Title	Modelling of Inhalation Exposures to Pharmaceutical Agents
Author(s)	Mc Donnell, Patricia
Publication Date	2011-09-09
Item record	http://hdl.handle.net/10379/2226

Reliability of the Advanced REACH Tool when used by health and safety professionals with no previous experience

Patricia E. Mc Donnell[1], Jody M. Schinkel[2], Marie A. Coggins[1], Hans Kromhout[3], Wouter Fransman[2], Erik L. Tielemans[4]

[1] School of Physics, National University of Ireland, Galway, University Road, Galway, Ireland
[2] Research Group Quality & Safety, TNO, Zeist, The Netherlands
[3] Institute for Risk Assessment Sciences, Division of Environmental Epidemiology, Utrecht University, Utrecht, The Netherlands
[4] Healthy Living, TNO, Zeist, The Netherlands

Author to whom correspondence should be addressed:
Email address: patricia.mcdonnell@nuigalway.ie

To be submitted to: Journal of Environmental Monitoring

Abstract
When developing an exposure model, reliability is an important aspect. This study focused on the reliability of the Advanced Reach Tool (ART) when used by eighteen health and safety professionals from the pharmaceutical industry, who had no previous experience with the tool. It aims to investigate: the reliability of rater's assessment, at the level of model determinants and of the resulting exposure estimate by comparing with a gold standard; inter-rater agreement; and the effect of providing information on agreement. Participants individually assessed eight activities comprising information on model parameters. Information and demonstration on ART were provided at two stages. The average not-chance corrected agreement values of the determinants with gold standard for the increased information stages were 58, 67, and 83% respectively. For five of six determinants there was an increased inter-rater agreement with provision of information. There was a broad range between the raters and gold-standard exposure estimates, with approximately 60% within ten-fold of the gold-standard. ART is an expert tool and use without training is not recommended. A more extensive study is planned. As models are increasingly used in the context of REACH and beyond, this study emphasises that validation and reliability studies are required.

INTRODUCTION

Both quantitative measurements of personal exposure to contaminants through air monitoring and professional judgement can be employed to assess occupational exposures. However occupational exposure monitoring can be expensive and labour intensive. While the use of professional judgement in the decision-making process has a role in exposure assessment, it is not defined and it is difficult to replicate or scientifically justify the exposure estimates or conclusions[1]. The professionals conducting the assessments may not be familiar with the jobs or industry and their background may influence how they assess exposure [2]. Exposure models may be a more transparent option to assist in the exposure assessment process and also for derivation of exposure scenarios as legislated under the Registration Evaluations Authorisation and restriction of Chemicals (REACH) Regulations 2006 [3]. The Advanced Reach Tool (ART) was developed for the purpose of the compilation of REACH exposure scenarios, to estimate exposure levels for specific groups of workers sharing operational conditions and risk management measures across different workplaces in Europe.

When developing exposure models both validity and reliability of the model are important issues to address [4].Validation compares the model estimates of exposure to real measurement data (accuracy and bias) while the reliability is a measure of the consistency of assessments or of the ability of assessors to reach the same conclusions about a specific case [5]. User variation in model estimates may occur if a user has a limited understanding of the exposure scenario (operational conditions and risk management measures), the exposure model and/or if the model is misused [6]. The impact of user-variation could have serious consequences for workers health, if an exposure scenario is incorrectly diagnosed as 'safe', or for the financial situation of the organisation if an exposure scenario is incorrectly diagnosed as 'unsafe', which leads to often very costly over-engineering. This might become a more widespread problem as a variety of commercial user-friendly software packages become available [6] . While studies have investigated the reliability or accuracy of subjective judgements of exposure [7-10], very few studies have been published on the reliability of occupational exposure models; an example is a study on the dermal exposure assessment model, DREAM [4]. This combination of '*serious consequences*' and '*limited reliability studies'* is a serious problem [6] and one that requires further investigation.

This study investigates the reliability of ART when workplace conditions were described with text and photos, i.e. users had to interpret information for some of the ART determinants. As the ART was developed for the purposes of assessing REACH exposure scenarios, ideally the reliability of the ART should be tested with REACH dossier documentation where relevant information is available and user interpretation of ART determinants should be minimal.

Health and safety professionals with experience in the pharmaceutical industry and with largely no previous experience using the ART or other exposure models attended a one day workshop. The professionals used the ART to assess four exposure scenarios, each comprising two activities, from the pharmaceutical industry individually at separate stages during the day. Introductory information and a demonstration on using the ART were delivered at two different stages and we considered if the provision of this information improved the agreement amongst the professionals. The aims of this study were to investigate: (i) do professionals' agree with the gold-standard assessment per determinant (ii) inter-rater agreement per exposure determinant and (iii) comparison of the professionals and the gold standard ART exposure estimate. This study was considered to be a pilot study that would provide an indication of the reliability of the ART amongst health and safety professionals with no previous training or experience using the tool.

METHODS

The ART framework

The ART framework incorporates a mechanistic model to predict inhalation exposure and exposure measurements combined using Bayesian statistics in order to produce more precise estimates of exposure. The ART mechanistic model is based on a source-receptor model [11–12] which describes a stepwise transport of a contaminant from the source to the receptor. ART incorporates seven independent principal modifying factors (MF) e.g., substance emission potential, activity emission potential, localised controls, dispersion, personal enclosure, segregation, and surface contamination. Relative multipliers have been assigned to the underlying categories per determinant of each MF which are used as multipliers in the mechanistic model algorithm [13]. This results in dimensionless relative exposure scores, which were calibrated using exposure measurements, enabling the mechanistic model to predict the geometric mean (GM) exposure level of an exposure scenario in mg/m^3 [14]. The mechanistic model was calibrated separately for the following exposure forms: vapour, mist, abrasive dust and dust. Fumes, gases and fibres are outside the current applicability domain of the ART. The ART version 1.0 is freely available online (www.advancedreachtool.com). This study focuses on the reliability of the inhalable dust mechanistic model of the ART and does not address the Bayesian application of the tool.

Study design

Health and safety professionals and specifically occupational hygienists are normally responsible for carrying out exposure assessment in the workplace and accordingly it was anticipated that they may be qualified users of the ART as an exposure assessment tool for REACH. A list of 61 health and safety professionals with experience in the pharmaceutical industry was compiled with the help of the Occupational Hygiene Society of Ireland (OHSI) and from personal contacts. Potential candidates were contacted via e-mail and 18 people agreed to participate in the study. In advance of the workshop a questionnaire was circulated to all participants to collect details such as: current job role; academic qualifications; experience to date working as an occupational hygienist; experience conducting chemical exposure assessments; experience using exposure models; and their familiarity with the ART. The questionnaires were disseminated to participants prior to the workshop. Of the 18 participants, nine participants had greater than ten years experience in occupational hygiene. One third of the group had some previous experience with using other exposure models. Only one of the 18 participants had previously used the ART, and this was for less than one hour; the remainder of the group had no previous experience with the ART.

A one day work shop was conducted at the National University of Ireland, Galway in July 2010 during which the professionals used the ART to assess four exposure scenarios. The exposure scenarios were representative of scenarios from the pharmaceutical industry and consisted of two activities per scenario during which a hazardous substance was handled. The scenarios were developed so as to contain similar levels of information on all the MFs as presented in Table 1. Prior to the assessment stages the participants were provided with a text copy of the exposure scenario consisting of (i) a brief text description outlining the substance, the handling activities, any localised controls, and work area size and ventilation rates, and (ii) pictures of the substance, processes and local control measures. For two of the scenarios the participants were provided with a container containing powder material, for which they subjectively assessed the dustiness of the materials.

Table 1 presents the MFs and classes that were included in the exposure scenarios. During the one day workshop it was only possible to assess eight activities i.e. four exposure scenarios, and those determinants considered most relevant for the pharmaceutical industry were included. In order for the workshop to be comparable to workplace use of the ART, subjective judgement was required for some of the determinants e.g. dustiness and emission source, while the information was more clearly provided for other determinants e.g. room size and air change per hour rates (ACH). Also while some MFs or determinants were not specifically referred to in the documentation e.g. segregation and separation, it was still necessary for participants to assess them all in the correct place in the tool to obtain an exposure estimate. Using the online version 1.0 of ART, the participants were asked to independently assess the four scenarios without discussions with the trainers or with each other. The participants were unaware of the gold-standard choice of determinants and exposure estimates i.e. the expert-assessment or how the developers of ART assessed the determinants and the resultant gold-standard exposure estimates. ART scores for each determinant were assigned by one member of the project team (PMD) and subsequently reviewed by a project team member (JS) and discussed until consensus was reached.

The workshop was structured so as to consider the effect of dissemination of information on the use of the ART on the reliability of the participant's while assessing the scenarios. Each participant individually assessed four exposure scenario at three stages throughout the day and each scenario was assessed only once: (i) one scenario was assessed *'without introduction'* to the theory of the ART or the functionalities of the tool; (ii) another scenario was assessed *'after introduction'* to the ART theory; (iii) and finally two scenarios were assessed *'after demonstration'* of the ART. To enable some comparison of participants' agreement per determinant without and after the introduction to ART, during the *'without introduction'* stage half the participants assessed Scenario 1 and the other half assessed Scenario 2; while during the *'after introduction'* stage each group assessed the opposite scenario. The participants were split into two groups so as both groups included individuals with similar years of experience in occupational hygiene.

During the introduction to ART (45 min), information was presented on the ART mechanistic model and on the model MFs. During the demonstration of using the ART (90 min) the assessment of the 'gold standard' of the previous Scenario 1 was presented. The demonstration also involved more detailed information on the use of the ART and scoring of the MFs, and discussions to clear up any uncertainties on the assessment of Scenario 1 or 2. In total only 2 hours and 15 minutes of information was presented and due to time limitations of this one day workshop, it was not possible to disseminate further detailed training on the ART. After the demonstration stage all the participants assessed both Scenario 3 and Scenario 4. Upon completion of the assessments hardcopies of all of the participant's assessments were printed and were subsequently transcribed into Microsoft Excel.

Table 1: Outline of information on determinants that were provided in the exposure scenarios

MF	Classes and determinants in exposure scenarios	Information provided	Rationale for providing information
Substance Emission Potential	Dustiness	Physical sample or pictures of the materials	Dustiness test results are not normally available, so have to use subjective assessment of dustiness
	% of API in the material	Specify the % of API in the materials	This information is readily available in the workplace (e.g. material safety data sheets)
	Moisture	Description of material and activity e.g. wet wipe or mopping of floor	In reality this is quite subjective but the ART description text should guide participants
Activity Emission Potential	Activity class • Compression • Movement and agitation • Transfer – falling • Handling • Fracturing	Brief description of activity (e.g. scooping, disposal, tabletting) and a picture of activity taking place	Participants had to decide on activity class and subclass and further determinants
		Will specify names of activity and corresponding duration of activities	This information is readily available in the workplace
	Quantity / use rate	Specify total quantities involved in the activity	Participants had to work out quantity/min
	Emission source	Description of activity e.g. at control panel located 5m away from tabletting machine or manually scooping tablets.	This information is readily available in the workplace; participants had to decide whether it is e.g. near-field (NF) or far-field source (FF) or a situation with NF and FF.
Local Controls	Sub-classes • No Local Controls • Containment • Glove-box • LEV	Briefly describe the local control measure and provide picture	This information is readily available in the workplace; participants had to decide on ART classification
Surface Contamination	• Questions regarding enclosure of process, housekeeping and maintenance.	Brief description referring to visible/no visible contamination on work surfaces and corresponding pictures for rest of scenario. Comment on housekeeping/maintenance	This information is readily available in the workplace e.g. if housekeeping, maintenance etc. are occurring
Dispersion	Indoors	Specify room size and air changes per hour (ACH)	This information is readily available on sites
	Down-flow room	Specify room size and ACH	This information is readily available on sites

Statistical Analyses
Statistical analyses were performed with SAS Statistical Software (version 9.1.3; SAS Institute, Cary, NC, USA). Each MF was assessed by all of the participants (hereafter referred to as raters) (n=18) for both activities in each of the four scenarios (activities n=8). The activities related to Exposure Scenarios 1 and 2, which were assessed alternatively by raters during without introduction and after introduction stages, were assessed separately (e.g. Activity 1a and 2a are without introduction; Activity 1b and 2b are after introduction).

To investigate the level of agreement of raters with the gold-standard, the percentage of ratings for the determinants of each activity that were in agreement with the gold-standard were calculated. To investigate if there were any differences in the agreement levels of raters of varying years of experience in occupational hygiene, the raters were separated into two groups, which were arbitrarily categorised as: <10 years of experience and ≥10 years experience in occupational hygiene.

While the percentage agreement results indicate the agreement with the gold-standard we were also interested in the inter-rater agreement. Cohen's kappa statistic (*k*) gives the exact proportion of agreement that cannot be expected by chance alone [15]. Kappa statistics were calculated as described by [16] for multiple raters and does not assume that the raters responsible for rating one subject are the same as those rating another subject or scenario.
For the surface contamination determinants, as 6 out of 8 activities occurred in an apparently clean work area, almost no variation in rater judgement was expected. As the MFs segregation and separation were not present in any activity, no variation in rater input was expected. For these determinants without varying rater inputs, kappa statistics were not relevant and the percentage agreement with the gold-standard results covers the conclusions about agreement. Also as the scores for the activity emission potential (AEP) MF are a result of multipliers for several determinants including for example; activity class, quantity of material, drop height and type of handling, it was not possible to calculate kappa statistics for this MF. Kappa statistics were calculated for the following determinants: dustiness, emission source, activity class, primary local controls, room size and ventilation rate.
The strength of the inter-rater agreement was qualified using terms defined by Landis and Koch: kappa statistic results ≤0 = no agreement (other than would be expected by chance), 0.01–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial and 0.81–1.00 = almost perfect [17]. Results are presented as the Cohen's kappa statistic and standard errors for each determinant throughout the various stages. Where a participant did not assess a second activity in a scenario as required, all determinants were assigned to a 'blank' category that was treated as not in agreement with the 'gold standard'. Where a determinant was not applicable it was also treated as an additional category (e.g. if Dispersion='Downward laminar flow booth' was chosen then the determinants 'room size' and 'ventilation rate' were not required) and did not influence the agreement measures.

Relative weights have been assigned to the underlying categories of each of these MFs which are used as multipliers in the mechanistic model algorithm [14] to result in exposure estimates in mg/m^3 [15]. As the exposure estimate will be used in the exposure assessment process we compared how the raters eventual exposure estimates compared to the gold-standard exposure estimate.

RESULTS

Table 2 presents the percentage of ratings in agreement with the ‘gold-standard’ for each of the determinants at the various stages of the workshop. The average agreement values of all the determinants with the gold standard for the without introduction, after introduction, and after demonstration stages were 58, 67, and 83% respectively, and the values per individual determinant are presented in Table 2.

As the AEP MF comprises several determinants we investigated if the raters chose the gold-standard activity class. Across all the information stages, the percentage of ratings in agreement with the gold-standard activity class was approx 80%. To obtain the AEP score, raters also had to assess two or more determinants (e.g. drop height and handling type). For the without introduction and after introduction stages relatively low percentage of the ratings were in agreement with the gold standard (14 and 17%). After the demonstration stage the percentage agreement with the AEP score increased to 45%. Surprisingly, some determinants (e.g. secondary local controls, separation and segregation) which were not referred to in the scenario documentation were assessed by a small number of raters in some of the scenarios, which resulted in a lower agreement for these MFs. While raters were provided with the information in the exposure scenario for ACH and room volume, there was also relatively low agreement during the without introduction stage. After introduction and demonstration stages improvements were evident for dustiness, emission source, AEP, localised controls, dispersion, surface contamination, segregation and separation. Also, three of the activities (denoted by * in Table 2), were not assessed by some of the raters (n=5). Although two activities were described per scenario, raters were possibly not aware of the second activity in the scenario or may not have known how to assess it in the tool.

Overall no apparent differences in percentage agreement with the gold-standard were seen between the group of experienced raters (>10 years experience in occupational hygiene) and less-experienced raters (< 10 years experience). Also there were no apparent differences between the participants whom had previous experience with other exposure models and those with no experience (results not presented).

Table 2: Percentage of ratings in agreement with the ‘gold-standard’ for each determinant per activity

Activity	N	Dustiness (%)	Emission source (NF-FF) (%)	AEP		Controls		Dispersion		Surface contamination			Segregation (%)	Separation (%)	Average (%) **
				Activity Class (%)	AEP Score (%)	1st LC (%)	2nd LC (%)	Room Size (%)	ACH (%)	1 (%)	2 (%)	3 (%)			
	Categories	5	2	7	n/a	21	21	8	9	2	2	2	5	5	
	Raters														
1.1	9	56	78	100	33	22	100	67	67	22	22	89	67	100	
1.2	6*	33	56	11	0	22	67	33	44	67	44	56	67	67	**58**
2.1	9	11	56	89	22	44	78	100	100	89	78	89	100	100	
2.2	9	11	44	89	0	33	67	78	78	89	89	89	100	100	
Average without introduction		**28**	**25**	**73**	**14**	**30**	**78**	**70**	**72**	**67**	**58**	**81**	**84**	**92**	
1.1	9	11	100	100	11	44	89	100	78	33	33	78	55	100	
1.2	9	33	100	100	11	67	78	100	78	89	79	89	100	100	**69**
2.1	9	33	89	89	33	11	89	78	78	100	79	78	100	100	
2.2	8*	33	78	56	11	0	89	67	61	89	89	89	89	89	
Average after introduction		**28**	**92**	**86**	**17**	**31**	**87**	**86**	**74**	**78**	**70**	**84**	**86**	**97**	
3.1	18	79	100	61	22	56	100	89	100	100	100	95	100	100	
3.2	17*	79	89	83	39	72	89	83	95	95	95	95	95	95	**85**
4.1	18	11	89	89	61	100	100	100	89	95	95	100	90	100	
4.2	18	42	100	61	56	89	78	100	83	100	100	100	90	100	
Average after demonstration		**53**	**95**	**74**	**45**	**63**	**92**	**93**	**92**	**98**	**98**	**98**	**94**	**99**	

1st LC = Primary localised controls
2nd LC = Secondary localised controls
* There were no entry/blank assessments for this activity
1= process fully enclosed
2=effective housekeeping
3=general housekeeping
**average of all columns with the exception of Activity Class (AEP determinant score included)

Table 3: Inter-rater reliability per determinant for the various stages (kappa statistics and standard error values in brackets)

Stage	Dustiness	Emission source (NF-FF)	Activity Class	Primary Local Controls	Dispersion	
					Room size	ACH
Average without introduction	0.34** (0.04)	0.15* (0.05)	0.60*** (0.04)	0.03* (0.03)	0.37** (0.05)	0.41*** (0.05)
Average after introduction	0.18* (0.05)	0.67**** (0.06)	0.69**** (0.04)	0.13* (0.04)	0.60*** (0.06)	0.31** (0.05)
Average after demonstration	0.41*** (0.03)	0.74**** (0.03)	0.47*** (0.02)	0.55*** (0.03)	0.78**** (0.03)	0.73**** (0.03)

Level of agreement
<0: none
0.01-0.20: slight*
0.21-0.4: fair**
0.41-0.60: moderate***
0.61-0.80: substantial****
0.81-1.0: almost perfect*****

Table 3 presents the kappa statistics (and standard errors) at the various stages for the determinants in the reliability study. Referring to the Landis and Koch agreement scale [17] during the without introduction stage, there was slight to moderate agreement per determinant (k range=0.03-0.60). After introduction stage, there was slight to substantial agreement per determinants (range 0.13-0.69). After demonstration stage, there was substantial agreement for: emission source, room size and ACH (k=0.74, 0.78, and 0.73 respectively); and moderate agreement for dustiness, activity class and primary local controls (k=0.41, 0.47 and 0.55 respectively).

Figure 1 presents the scatter of exposure estimates in relation to the gold-standard estimate.

Fig 1: The fold differences between rater's exposure estimates and gold-standard exposure estimates, expressed as a percentage of assessments

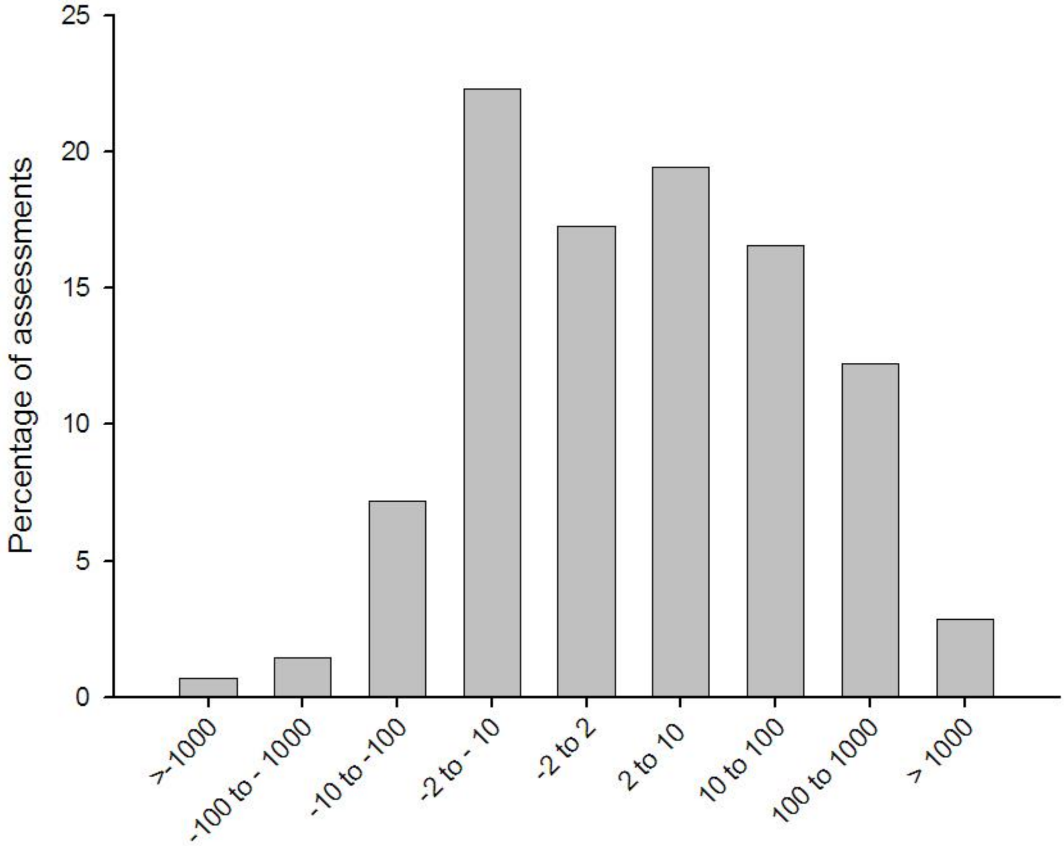

Figure 1 presents the difference between the raters' exposure estimates and the gold-standard exposure estimates, which were determined by two members of the project team (PMD and JS). The figure shows a very broad range with approximately 60% of raters' exposure estimates within ten-fold of the gold-standard exposure estimate. Notably, the raters tended to overestimate exposure since more than 30% of the assessments were greater than ten-fold higher than the gold-standard; while only approx 10% of the assessments were more than ten-fold lower than the gold-standard. In contrast with the percentage agreement and kappa statistics results, there was no apparent effect of the information stages on the agreement of the overall exposure estimates (results not presented).

DISCUSSION

This paper presents the results of a pilot study on the reliability of the ART when used by health and safety professionals from the pharmaceutical industry. With the exception of one participant, none of the professionals had previous training or working experience with the tool. It considered the effect of disseminating information on the ART on study participant's reliability to assess eight activities included in four exposure scenarios from the pharmaceutical industry. The results indicate that the ART cannot be used with sufficient reliability by health and safety professionals without the provision of further information on the tool.

The results of this study give an indication of the effect of provision of information on the reliability of untrained users of the ART. However it is possible that the level of information provided or complexity of the MFs in the exposure scenarios differed per stage of the workshop. Also, the assessments of the exposure scenarios before and after delivery of the introduction were made by different groups of randomly chosen health and safety professionals. Nonetheless we believe that the results indicate that provision of information will improve user reliability. For all model determinants (n=12), the average percentage of ratings in agreement with the gold-standard increased with the provision of information. When assessing room size and ACH determinants, the raters had the option to choose the exact room size and ventilation rates or choose from categories (e.g. $100m^3$ or small work room); while these choices are linked to the same model score, only the exact information specified in the scenario documentation was assessed as being in agreement with the gold-standard. Therefore for room size and ACH determinants, the percentage agreement was indeed slightly higher for some of the activities than the results presented. There were no apparent differences in percentage of ratings in agreement with the gold-standard with regard to years of experience in occupational hygiene practice. For five of the six model determinants, there was an increased inter-rater agreement with the provision of information, with substantial to moderate agreement for all determinants after the workshop facilitators delivered a demonstration of ART. There was a very broad range between the raters' exposure estimates and the gold-standard exposure estimates. Approximately 60% of rater's exposure estimates were within ten-fold of the gold-standard exposure estimate. Notably, raters tended to overestimate exposure and previous studies have shown that occupational hygienists tend to overestimate exposure when using exposure assessment tools [18]. As some MFs i.e. separation and segregation and were not varied in the exposure scenarios it is possible that the reliability results are underestimated as a result, however as shown in Table 2, some raters still assessed these as being applicable.

It is likely that the raters made errors in their assessments due to two broad issues: technical errors when using the ART website; and exposure assessment judgement errors. Firstly, it is likely that the participants did not receive enough adequately detailed information on how to use the website, and so even if they did understand the given scenarios, they did not know how to use the tool properly or how to input the required information in the correct places. Also while a lot of guidance text and photographs are provided on the ART website to assist users choose the correct categories, it is possible that due to time constraints, the participants did not locate or consider them adequately. Secondly, it is likely that the participants made mistakes in their exposure assessment judgements or with the theoretical use of the tool. Results of this study indicate that some of the model determinants were more problematic for participants to assess e.g. dustiness is acknowledged as being subjective. Also a limited amount of the information and photographs were provided in the exposure scenario documentation. Therefore it is possible that the raters encountered difficulties with assessing e.g. the level of contamination on objects, or the containment levels of the localised controls and so choose the conservative options during their assessments. Such choices could have a big influence on the exposure score and the reliability results. Consider the following example of a scenario where the gold-standard involves scooping coarse dust in a high specification glove-box. Referring to the available pictures and text in the scenario and the guidance within the tool, a rater assessed the scenario as involving fine dust in a low specification glove-box as opposed to coarse dust in a high specification glove-box, with all other determinants assessed correctly. The conservative rationale behind the choices of this rater is logical, but following the mechanistic model of ART [14], the choices results in 30-fold differences with the gold standard exposure estimate. Due to the multiplicative nature of the algorithm, wherein determinants result in multipliers possibly ranging from 0.0001 to 100, differences in assessment of each determinant can potentially have a large effect on the eventual ART exposure estimate. Users of the ART should be cognisant of the multiplicative algorithm, and that even if they assess all other determinants correctly, choosing one incorrect category can drastically affect the resulting ART exposure estimates. Also in some cases, even when the information was explicitly provided in the documentation e.g. room size and ACH, the participants inputted the wrong information. This observation highlights a more generic problem of user error, i.e. experts misinterpreting the assessment process, which may not be directly related to the ART.

The information provided in this workshop was quite limited and was only 135min total in duration. In the future proper, more detailed training sessions for users of the ART should be provided. They should focus on: technical aspects of using the website and availing of the substantial guidance available on the tool; the multiplicative algorithm of the ART mechanistic model; improving exposure assessment judgement including, assessment of determinants, particularly those highlighted as problematic to assess (e.g. dustiness and localised controls); demonstration use of the tool and of assessments; and feedback on assessments. Also the results of this study could be used to improve the user guidance for specific MFs, which may reduce the sources of variation between users of the ART.

While reliability is not a sufficient condition of, it is an important and a necessary component of the validity of occupational exposure models [19]. Nonetheless, there are few published reliability studies for occupational exposure models. This study investigated the reliability of the ART when workplace conditions were described in paper documents with limited text and photos and as a result it was necessary for raters to interpret information which would influence reliability. It is possible that the reliability of the ART may be better when used for the purposes of assessing REACH exposure scenarios, where users have access to relevant information in the REACH dossier documentation and so user interpretation of ART determinants should be minimal. The results of this study are not directly comparable to a study on the reliability of the dermal exposure assessment (DREAM) method, in which occupational hygienists used DREAM while performing side by side observations of different workplace tasks [4]; this observational approach would likely involve less of a need to interpret workplace conditions which was required in this study.

While the pharmaceutical industry is largely exempt from the requirements of REACH (as per Article 2 (5) (a) of the Regulations), it was anticipated that the ART could have useful applications for exposure assessment and risk management within this industry. A validation study of the ART using a dataset for the pharmaceutical industry reported that for 90% of the scenarios the exposure estimates were within the 90% uncertainty factor of 5.5 [11]. Results of this reliability study indicate that while approximately 60% of rater's exposure estimates were within ten-fold of the gold-standard exposure estimate, the remainder varied widely. This variability is unacceptable, particularly for assessment of exposures to harmful chemicals, such as those in use in the pharmaceutical industry. Results from this study signifies that it is not reasonable to assume that health and safety professionals, regardless of their years experience in occupational hygiene, will know how to use the ART without any prior training on the tool.

In conclusion, the ART is an expert tool and use without extensive training is not recommended. Even with the information that was provided in this workshop the reliability of untrained users of the tool is not sufficient. This study highlights model determinants and aspects associated with the tool that require particular attention during training. A more extensive study is required and planned to investigate the reliability of ART after provision of training. Moreover, as models are increasingly used in the context of REACH and beyond, this study emphasises that proper evaluation of reliability and accuracy of models needs much more attention in the exposure science community.

Acknowledgements

This study was supported by a research grant from GlaxoSmithKline and the College of Science, National University of Ireland Galway. We are very grateful to all of the GlaxoSmithKline employees who participated in the collation of the exposure data and to the participants of the workshop for their time and involvement in this study.

REFERENCES

1. J. Cherrie, T. Schneider, S. Spankie and M. Quinn, A new method for structured, subjective assessments of past concentration, *Occup. Hyg.*, 1996, **3**: 75-83.
2. K. Steinsvåg, M. Bråtveit, B. Moen and H. Kromhout, Inter-rater agreement in the assessment of exposure to carcinogens in the offshore petroleum industry, *Occup. and Environ. Med.*, 2007, **64**(9): 582-588
3. E. Tielemans, N. Warren, T. Schneider, M. Tischer, P. Ritchie, H. Goede, H. Kromhout, J. VanHemmen and J. Cherrie, Tools for regulatory assessment of occupational exposure: development and challenges, *J. of Expo. Sci. and Env. Epid.*, 2007, **17** S72-S80.
4. B. Van Wendel De Joode, J.Van Hemmen, T. Meijster, V. Major, L. London and H. Kromhout, Reliability of a semi-quantitative method for dermal exposure assessment (DREAM), *J. of Expo. Analysis and Env. Epi.*, 2005, **15**: 111-120.
5. D. Kunac, D. Reith, J. Kennedy, N. Austin and S. Williams, Inter- and intra-rater reliability for classifiction of medication related events in padeiatric inpatients, Q*ual. Saf. Health Care,* 2006 **15**: 196-201.
6. F. Swartjies, Human Exposure Model Comparison Study: State of Play, *Land Contamination and Reclamation* , 2001, **9**(1): 101-106.
7. H. Kromhout, Y. Oostendorp, D. Heederik and J. Boleij, Agreement between qualitative exposure estimates and quantitative exposure measurements. *Am. J. Ind. Med., 1987,* **12,** 551-562.
8. W. Post, H. Kromhout, D. Heederik, D. Noy and R. Smit Duijzentkunst, Semiquantitative Estimates of Exposure to Methylene Chloride and Styrene: The Influence of Quantitative Exposure Data. *App. Occup. and Environ. Hyg., 1991,* **6,** 197-204.
9. J. de Cock, D. Heederik and J. Burema, Experts' subjective assessment of pesticide exposure in fruit growing, *Scand. J. Work. Environ. Health, 1996,* **26**, 44-51
10. G. Ramachandran, Toward Better Exposure Assessment Strategies--The New NIOSH Initiative, *Ann. Occup. Hy.g,* 2008, **52**(5): 297-301.
11. P. Mc Donnell, J. Schinkel, M. Coggins, W. Fransman, H. Kromhout, J. Cherrie and E. Tielemans, Validation of the inhalable dust algorithm of the Advanced REACH Tool using a dataset from the pharmaceutical industry, 2011, *J. of Env. Monitoring* **16,** 1597-1606.
12. J. Cherrie and T. Schneider, T., Validation of a new method for structured subjective assessment of past concentrations. *Ann. Occup. Hyg.*, 1999, **43**(4): 235-245.
13. E. Tielemans, T. Schneider, H. Goede, M. Tischer, N. Warren, H. Kromhout, M.Van Tongeren, J. Van Hemmen and J. Cherrie, Conceptual Model for Assessment of Inhalation Exposure: Defining Modifying Factors, *Ann. Occup. Hyg.*, 2008, 52(7): 577-586.

14. W. Fransman, M, van Tongeren, J. Cherrie, M. Tischer, T. Schneider, J. Schinkel, H. Kromhout, N. Warren, H. Goede and E. Tielemans, Advanced Reach Tool (ART): Development of the mechanistic model, *Ann. Occup. Hyg.,submitted.*
15. J. Schinkel, N. Warren, W. Fransman, M.van Tongeren, P. McDonnell, E. Voogd, J.Cherrie, M.Tischer, H. Kromhout and E. Tielemans, Advanced REACH Tool (ART): Calibration of the mechanistic model. *J. of Env. Monitoring*, 2011, **13,** 1374-1382.
16. J. Cohen, A coefficient of agreement for nominal scales, *Edu. and Psych. Measurement*, 1960, **20** (1): 37–46.
17. J. Fleiss, The Measurement of Interrater Agreement. *Statistical Methods for Rates and Proportions***,** 2003, 2.
18. N. Hawkins and J. Evans, Subjective estimation of toluene exposures: a calibration study of industrial hygienists, *Appl. Ind. Hyg.,* 1989, **6,** 61–68.
19. R. Landis and G. Koch, The Measurement of Observer Agreement for Categorical Data, *Biometrics,* 1977, **33,** 159-174.
20. J. Siemiatycki, L. Fritschi, L.Nadon and M. Ge´rin, Reliability of an Expert Rating Procedure for Retrospective Assessment of Occupational Exposures in Community-Based Case-Control Studies, *American J. of Ind. Med.,* 1997, **31,** 280-286.