to fulfill the Registration, Evaluation, Authorization, and restriction of Chemicals (REACH) regulation requirements (ECHA, 2016). A Tier 1 approach involves the use of conservative exposure tools including the European Centre for Ecotoxicology and Toxicology of Chemicals Targeted Risk Assessment (ECETOC TRA), Metals' Estimation and Assessment of Substance Exposure (MEASE), and an easy-to-use workplace control scheme for hazardous substances (EMKG-EXPO-TOOL) that can be easily used by inexperienced assessors. If assessors believe that the quantitative occupational exposure estimates of Tier 1 tools are too high, a higher tier approach designed to estimate exposures more accurately than the Tier 1 tools are recommended. Stoffenmanager® and the Advanced REACH Tool (ART) are known as higher-tier tools and were developed based on the same concept of a source-receptor approach. Both tools predict quantitative exposures based on a multiplicative model of several exposure determinants, where each determinant is categorized into exposure scores (dimensionless), and then the assigned score is put into a mixed effect regression model (internally calibrated) to estimate exposures (unit in milligrams per cubic meter, mg m^{-3}). The exposure estimates include the 50th, 75th, and 90th percentiles for Stoffenmanager® and the 50th, 75th, 90th, 95th, and 99th percentiles with confidence intervals (CIs) including interquartile, 80, 90, and 95% of each percentile estimate for ART. ECHA recommends using the 90th percentile estimate to predict reasonable worst-case situations (ECHA, 2016). Detailed information on the development of each tool is available elsewhere (Marquart *et al*., 2008; Tielemans *et al*., 2008, 2011; Cherrie *et al*., 2011; Fransman *et al*., 2011; Schinkel *et al*., 2011, 2013; 2014; van Tongeren *et al*., 2011; McNally *et al*., 2014; Schlüter *et al*., 2022a). In addition, Hesse *et al*. (2015) reviewed the background information and conceptual evaluation of Tier 1 tools and Stoffenmanager® to determine the level of confidence in terms of accuracy and reliability of the tools' estimates. They concluded that because the tools cannot be used for all realistic situations, users should be aware of their underlying concepts, strengths, and weaknesses. Later, Koivisto *et al*. (2021) reviewed the theoretical background of the Stoffenmanager® and ART tools and reported that both tools' algorithms do not exhibit the physical concept as originally presented by the tool developers. A subsequent response to the Koivisto *et al*. paper was provided by Fransman *et al*. (2022) criticizing that some statements in the Koivisto *et al*. paper were incorrect, such as unavailability of the calibration database, and the critique was biased. Recently, Schlüter *et al*. (2022b) reported the status of exposure modeling in Europe, challenges, and future needs as part of the activities of the exposure science modeling working group under the Europe Regional Chapter of the International Society of Exposure Science: the group presented sets of strategic objectives and action plans to improve models and its use and being adopted as regulatory requirements.

Many researchers have investigated the reliability of these tools by comparing the tools' estimates with measured exposures. Spinazzè *et al*. (2019) published a review paper on the current state of knowledge on Tier 1 and higher tier tools in terms of accuracy (i.e. bias or the mean difference between tools' estimates and measured exposures on a log scale), precision (i.e. variability of the bias), and conservatism (i.e. proportion of the measured exposures exceeding the tools' estimates), as well as identifying needs to be considered in the future. In their results, they reported Stoffenmanager® as the most balanced and robust tool based on the reported accuracy, precision, and conservatism and ART as generally the most accurate and precise tool along with a medium conservatism. These authors concluded that none of the tools was completely evaluated. They recommended comprehensive evaluation and validation studies of the tools, along with recalibration of the tools.

In a previous study conducted by the author and colleagues (Lee *et al*., 2019a), two higher-tier tools—Stoffenmanager® v4.5 and ART v1.5—were evaluated with measured personal exposures collected from workplaces along with the exposure determinants required to simulate the tools. They used 42 exposure situation (ES) scenarios containing 251 personal exposure measurements. Among the 42 ESs, 23 ESs (~55%), where each ES represents a task, included multiple subtasks with the same or different control methods while handling the same chemical. For example, a worker conducting a batch-making task in a paint manufacturing company performed four subtasks during his shift: (i) adding solid materials manually to a batch, (ii) transferring mixed chemicals to other containers using an automated system, (iii) manually cleaning the emptied batches either enclosed or with no ventilation present, and (iv) filling, mixing, and shipping the end products in fully enclosed systems. To predict Stoffenmanager® estimates, Lee *et al*. (2019a) used an algorithm developed for the Evaluation of Tier 1 Exposure Assessment Models (ETEAM) project under the permission of the legal owner of Stoffenmanager®. Because the algorithm was developed to consider only one task per ES, the reported estimates were based on one task (determined by the company) by applying the lowest control measure, while ART's estimates were predicted by considering multiple subtasks (up to four subtasks). Thus, Lee *et al*. (2019a) was not able to directly compare the performance results between Stoffenmanager® and ART. To the best of this author's knowledge, none of the previous studies has compared the performance of Stoffenmanager® and ART when workers perform multiple subtasks with different control measures during their shifts.

ART consists of two parts—the mechanistic model and the Bayesian approach. In the absence of exposure data, ART allows the prediction of inhalation exposure based on the mechanistic model only. Lee *et al*. (2019a) reported that the ART mechanistic model underestimated exposures compared with measured exposures. Other previous studies demonstrated either consistent (i.e. underestimation) or inconsistent (i.e. overestimation) results compared with the findings by Lee *et al*. (2019a). For example, Mc Donnell *et al*. (2011) reported an underestimation of exposures when the ART 50th percentile estimates were compared with the 50th percentile value of measured exposures. On the other hand, Hofstetter *et al*. (2013) found that ART overestimated exposure when ART 50th percentile estimates were compared with personal exposure data. Furthermore, Savic *et al*. (2017) and Lee *et al*. (2019a) suggested considering the 90% upper confidence limit (CL) of the 90th percentile estimate to assess exposures using ART. Surprisingly, only a few studies evaluated ART estimates that were predicted based on the combination of the mechanistic model and the Bayesian approach compared with the measured exposures (Landberg *et al*., 2018; LeBlanc *et al*., 2018; Ribalta *et al*., 2019). Overall, ART exposure estimates predicted with the Bayesian approach performed better than ART estimates predicted solely from the mechanistic part. Since limited evaluation studies of ART with the Bayesian approach are available, additional studies to support the previous findings are needed.

This study was conducted to evaluate the performance of Stoffenmanager® and ART by employing two approaches: (i) Stoffenmanager® estimates accounting for multiple subtasks if a task has more than one ongoing activity during handling the same chemical, and (ii) ART estimates with and without the Bayesian approach. The same ESs, along with the measured exposure data reported by Lee *et al*. (2019a), were used for this evaluation study. Also note that the ART's estimates without Bayesian approach and the measured full-shift exposures that were reported in Lee *et al*. (2019a) are used again in this study to compare with the estimates predicted using Stoffenmanager® and ART with Bayesian approach.

Methods

Exposure situation scenarios

The National Institute for Occupational Safety and Health (NIOSH) collected personal exposures to volatile liquids with vapor pressures (VPs) greater than 10 Pascal (Pa), along with the contextual information required for the tools' input parameters, from various workplaces (including paint manufacturing industries, hospital labs, print shops, and others) in the USA. For some workplaces, personal exposure measurements were collected only for the duration that workers were handling the chemicals of interest with a sampling time ranging from 32 to 712 min (88% of the measurements >240 min). For comparison purposes to the tools' estimates, the time-weighted average exposure data were converted to full-shift (i.e. 8 h) exposures by assuming zero exposure for the uncollected time.

After field surveys, the NIOSH senior industrial hygienist (IH) then developed ESs based on the job tasks and chemicals handled during the shift. An ES template used during the field survey is listed in Supplementary Table S1 (available at *Annals of Work Exposures and Health* online). For a task reporting exposure to multiple chemicals from the same sampling media, only the chemical exposure having the highest proportion in the mixture of products used was considered. The present study considered 42 ESs [number of exposure measurements (N) = 251], and due to the difficulties of finding workplaces, this study was limited to the exposure data of volatile liquids. Among 42 ESs, 23 ESs (~55%; N = 130) included two or more subtasks. Thirteen ESs (N = 75), 3 ESs (N = 33), and 7 ESs (N = 22) were comprised of two, three, and four subtasks, respectively. Detailed information about the type of industries, tasks, number of subtasks, and number of samples for each ES are listed in Supplementary Table S2 (available at *Annals of Work Exposures and Health* online) in Lee *et al*. (2019a). Additional information for the sampling campaigns and the development of the ESs are available elsewhere (2019b).

Generation of the tool estimates

After developing ESs, six assessors from the UK, Germany, Switzerland, and the USA, who are familiar with both tools, were gathered to make consensus decisions on all input parameters for each tool (by the NIOSH senior IH lead) (2019a). Once consensus decisions were made, the NIOSH senior IH estimated exposures for each ES using the specified tools below.

Stoffenmanager®

In this study, a web-based tool of Stoffenmanager® v7 (herein after referred to as Stoffenmanager®; https://stoffenmanager.nl/) was employed, and as indicated in the Introduction, up to four multiple subtasks were considered for each ES (if any) to predict full-shift exposure. For the direct comparison of the performance between Stoffenmanager® and ART, the 50th, 75th, and 90th percentile estimates of full-shift exposures were obtained.

ART mechanistic modeling

A web-based tool, ART v1.5 (herein after referred to as ART; https://www.advancedreachtool.com/), was used to estimate full-shift exposures using the mechanistic model only. The exposure estimates from this model were based upon users' inputs on exposure determinants, which later applied to a calibrated model (expressed as a lognormal mixed effects model), and variabilities of between-company, between-worker, and within-worker sources determined from previous literature (McNally *et al.*, 2014). In Bayesian terminology, this is referred to as 'prior' distributions. For each ES, the 50th, 75th, and 90th percentile estimates along with the 90% upper CL of each percentile estimate were obtained (Lee *et al.*, 2019a).

ART Bayesian modeling

ART allows the prior distributions to be updated to posterior distributions using measurement data collected on the exposure scenario (McNally *et al.*, 2014). Thus, inhalation exposure estimates using Bayesian approach (i.e. posterior distribution; herein after referred to as ART+B) were predicted using the output from the mechanistic model (i.e. prior distribution) of the ES and adding measured exposure data (i.e. likelihood distribution). The Markov Chain Monte Carlo (MCMC) sampling procedure, which is supported by the ART web tool, was employed. A user uploads measurement data to ART, the MCMC runs in the background, and summary statistics (i.e. point estimates of various percentiles and CL of those point estimates) can then be selected by the user. In the present study, a more efficient process was adopted to rapidly cycle through the ESs. The MCMC sampling was performed using R 4.0.2 (R Foundation for Statistical Computing) in conjunction with the R2Winbugs package, which is called a Winbugs encoding of the ART statistical model. For each ES, the model was parameterized using the mechanistic model estimate for the scenario, the mechanistic model uncertainty and variance components (McNally *et al.*, (2014), Tables 1 and 2), and the measurement data associated with the exposure scenario selected. In this study, an approach splitting the measured exposure data into two parts—one part to be used in the Bayesian approach (i.e. likelihood) and the other part to compare with ART+B outputs—was considered. However, each ES is unique in terms of task, chemical, ventilation, etc. and the sample size for each ES

Table 1. Comparison of the measured exposure data and tools' estimates by calculating bias, precision, and robustness (All ESs).

Tool	ES No (N)	Full-shift exposure measurement results (mg m^{-3})			Tool estimates (full-shift)							
					Median estimates (mg m^{-3})			Bias	Precision	%M>T		
		AM	GM	Range	AM	GM	Range			>50th (>UCL50)	>75th (>UCL75)	>90th (>UCL90)
Stoffenmanager®	42 (251)	60.9	7.88	0.01–1455	12.5	6.80	0.31–47.5	−0.42 (−0.50)[a]	1.81 (1.82)[a]	49 (NA)	28 (NA)	16 (NA)
ART	42 (251)				21.4	3.73	<0.01–100	−1.45 (−1.52)[a]	2.13 (2.03)[a]	65 (34)	53 (24)	41 (18)
ART+B	42 (251)				27.7	6.00	0.02–380	−0.62 (−0.70)[a]	1.10 (1.05)[a]	68 (27)	32 (6.0)	13 (0.8)

ES No, number of exposure situations (ESs) developed by NIOSH; *N*, number of personal exposure measurements; AM, arithmetic mean; GM, geometric mean; %M>T, percentage of exposure measurement data exceeding the tool estimates; UCL50, UCL75, and UCL90, the 90% upper confidence limit of the 50th, 75th, and 90th percentile estimates, respectively; NA, Not available.

[a]Bias and precision calculated using the median of tool estimates and median of the measured exposures for each ES.

Table 2. Summary of the performance by the vapor pressure input options.

Category		ES No	n	Tool	r_p	Bias	Precision	%M>T >50th (>UCl50)	>75th (>UCl75)	>90th (>UCl90)
Vapor pressure[a]	High	6	37	Stoffenmanager®	0.58	−0.38	0.94	70 (NA)	38 (NA)	19 (NA)
				ART	0.46	−0.02	0.67	57 (16)	35 (3)	22 (0)
				ART+B	0.92	−0.08	0.37	65 (24)	16 (3)	0 (0)
	Medium	31	190	Stoffenmanager®	0.57	−0.81	1.80	50 (NA)	28 (NA)	18 (NA)
				ART	0.74	−1.77	2.23	62 (34)	51 (29)	39 (22)
				ART+B	0.94	−0.88	1.16	68 (27)	35 (5)	14 (1)
	Low	5	24	Stoffenmanager®	−0.41	1.27	2.05	13 (NA)	8 (NA)	0 (NA)
				ART	−0.20	−1.76	0.65	100 (63)	96 (27)	88 (13)
				ART+B	0.94	−0.31	0.10	75 (29)	38 (17)	21 (0)

ES No, number of exposure situations (ESs) for which data were available; n, number of exposure measurements; r_p, Pearson correlation coefficient; %M>T, percentage of exposure measurement data exceeding the tool estimates; UCL50, UCL75, and UCL90, the 90% upper confidence limit of the 50th, 75th, and 90th percentile estimates, respectively; NA, Not available.

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

is too small to separate data into two parts. Among the 42 ESs, ~74% of ESs (31/42 ESs) has ≤6 exposure measurements and the evaluation study for ART+B using divided data might not become reliable because of a few samples used. Thus, it was decided to run the Bayesian approach with all measurement data for each ES. The MCMC was then run for 50 000 iterations following a burn-in of 1000 iterations, with every 20th iteration retained for inference to generate ART+B estimates. The same estimates as ART described above were obtained for the comparison.

Data analyses

Several methods were utilized to compare the tools' performance using all ESs combined and ESs divided into three VP groups (i.e. <500 Pa, 500 ≤ VP ≤ 10 000 Pa, and >10 000 Pa for the low, medium, and high VP, respectively). First, bias and precision were calculated to determine the level of agreement between the tools' estimates and measured exposures using the following equations:

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

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

where $\hat{y}$ = predicted (median) exposure level for the ith set of exposure factors in the validation set (log-transformed), y_i = measured exposure for the ith set of exposure factors (log-transformed), and n_0 = number of measurements in the validation set. A positive bias indicates an overestimation of the exposure while a negative bias indicates underestimation. Precision specifies the standard deviation (SD) of the bias. For each ES, the median (50th percentile) tool estimate and geometric mean (GM) of the measured exposures were used to compare the results with those from previous studies. In addition, the median values of tool estimate and measured exposures were employed to calculate bias and precision. For ESs having only one exposure measurement, that exposure level was used as the GM (or median).

Second, each tool's conservatism (i.e. robustness) was determined by calculating the percentage of individual exposure measurement data exceeding the corresponding tool's estimates (%M>T). For each ES, the tools' estimates of the 50th, 75th, and 90th percentiles and the 90% upper CL of each percentile estimate (for ART and ART+B) were compared with the measured data. The level of conservatism was considered low, medium, or high, if %M>T was >25%, >10%–25%, or ≤10%, respectively (van Tongeren *et al*., 2017).

Third, the relationship between the log-transformed exposure data (90th percentile) and log-transformed tool estimates (90th percentile) was determined by calculating the Pearson correlation coefficient (r_p) and performing linear regression analysis using SAS v9.4 (SAS Institute, Inc., Cary, NC).

Lastly, residuals were calculated by subtracting individual exposure data (log-transformed) from the tools' 90th percentile estimates (log-transformed) using the following equation:

$$\text{Residual} = \hat{y}_i - y_i \tag{3}$$

Here, the 90th percentile estimates of Stoffenmanager®, ART, and ART+B and the 90% upper CL of the 90th percentile estimate (UCL90; ART and ART+B only) were employed to determine if any pattern of tools' estimates (i.e. underestimation or overestimation) was observed.

Results

Table 1 presents a summary of the measured exposures and median estimates of Stoffenmanager®, ART, and ART+B for all ESs. The measured exposure data ranged from 0.01 mg m^{-3} to 1455 mg m^{-3}, showing a wide spread of data because sample collection was carried out across different chemicals and ESs. Each ES included a different number of exposure measurements ranging from 1 to 24; the number of measurements for each ES is provided in Supplementary Table S2 (available at *Annals of Work Exposures and Health* online) in Lee *et al.* (2019a). Unlike the measured exposures, the ranges of the tools' median estimates were smaller but still showed widespread distributions (e.g. 0.02–380 mg m^{-3} for ART+B).

The calculated bias using the GM of the measured exposures for each ES was negative for all tools, implying underestimation of exposures (Table 1). Among the three tools, Stoffenmanager® (–0.42) demonstrated the most accurate, while ART (–1.45) showed the least accurate. The bias of ART+B (–0.62) was close to that of Stoffenmanager®. The comparison of precisions indicated that ART+B (1.10) was the most precise and ART (2.13) was the least precise. No noticeable differences were observed when bias and precision were calculated using the median of measured exposures compared with those using the GM of measured exposures.

The percentages of measured exposures exceeding the tools' percentile estimates (%M>T) are reported in Table 1. Overall, regardless of which point estimates (i.e. 50th, 75th, and 90th percentiles) were compared, the %M>Ts for ART were considerably higher than those for Stoffenmanager® and ART+B, implying underestimation of exposure. With the tools' 90th percentile estimates, the levels of conservatism were medium for both ART+B (13%M>T) and Stoffenmanager® (16%M>T) and low for ART (41%M>T). When the %M>Ts using the UCL90 estimates were compared with that of the 90th percentile estimates, the conservatism changed from low (41%M>T) to medium (17%M>T) for ART and from medium (13%M>T) to high (0.8%M>T) for ART+B.

For all tools, statistically significant linear relationships between the measured exposures (90th percentile) and tools' 90th percentile estimates (all p-values <0.05) were observed (Fig. 1). Stoffenmanager® and ART showed higher variability (adjusted R^2 = 0.25 for Stoffenmanager® and 0.45 for ART) than ART+B (adjusted R^2 = 0.87). In addition, correlations between the log-transformed measured exposures (90th percentile) and the log-transformed estimates (90th percentile) were moderate for Stoffenmanager® (r_p = 0.52) and ART (r_p = 0.68) and high for ART+B (r_p = 0.94). When considering the impact of the chemicals' VP, the tool's performance for the medium VP (500 ≤ VP ≤ 10 000 Pa) input was similar to that without dividing ESs by VP (Table 2). For example, Stoffenmanager® and ART+B were determined as most accurate and precise, respectively, and the correlations between the measured exposures and the tools' estimates were moderate for Stoffenmanager® (r_p = 0.57) and ART (r_p = 0.74) and high for ART+B (r_p = 0.94). Regardless of VP inputs, ART+B was most precise and showed higher correlation compared to the other tools. The %M>Ts by the input parameter of VP demonstrated mixed results (Table 2). For example, the %M>T for the 90th percentile for Stoffenmanager® showed medium conservatism for the high and medium VP options and high conservatism for the low VP. On the other hand, the conservatism of ART+B was medium for the medium and low VP options and high for the high VP input. Overall, %M>Ts for ART were always higher than those for Stoffenmanager® and ART+B, showing either medium or low conservatism.

Figure 2 presents residual plots of the differences in the tools' estimates and measured exposures (both log-transformed) as a function of the measured exposures by the VP input. The log mean ± SD differences of all combined data were 2.05 ± 1.97, 0.34 ± 2.24, and 1.13 ± 1.30 for Stoffenmanager®, ART, and ART+B, respectively, when the tools' 90th percentile estimates were compared with the individual exposure measurement data. The log mean difference shifted up when the UCL90 estimates were compared with the measured data for ART (2.38 ± 2.24) and ART+B (2.27 ± 1.52). Overall, a clear pattern of overestimating low exposures and underestimating high exposures was observed regardless of the VP input options.

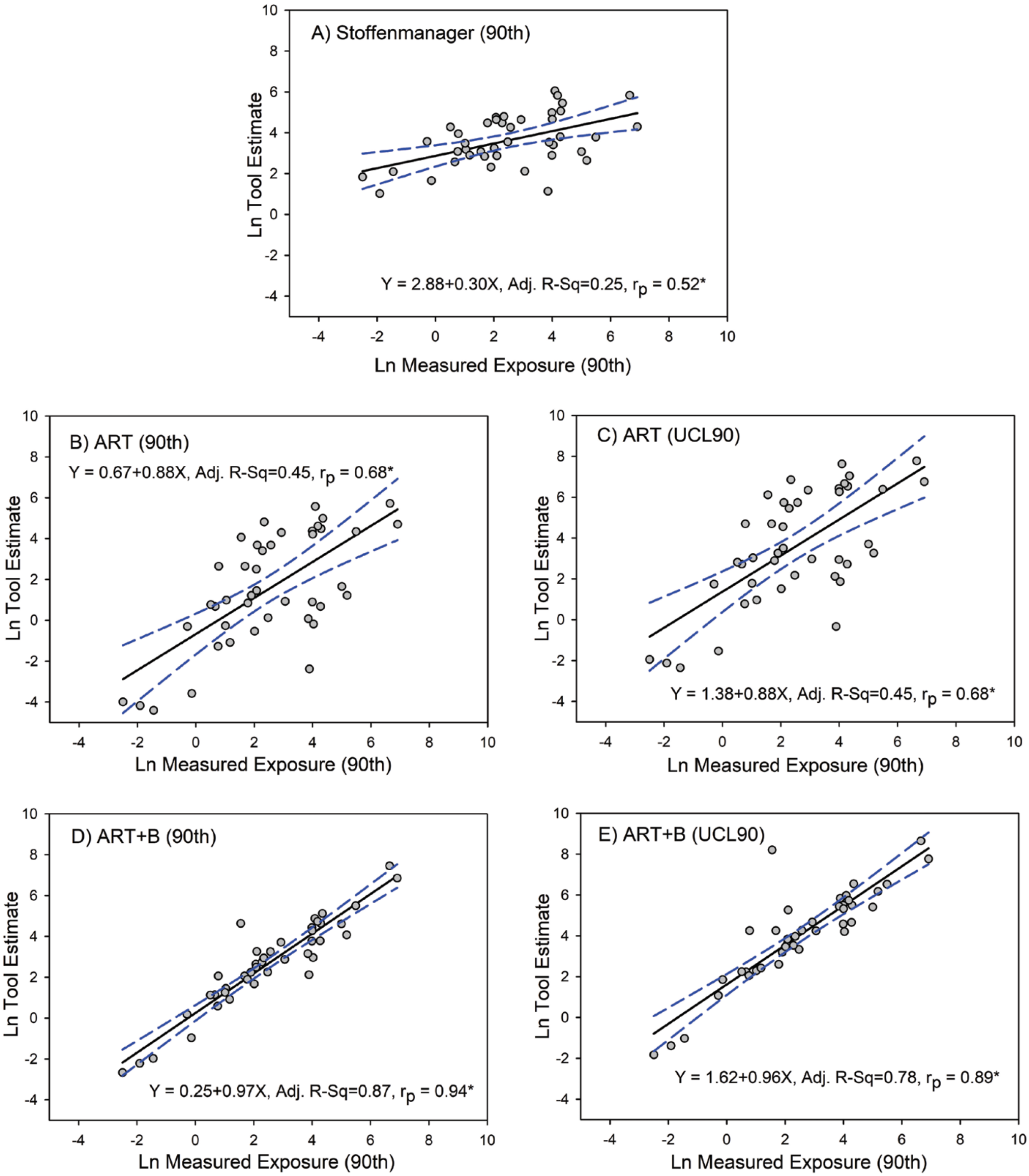

Figure 1. Comparison of the tool's estimates with the 90th percentile values of the measured exposures (both log-transformed): (A) Stoffenmanager (90th), (B) ART (90th), (C) ART [90% upper confidence limit of the 90th percentile estimate (UCL90)], (D) ART+B (90th), and (E) ART+B (UCL90). The solid line indicates a regression line, and the medium dashed lines indicate the 95% confidence limits; r_p and * represent Pearson correlation coefficient and p-value <0.05, respectively.

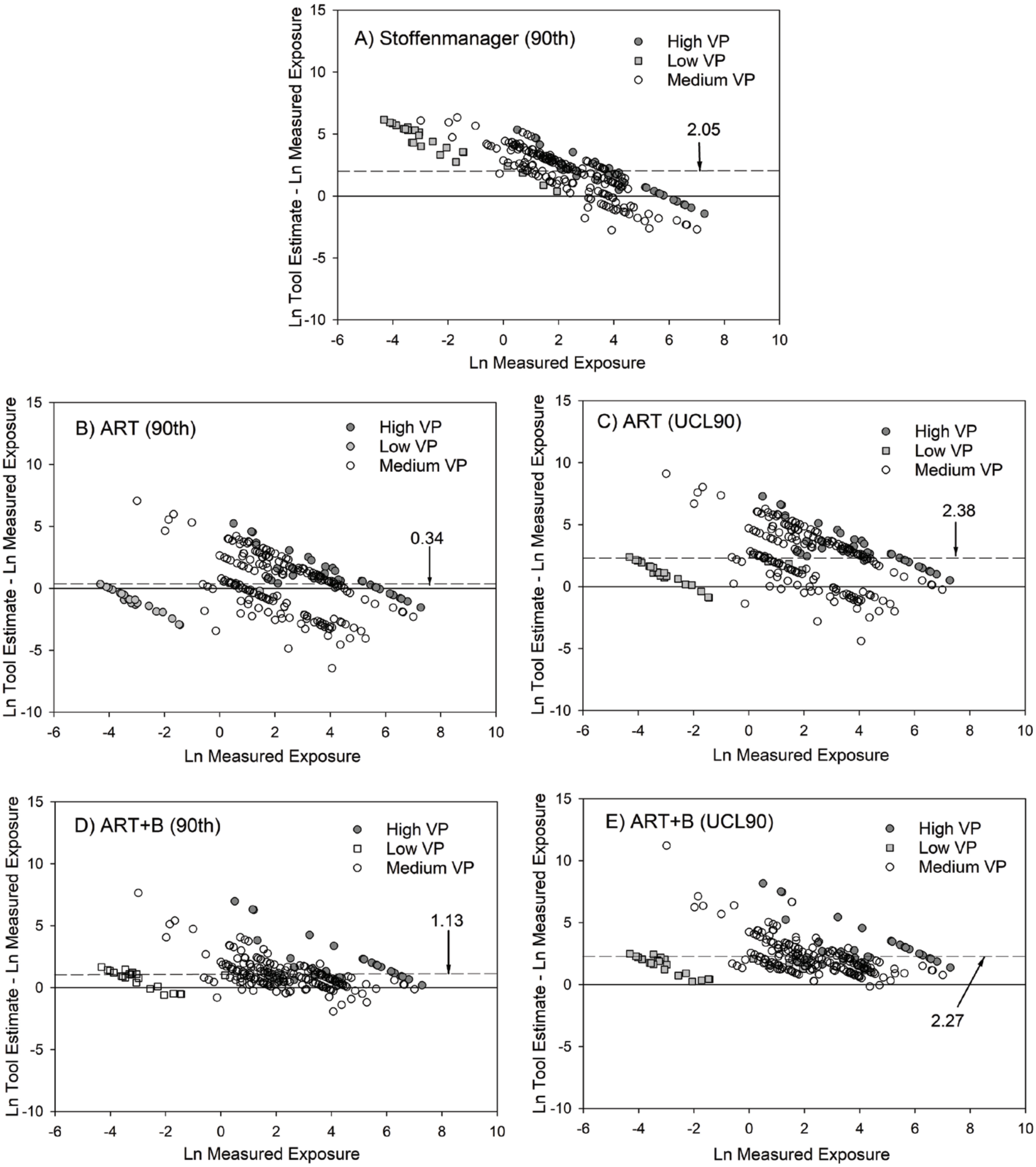

Figure 2. Residual plots of the differences in each tool's estimates and measured exposures (both log-transformed) as a function of the measured exposures: (A) Stoffenmanager (90th), (B) ART (90th), (C) ART [90% upper confidence limit of the 90th percentile estimate (UCL90)], (D) ART+B (90th), and (E) ART+B (UCL90). The dashed line indicates an overall mean difference of the tool's estimates and measured exposures.

Discussion

The performance of Stoffenmanager®, ART, and ART+B was evaluated in terms of accuracy (i.e. bias), precision, correlation, and conservatism by comparing the tools' estimates with measured exposures. The findings of ART's estimates without the Bayesian approach (i.e. ART) have been already reported by Lee *et al.* (2019a) and are repeated here for comparison with those of other tools. One exception is that the reported biases and precisions in the previous study (Lee *et al.* (2019a), Table 1) and in the present study are different because the current study used the GM values of measured exposures (instead of using individual exposure measurement data) for the calculations. The GMs were used here for comparison with the results of other previous studies (Landberg *et al.*, 2017; Lee *et al.*, 2019c; Savic *et al.*, 2020). It should also be noted that the Stoffenmanager® estimates in the previous study (Lee *et al.*, 2019a) were task-based and compared with task-based exposure data. In this study, the estimates of Stoffenmanager® represent daily exposures and thus are compared with the full-shift exposure data. Multiple subtasks were not considered in the exposure estimation of Stoffenmanager® in Lee *et al.* (2019a), whereas the ART estimates reflected subtasks. Thus, no direct comparison of the performance between Stoffenmanager® and ART was conducted in the previous study, as they are in the present study.

The bias results using the 50th percentile estimates of tools and the GM of the measured exposures demonstrated that, overall, all tools underestimated exposures. Among these, Stoffenmanager® was the most accurate compared with ART and ART+B (Table 1). ART+B was then determined to be the most precise compared with the others. No consistent observations were reported by previous studies when the comparison was made using the same percentile estimates (50th percentile) and exposure data (GM). Savic *et al.* (2020) observed an overestimation of exposures for both Stoffenmanager® and ART, while Lee *et al.* (2019c) reported an underestimation of exposures. Landberg *et al.* (2017) then showed an overestimation of exposures for Stoffenmanager® and an underestimation for ART for liquids only. For accuracy, Savic *et al.* (2020) also observed the same result as the present study, which is that Stoffenmanager® is more accurate than ART (bias = 0.09 for Stoffenmanager® and 0.62 for ART; no precision was calculated). Conversely, Lee *et al.* (2019c) reported that ART is more accurate than Stoffenmanager®, although Stoffenmanager® is more precise than ART for the exposure data collected from 10 organic solvents. It is unclear why the results among different studies were inconsistent. The combined effect of different types of compounds, tasks, and exposure ranges (ordinary work vs. maintenance) might be factors affecting the observed differences. For example, in this study, only liquids with VPs greater than 10 Pa were used, while other studies investigated the tools' performance using other types of compounds (e.g. wood dust, powders, volatile, and non-volatile liquids). This is the first study to report the lack of agreement (bias and precision) using estimates of ART+B. Overall, the findings of this study showed Stoffenmanager® as the most accurate tool and ART+B as the most precise tool compared with the other tools (Table 1).

The level of conservatism (i.e. %M>Ts for the 90th percentile estimates) was medium for Stoffenmanager® and ART+B and low for ART. For Stoffenmanager®, the same level of conservatism was reported from previous studies (Koppisch *et al.*, 2012; Spinazzè *et al.*, 2017; Van Tongeren *et al.*, 2017; Landberg *et al.*, 2018; Lee *et al.*, 2019b, 2019c) when the 90th percentile estimates were compared with the measured exposures (%M>Ts range of previous studies: 11–22%). Other previous studies showed either high conservatism [7%M>T by Schinkel *et al.* (2010) and Koppisch *et al.* (2012)] or low conservatism [27%M>T by Landberg *et al.* (2015). For ART, the low level of conservatism was in agreement with Spinazzè *et al.* 2017; 29%M>T]. When considering the UCL90 estimates of ART, %M>T was noticeably reduced from 41 to 17%. Surprisingly, when the measured exposure data were used to simulate the 90th percentile estimates (i.e. ART+B), the conservatisms changed from low to medium and high for the UCL90 estimates (Table 1). Landberg *et al.* (2018) reported 5%M>T using the 95% upper CL of the 90th percentile estimate for ART and 0%M>T for ART+B. LeBlanc *et al.* (2018) predicted benzene exposures during parts washing using ART and ART+B and found that the 50th percentile estimate of ART+B was closer to the measured exposure value than ART. Ribalta *et al.* (2019) found that ART underestimated exposures (three out of five cases) while ART+B overestimated all five cases with factors up to six. The findings of the present study appeared to be the same as those reported by the previous studies.

In addition, the results of %M>Ts for the 50th, 75th, and 90th percentile estimates showed that the variability of Stoffenmanager® seemed to agree with the predictions of each point percentiles. That is, about 49%, 28%, and 16% of the measured exposures were below the 50th, 75th, and 90th percentile estimates, respectively. For ART, the %M>Ts showed no agreement of variability regardless of which point estimate (i.e. 50th, 75th, and 90th percentiles) was compared. Tielemans *et al.* (2011) stated that ART's percentile estimates of the exposure distribution (i.e. 50th, 75th, 90th, 95th, or 99th) included exposure variability (i.e. combined results of a calibrated model and variabilities of between-company, between-worker and within-worker sources), while CL for each percentile (i.e. interquartile, 80, 90, or 95% of each percentile) represented exposure uncertainty. When the uncertainty of each percentile estimate was included (i.e. obtaining UCL values), the %M>Ts were 34%, 24%, and 17% for the UCL50, UCL75, and UCL90, respectively, indicating good agreement between the tool's estimates and measured exposures. Thus, as discussed in Savic *et al.* (2017) and Lee *et al.* (2019a), users are encouraged to adopt the UCL90 estimate by including the uncertainty around the percentile estimate.

McNally *et al.* (2014) mentioned that ART's uncertainty can also be reduced by combining the results of the mechanistic model with exposure measurement data (either from the tool's internal database or users' own data) using the Bayesian approach provided in ART. In this study, an agreement between the 90th percentile estimates of ART+B and measured exposures was considerably improved compared with ART (%M>T = 41% for ART and 13% for ART+B).

As shown in Fig. 1, the relationships between the tools' estimates and measured data (both log-transformed) demonstrated statistically significant differences for all comparisons (all P-values <0.001) with a moderate correlation for Stoffenmanager® and ART and high correlation for ART+B. Also notable is a systematic tendency of overestimating low exposures and underestimating high exposures observed for all three tools, which agrees with previous studies (Bekker *et al.*, 2016; Landberg *et al.*, 2017; Savic *et al.*, 2017, 2020; Lee *et al.*, 2019c, 2020). This observation suggests that the tool developers might review calibration results embedded in the tools and uncertainty factors applied to estimate predictions.

When considering the impact of VP input, the reported results for Stoffenmanager® were not considerably different compared to those by Lee *et al.* (2019a). For example, Stoffenmanager® showed medium conservatism for the allocation of high or medium VP and low for the low VP option. van Tongeren *et al.* (2017) reported similar conservatisms for the medium (19%, n = 887) and low (5%, n = 131) VP inputs, while a better conservatism (3%M>T) for the high VP was shown compared to this study (19%M>T; Table 2). Regardless of the VP options, ART's performance was poorer compared to the other tools. ART+B demonstrated high conservatism for the medium and low VP and high for the high VP. It should be noted that the sample size of ESs along with the number of exposure measurements were substantially different among the VP groups [i.e. 31 ESs (medium VP) versus 5 or 6 ESs (high and low VP); Table 1]. To firmly conclude these results, additional number of ESs and exposure measurements would be needed.

Overall, although Stoffenmanager® is less precise than ART+B, it is more accurate than ART and ART+B and shows the same medium level of conservatism as ART+B and moderate correlation when compared with the measured exposure. These results indicate that Stoffenmanager® is sufficient to be used for predicting inhalation exposures to volatile liquids. Stoffenmanager® has an advantage compared with ART and ART+B as it requires fewer input parameters. That is, Stoffenmanager® can be easily used by users who don't have advanced knowledge in occupational exposure assessment. In the present study, v7 was used and currently v8.3 is available. The major difference is that v8.3 can perform risk assessments at different process temperatures, which is irrelevant to the exposure data utilized in this study. The results of exposure estimates (without considering a respiratory protection factor) would be the same for both v7 and v8.3. Thus, the findings of this study can be applied to an evaluation study with the current version.

Similarly, ART+B can also be a promising tool because, although it is less accurate than Stoffenmanager®, it is the most precise and has the same conservatism as Stoffenmanager®. If exposure measurement data are available (even a few data), users are encouraged to predict inhalation exposures with ART+B. However, it should be noted that in this study, the same exposure data used for the comparison of the tools' estimates with the measured data were incorporated into the Bayesian approach to predict estimates, potentially leading to artificially good performance. This is a limitation of this study. If researchers are interested in evaluating the performance of ART+B, it would be ideal to use a different set of data from similar ESs into the Bayesian approach. Alternatively, users are encouraged to consider the 90% upper CL of the 90th percentile estimate.

Conclusions

The present study investigated the performance of Stoffenmanager® and ART with and without the Bayesian approach in terms of accuracy, precision, and conservatism. In conclusion, the findings of this study indicate that the 90th percentile estimates of Stoffenmanager® and ART+B are sufficiently accurate, precise, and conservative with the measured exposures for estimating inhalation exposures of volatile liquids. Alternatively, it is strongly recommended that users select the 90% upper CL of the 90th percentile estimate for ART if only the mechanistic part is considered. In addition, a systemic pattern (i.e. overestimating low exposures and underestimating high exposures) for all tools' estimates indicate that the tool developers might want to revisit their calibration results embedded in the tools and uncertainty factors applied to the development of tools. The strengths of this study are that the required tools' input parameters were obtained during the sample collection, and unlike other previous studies, more realistic activities (such as considering multiple subtasks) were considered in the estimation of exposures using higher-tier tools recommended by ECHA. In future studies, other types of materials (e.g. solid, metals), ES scenarios not covered in this study, and a wide range of exposure levels should be considered.

Acknowledgments

The author is sincerely thankful to Dr. Kevin McNally in Health and Safety Laboratory, Buxton, UK for his assistance in processing ART Bayesian analyses. The author also would like to thank Drs. M. Abbas Virji (Field Studies Branch/Respiratory Health Division/NIOSH) and Melissa Edmondson (Risk Evaluation Branch/Division of Science Integration/NIOSH) for reviewing this manuscript prior to journal submission.

Funding

This project was funded internally by the National Institute for Occupational Safety and Health (Project CAN number: 9278193).

Disclaimer

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

Conflict of interest

The author declares no conflict of interest.

Data Availability

Data request can be made to Eun Gyung Lee (DTQ5@CDC.GOV) for those interested in accessing the NIOSH exposure data.

Supplementary Data

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

References

Bekker C, Voogd E, Fransman W *et al*. (2016) The validity and applicability of using a generic exposure assessment model for occupational exposure to nano-objects and their aggregates and agglomerates. *Ann Occup Hyg*; **60**: 1039–48.

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.

ECHA. (2016) *Guidance on information requirements and chemical safety assessment-Chapter R.14: occupational exposure assessment (Version 3.0). Book guidance on information requirements and chemical safety assessment-Chapter R.14: occupational exposure assessment (Version 3.0)*. Helsinki, Finland: European Chemicals Agency.

Fransman W, Arnone M, Borghi F *et al*. (2022) Response letter to Koivisto *et al*. 'Evaluating the theoretical background of STOFFENMANAGER® and the advanced REACH tool'. *Ann Work Expo Health*; **66**: 543–49.

Fransman W, Van Tongeren M, Cherrie JW *et al*. (2011) Advanced Reach Tool (ART): development of the mechanistic model. *Ann Occup Hyg*; **55**: 957–79.

Hesse S, Schroeder K, Mangelsdorf I *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, Germany, BAUA: Federal Institute for Occupational Safety and Health.

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.

Koivisto AJ, Jayjock M, Hämeri KJ *et al*. (2021) Evaluating the theoretical background of STOFFENMANAGER® and the advanced REACH tool. *Ann Work Expo Health*; **66**: 520–36.

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.

Landberg HE, Axmon A, Westberg H *et al*. (2017) A study of the validity of two exposure assessment tools: Stoffenmanager and the advanced REACH tool. *Ann Work Expo Health*; **61**: 575–88.

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.

Landberg HE, Westberg H, Tinnerberg H. (2018) Evaluation of risk assessment approaches of occupational chemical exposures based on models in comparison with measurements. *Saf Sci*; **109**: 412–20.

LeBlanc M, Allen JG, Herrick RF *et al*. (2018) Comparison of the near field/far field model and the advanced REACH tool (ART) model V1. 5: exposure estimates to benzene during parts washing with mineral spirits. *Int J Hyg Environ Health*; **221**: 231–8.

Lee EG, Lamb J, Savic N *et al*. (2019a) Evaluation of exposure assessment tools under REACH: part II—higher tier tools. *Ann Work Expo Health*; **63**: 230–41.

Lee EG, Lamb J, Savic N *et al*. (2019b) Evaluation of exposure assessment tools under REACH: part I—tier 1 tools. *Ann Work Expo Health*; **63**: 218–29.

Lee S, Choi S, Lee K. (2020) Evaluation of Stoffenmanager and a new exposure model for estimating occupational exposure to styrene in the fiberglass reinforced plastics lamination process. *Int J Environ Res Public Health*; **17**: 4486–4506.

Lee S, Lee K, Kim H. (2019c) Comparison of quantitative exposure models for occupational exposure to organic solvents in Korea. *Ann Work Expo Health*; **63**: 197–217.

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.

Mc Donnell PE, Schinkel JM, Coggins MA *et al.* (2011) Validation of the inhalable dust algorithm of the Advanced REACH Tool using a dataset from the pharmaceutical industry. *J Environ Monit*; **13**: 1597–606.

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.

Ribalta C, López-Lilao A, Estupiñá S *et al.* (2019) Health risk assessment from exposure to particles during packing in working environments. *Sci Total Environ*; **671**: 474–87.

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.

Savic N, Lee EG, Gasic B *et al.* (2020) TREXMO plus: an advanced self-learning model for occupational exposure assessment. *J Expo Sci Environ Epidemiol*; **30**: 554–66.

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.

Schlüter U, Arnold S, Borghi F *et al.* (2022a) Theoretical background of occupational-exposure models—Report of an expert workshop of the ISES Europe working group "Exposure Models". *Int J Environ Res Public Health*; **19**: 1234–46.

Schlüter U, Meyer J, Ahrens A *et al.* (2022b) Exposure modelling in Europe: how to pave the road for the future as part of the European Exposure Science Strategy 2020-2030. *J Expo Sci Environ Epidemiol*; **32**: 499–512.

Spinazzè A, Borghi F, Campagnolo D *et al.* (2019) How to obtain a reliable estimate of occupational exposure? Review and discussion of models' reliability. *Int J Environ Res Public Health*; **16**: 2764–2792.

Spinazzè 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.

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.