Annals of Work Exposures and Health, 2019, Vol. 63, No. 7, 814–820
doi: 10.1093/annweh/wxz040
Advance Access publication 23 May 2019
Short Communication

Short Communication

Inter-assessor Agreement for TREXMO and Its Models Outside the Translation Framework

Nenad Savic,[1,*] Eun Gyung Lee,[2] Bojan Gasic,[3] and David Vernez[1]

[1]Centre for Primary Care and Public Health (Unisanté), University of Lausanne, CH-1066 Epalinges-Lausanne, Switzerland; [2]National Institute for Occupational Safety and Health (NIOSH), Health Effects Laboratory Division (HELD), Exposure Assessment Branch (EAB), 1095 Willowdale Road, Morgantown, WV 26505, USA; [3]Chemicals and Occupational Health Unit, Swiss State Secretariat for Economic Affairs (SECO), Holzikofenweg 36, CH-3003 Bern, Switzerland

*Author to whom correspondence should be addressed. Tel: +41-21-314-37-82; e-mail: nenad.savic@chuv.ch

Submitted 30 January 2019; revised 23 April 2019; editorial decision 29 April 2019; revised version accepted 3 May 2019.

Abstract

Within the framework of Registration, Evaluation, Authorization, and restriction of CHemicals (REACH), occupational exposure models are often used to predict the levels of exposure at a workplace. Poor inter-assessor agreement with their use poses a concern that may lead to different and dangerous risk conclusions. TREXMO (v1, 2016), a web tool providing parameter translations between six exposure models under REACH, is expected to improve the inter-assessor agreement. In this study, 18 assessors performed exposure assessment for six exposure situations within and outside the framework of this translation tool. In more than half of the evaluated cases, the results showed better agreement between assessors selecting the exposure parameters within the framework of TREXMO than when manually coding. The most affected were the parameters related to activity (such as "handling types" of Stoffenmanager) and exposure control (such as local controls). Furthermore, the agreement between the estimates calculated by different assessors was also improved when performing the translations between the models. For Stoffenmanager, for example, the relative standard deviation of 70–121%, obtained for vapors without applying the translation system, was 29–94% with the translations from ART. These findings showed a potential of TREXMO to impact the inter-assessor agreement. Because the study was limited to 18 assessors and only six exposure situations were assessed, further investigations are suggested.

Keywords: Advanced REACH Tool; ECETOC TRA; inter-rater agreement; occupational exposure models; Stoffenmanager; TREXMO

Introduction

Occupational exposure models, within the context of the Registration, Evaluation, Authorization and restriction of CHemicals (REACH) legislation (ECHA, 2016), are often applied by registrants to establish exposure scenarios. However, the different educational backgrounds and levels of experience of assessors evaluating the same set of conditions in an exposure scenario may strongly affect the reliability of their exposure estimates (Kromhout *et al.*, 1987; Cherrie and Schneider, 1999; Friesen *et al.*, 2011).

Inter-assessor agreement presents a degree of concurrence between the assessors evaluating the same exposure situation (ES) (Fleiss, 1971). Previous studies (Schinkel *et al.*, 2014; Landberg *et al.*, 2015; Lamb *et al.*, 2017) have reported variations by several orders of magnitude between the exposure estimates calculated by different assessors. Schinkel *et al.* (2014) compared assessors' agreement regarding the Advanced REACH Tool (ART) and found that training improved slightly the agreement. However, substantial variation between the assessors was observed in evaluating the same exposure situation, mostly due to the variations in values chosen for activity-related parameters. Landberg *et al.* (2015) investigated the agreement for Stoffenmanager and reported significant variations for activity-related parameters and personal protection. More recently, a broader study, known as the eteam (Lamb *et al.*, 2017), evaluated the inter-assessor agreement for several models, that is, EMKG-EXPO-TOOL, MEASE, ECETOC TRAv2 (an older version of the model; ECETOC, 2009) and TRAv3 (ECETOC, 2012) and Stoffenmanager. Again, the authors found significant variations between the assessors, particularly when assessing the activity-related parameters in the models (such as PROC in TRA). No systematic trend, however, was observed between less and more experienced participants in the study or between those employed at different types of organizations.

The present work shows the results obtained for the agreement evaluated for several professionals in occupational safety and health when using TREXMO (Savic *et al.*, 2016). By using between-model translations, TREXMO supports parameter selection and it is expected to provide reliable estimates for a wide range of ESs. The agreement results obtained from TREXMO were compared with those found when coding the models' parameters manually. The term "*manual*" coding is used to refer to the common use of the models outside the framework of TREXMO.

Method

Eighteen assessors employed in occupational and safety organizations in Switzerland and the EU participated in this study and completed the exposure assessment. Only those, who were to some extent experienced with the exposure models, were invited. On average, the assessors had 5 years of experience in occupational hygiene. Before this study, nine assessors had applied ART or Stoffenmanager in >10 instances; four were more experienced with TRAv3; three had used one of these models 5–10 times, while the remaining two had conducted only 1–3 exposure assessments by using any of the considered models.

Each assessor was provided with six ESs, which were selected from a NIOSH database (Lee *et al.*, 2019a,b). Among these six ESs, three ESs addressed exposure to vapors, including spreading of liquids over a surface (ES-1), liquid transfer (ES-2), and spraying of liquids (ES-3). The remaining three ESs addressed exposure to dusty products (powders), including powder handling (ES-4), powder transfer (ES-5), and moving/agitation of powders (ES-6). More details on the ESs can be found in Supplementary Tables S1 and S2, available at *Annals of Work Exposures and Health* online. The assessors were asked to code the most appropriate parameters by interpreting the exposure conditions in the ESs and evaluate exposure by using the three following methods:

1. Manual coding, where the models' parameters were coded outside the framework of TREXMO and thus with no parameter translations applied. A Microsoft Office Excel (Visual Basic Application) file was provided, containing all input parameters of the six models including Stoffenmanager, ART, MEASE, EMKG-EXPO-TOOL, EASE, and ECETOC TRAv3.
2. Translating from Stoffenmanager (SM), where SM was used as the starting model. This means that the assessors interpreted the information in the provided ESs and selected the corresponding parameters in this model. The established coding was used to get the translation outcomes in the other models. Due to a low translation efficiency to ART (Savic *et al.*, 2016), SM was not used to deliver translation outcomes to ART.
3. Translating from ART, where ART was used as the starting model in TREXMO to obtain the parameter sets in the other models.

The study was designed in a way that an assessor did not evaluate an ES more than once. This was done to redistribute the total work amount evenly over the 18 assessors and to make the study more time-efficient. Therefore, based on the backgrounds and experience, the assessors were split into three equal groups to minimize a possible between-group variability. Each group assessed six ESs following a different order of the three methods above. For example, group I evaluated ES-1 (spraying of liquids) and ES-4 (powder handling) by using manual coding, while ES-2 (liquid transfer) and ES-5 (powder transfer) by translating the parameters from ART in TREXMO and ES-3 and ES-6 by translating from SM in TREXMO. Group II, by following a different order, evaluated ES-1 and ES-4 by translating from ART in TREXMO, while group III evaluated these two ESs by translating from SM in TREXMO. Finally, each vapor and dust ES was evaluated by all three methods, but different assessors. This means that every ES-method combination was evaluated six times by six different assessors. The only exception is ES-3 (spraying of liquids), which was not assessed in EMKG-EXPO-TOOL and EASE, because this activity is outside of the domain of two models.

Statistics

For each assessment method, the agreement between the exposure parameters, which were coded by different assessors, and between the obtained exposure estimates was analyzed in R software (version 3.3.1). The agreement was analyzed for all six models of TREXMO. We, however, present only the results of SM and TRAv3, for which more evaluation studies are available (Schinkel *et al.*, 2010; Landberg *et al.*, 2017; Spinazzè *et al.*, 2017; van Tongeren *et al.*, 2017; Lee *et al.*, 2019a) compared with the other models (supplementary table S1 of Lee *et al.*, 2019a,b), in this short communication with the limited length. The results of MEASE, EMKG-EXPO-TOOL, EASE, and ART (only for manual coding) are presented in Supplementary Tables S3–S8, available at *Annals of Work Exposures and Health* online.

Different measures were used to quantify the levels of agreement per parameter and between the modeled estimates. For the former, the percentage agreement [Gisev *et al.*, 2013; equation (1)] and S-index [Falotico and Quatto, 2010; equation (2)] were calculated. The percentage agreement is calculated as the ratio between the number of times any two responses (i.e., exposure parameters) were in concordance and the total number of responses. The reasoning for using S-index as an alternative to Cohen's kappa (Cohen, 1960) as the agreement measure has been explained elsewhere (Falotico and Quatto, 2010).

$$P\,(\%) = \frac{N\,(\text{concordant responses})}{N\,(\text{total})} \times 100 \qquad (1)$$

$$S = \frac{\frac{P}{100} - \frac{1}{N}}{1 - \frac{1}{N}} \qquad (2)$$

In equation (2), N is the number of categories (e.g., five dustiness categories in SM) available per given exposure parameter.

For the agreement between the modeled estimates, the relative standard deviation (RSD) (equations (3) and (4)) was calculated.

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}} \qquad (3)$$

$$\text{RSD} = \frac{\sigma}{\bar{X}} \times 100\% \qquad (4)$$

where, σ is the standard deviation of the modeled estimates and $\bar{X}$ is their average value. The higher RSD, the smaller agreement is between the assessors, and vice-versa. For each coding method, a mean (with its standard deviation) and a median were calculated for the six RSD obtained for the six ESs.

Results and discussion

Table 1 shows the agreement per input parameter for SM and TRAv3. The parameters are classified into source, activity, control, and time group, describing different aspects of exposure. The same classification has been already applied in Savic *et al.* (2016). Table 2 contains the calculated RSD for the estimates of the two models. The results for the other four investigated models (i.e., ART, MEASE, EMKG-EXPO-TOOL, and EASE) are given in Supplementary Tables S3–S8, available at *Annals of Work Exposures and Health* online. Although somewhat lower, the RSD values for MEASE were similar to those of TRAv3. The translations from ART were the least efficient for EMKG-EXPO-TOOL and EASE (Supplementary Table S6, available at *Annals of Work Exposures and Health* online), for which the RSD values were often even lower compared with the manual coding. Also, when coding manually, the lowest RSD was obtained for EMKG-EXPO-TOOL (median = 71; Supplementary Table S6, available at *Annals of Work Exposures and Health* online). MEASE and ART showed slightly higher RSD (both with median = 75; Supplementary Tables S6 and S8, available at *Annals of Work Exposures and Health* online) than

Table 1. Agreement per parameter for SM and TRAv3, showing assessed percentage agreement (P) and S-index (S) for the three coding methods: manual, when translating from ART (TREXMO-ART) and SM in TREXMO (TREXMO-SM).

Parameter group	Stoffenmanager (SM)					ECETOC TRAv3						
	Parameter	Manual		TREXMO-ART		Parameter	Manual		TREXMO-SM		TREXMO-TRA	
		P, %	*S*	*P*, %	*S*		*P*, %	*S*	*P*, %	*S*	*P*, %	*S*
Source	Dustiness	47	0.36	47	0.36	Dustiness	58	0.37	69	0.54	71	0.58
	Vapor pressure	80	[a]	71	[a]	Vapor pressure	76	[a]	80	[a]	100	[a]
	Concentration	55		71		Concentration	89	0.85	84	0.79	91	0.88
	Cleaning of workplace	64	0.28	52	0.04							
	Inspections and maintenance	44	0.01	72	0.44							
Activity	Type of handling (powder)	16	0.03	40	0.31	PROC (powder)	13	0.10	13	0.10	31	0.28
	Type of handling (vapor)	53	0.47	58	0.52	PROC (vapor)	89	0.88	62	0.60	64	0.63
	Breathing zone	77	0.54	66	0.32	Sector of use	91	0.82	91	0.82	91	0.82
	Number of employees	78	0.56	78	0.56							
	Evaporation, drying or curing	58	0.15	81	0.62							
Controls	General ventilation	94	0.93	74	0.66	Ventilation	42	0.33	66	0.60	71	0.66
	Room volume	78	0.70	67	0.56							
	Localized controls	76	0.69	82	0.78							
	Immission	94	0.92	94	0.92							
Time	Task duration	100	[a]	100	[a]	Task duration	94	0.92	89	0.85	94	0.92

[a] *S*-index was not calculated for the continuous parameters.

Table 2. Relative standard deviation (%) for the modeled exposure estimates by using the three coding methods.

Exposure type	ES	Stoffenmanager (SM)		ECETOC TRAv3		
		Manual	TREXMO-ART	Manual	TREXMO-SM	TREXMO-ART
Vapors	1	70	94 (–)	94	42 (+)	44 (+)
	2	87	29 (+)	141	200 (–)	0 (+)
	3	121	46 (+)	61	63 (–)	43 (+)
Powders	4	115	78 (+)	90	45 (+)	49 (+)
	5	156	34 (+)	92	75 (+)	16 (+)
	6	74	208 (–)	106	220 (–)	133 (–)
Number of "+"			4/6		3/6	5/6
Mean RSD		104	82	97	108	48
SD[a] RSD		33	67	26	81	46
Median		101	62	93	69	43.5

+ (or –) means that the agreement of exposure estimates using TREXMO-SM or TREXMO-ART was improved (or decreased) compared with the using the manual coding.

[a]SD, standard deviation for RSD.

EMKG-EXPO-TOOL. This might indicate the best level of agreement for the three given models. Since, however, only 18 assessors participated in the study, this statement cannot be considered reliable enough.

Since no translations were performed from other models to ART, only the results for manually coding were calculated (Supplementary Tables S7 and S8, available at *Annals of Work Exposures and Health* online). Compared with the manual coding in the other models, the agreement per parameter for ART was considerably better. Probably, this might be because ART provides supporting information using pictures and/or detailed description with examples for exposure parameters. Excluding dustiness, for which the agreement was the lowest (S = 0.33), the S-index for other parameters was within the range of 0.51–1.00. The better starting agreement for ART may explain why it resulted in better results, as compared with SM as the starting models. In addition, the corresponding RSD found for ART were similar to those of SM and TRAv3. Compared with SM and TRAv3, Riedmann *et al.* (2015) have shown that the estimates in ART are more sensitive to the changes in input parameters. This could explain why the better agreement per parameter did not result also in a better agreement between the estimates of ART.

Stoffenmanager

Compared with the manual coding, translations from ART improved the agreement for *type of handling* (S-index from 0.03 to 0.31 for powders; only slight for vapors), *evaporation, drying, or curing* (i.e., presence of the far-field exposure source; S-index from 0.15 to 0.62), and somewhat less for *localized controls* (S-index from 0.69 to 0.78) (Table 1). These parameters have previously found to be most decisive factors in estimating exposures for SM (Riedmann *et al.*, 2015; Savic *et al.*, 2018). This is also obvious from the results shown in Table 2, where the RSD was smaller for two out of three ESs for both vapors and dusts when translating from ART. Excluding *inspections and maintenance*, which has almost no impact on estimating exposures, the other parameters in the source group were not altered by the translations from ART. This was an expected outcome, since SM and ART share similar descriptions of the parameters used to model the source group. For example, the five dustiness categories (such as *fine dust*) are the same in both models.

The translations from ART, however, reduced the level of agreement for some parameters, such as *breathing zone* (i.e., the parameter that determines the presence of a near-field source), *general ventilation*, and *room size*. A detailed analysis of the assessors' selections showed that this outcome originates from the coding provided in the starting model, that is, ART. The two coding methods (i.e., manual and translating from ART) were not performed by the same group of assessors. In this case, it was found that the assessors who were translating from ART were less concordant when selecting the parameters in this starting model. This later resulted in different translation outcomes in SM. In addition, in a few cases related to *general ventilation*, the assessors did not apply the translation outputs obtained from ART, but rather selected other ventilation categories available in SM. If the assessors applied the obtained translations, the inter-assessor agreement for general ventilation would not be different for the different coding methods.

ECETOC TRAv3

Translations into TRAv3 were obtained from both ART and SM as the starting models. When translating form ART slightly better results were found than when translating from SM. For example, translations from SM did not alter the agreement for PROC (S-index = 0.10, for powders), while with ART, the S-index was increased meaningfully (=0.28). Compared with the manual coding, the smaller RSD was obtained for only three ESs when translating from SM, while the agreement was improved for five out of six ESs with the translations from ART.

In a few cases, again, the translations led to lower agreement. The most important was for PROCs coded for vapors, where S-index of 0.88 found when manually coding was reduced to 0.60 and 0.63 when translating from SM and ART, respectively. The reason for this outcome was the definition of the translation rules (Savic *et al.*, 2016) in case of the spraying ES. Since neither ART nor SM have *sector of use* defined as *industrial* and *professional* use, the translation of the spraying activity/handling from these models results in two translation outcomes in TRAv3, that is, *PROC 7-industrial spraying* and *PROC 11-professional spraying*. One might believe that the assessors' choice between two PROCs in TRAv3, which followed the translations, was rather random. However, all the assessors, who coded the given ES manually in TRAv3, selected PROC 7, as the corresponding ES report stated clearly that the ES addresses industrial use.

Overall, the translations from ART led to better results than those from Stoffenmanager. The explanation for this might be found in the concept behind the translation rules (Savic *et al.*, 2016). ART is a more complex model than any other tools and often has more options in an exposure parameter (e.g., five categories of dustiness). Since lower-level models (such as TRAv3) have less number of options in exposure parameters, it is often a case that two or more options in exposure parameters of ART share the same translation path by leading to a single outcome in another model. The simplest example would be dustiness, where for example, both "fine dust" and "extremely fine dust" are translated as "high dustiness" in TRAv3. This means that if two assessors select different options for dustiness parameter in ART, the translations might still result in the same outcome in TRAv3. Regarding the inter-assessor agreement preceded by a translation from ART, the translation of multi options in a parameter into a single option in the lower-level models (n→1 translation) might compensate the complexity of ART.

Conclusion

In this study, we evaluated the inter-assessor agreement for six different ESs when assessed within and outside the framework of TREXMO. The study was limited to a small number of ESs (i.e., six) and only 18 occupational professionals. The findings of this study might have also been biased by various factors, outside of the tools or the methods used, leading to between-assessor variability in terms of, for example, their backgrounds, their subjective judgment (such as whether they prefer middle values or extremes), and their motivation for performing the assessments. This impact, however, stays unknown due to the small number of the assessors. It would be thus difficult to generalize the obtained results and draw straightforward conclusions. The results, however, imply a potential of TREXMO to affect the assessors' agreement and suggest further investigations to be conducted. Another limitation of the study was the use of TREXMO 1.0, which was not offering a user-friendly graphical environment. The study should be thus repeated with the most recent version of the tool and with a greater number of volunteers and ESs. Finally, it should be noted that TREXMO integrates version 4.0 of Stoffenmanager and that some differences, regarding the inter-assessor agreement, might exist with newer versions (Heussen and Hollander, 2017; Savic *et al.*, 2017).

Supplementary Data

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

Disclaimer

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

This research was supported by the Swiss Centre for Applied Human Toxicology (SCAHT) and the State Secretariat for Economic Affairs (SECO). The authors declare that they have no competing interests that might be perceived to influence this article.

References

Cherrie JW, Schneider T. (1999) Validation of a new method for structured subjective assessment of past concentrations. *Ann Occup Hyg*; **43**: 235–45.

Cohen J. (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas*; **20**: 37–46.

ECETOC. (2009) *Addendum to ECETOC targeted risk assessment*. Report No. 93. Brussels, Belgium: European Centre for Ecotoxicology and Toxicology of Chemicals.

ECETOC. (2012) *ECETOC TRA version 3: Background and rationale for the improvements*. Technical report No. 114. Brussels, Belgium: European Centre for Ecotoxicology and Toxicology of Chemicals.

ECHA. (2016) *Guidance on information requirements and chemical safety assessment—chapter R.14: occupational exposure assessment*. Helsinki, Finland: European Agency for Chemicals.

Falotico R, Quatto P. 2010. On avoiding paradoxes in assessing inter-rater agreement. *Stat Appl—Ital J Appl Stat*; **22**: 151–60.

Fleiss JL. (1971) Measuring nominal scale agreement among many raters. *Psychol Bull*; **76**: 378–82.

Friesen MC, Coble JB, Katki HA *et al*. (2011) Validity and reliability of exposure assessors' ratings of exposure intensity by type of occupational questionnaire and type of rater. *Ann Occup Hyg*; **55**: 601–11.

Gisev N, Bell JS, Chen TF. (2013) Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Res Social Adm Pharm*; **9**: 330–8.

Heussen GAH, Hollander AL. (2017) Stoffenmanager® exposure model algorithms within TREXMO and Stoffenmanager® lead to different outcomes. *Ann Work Expo Health*; **61**: 604–6.

Kromhout H, Oostendorp Y, Heederik D *et al*. (1987) Agreement between qualitative exposure estimates and quantitative exposure measurements. *Am J Ind Med*; **12**: 551–62.

Lamb J, Galea KS, Miller BG *et al*. (2017) Between-user reliability of tier 1 exposure assessment tools used under REACH. *Ann Work Expo Health*; **61**: 939–53.

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.

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

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

Riedmann RA, Gasic B, Vernez D. (2015) Sensitivity analysis, dominant factors, and robustness of the ECETOC TRA v3, Stoffenmanager 4.5, and ART 1.5 occupational exposure models. *Risk Anal*; **35**: 211–25.

Savic N, Gasic B, Vernez D. (2017) Stoffenmanager® algorithm within version 6 differs from the published algorithm within old versions and TREXMO. *Ann Work Expo Health*; **61**: 607–10.

Savic N, Gasic B, Vernez D. (2018) ART, Stoffenmanager, and TRA: a systematic comparison of exposure estimates using the TREXMO translation system. *Ann Work Expo Health*; **62**: 72–87.

Savic N, Racordon D, Buchs D *et al*. (2016) TREXMO: a translation tool to support the use of regulatory occupational exposure models. *Ann Occup Hyg*; **60**: 991–1008.

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.

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.

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.