

Assessment of occupational exposure to chemicals by air sampling for comparison with limit values: the influence of sampling strategy

Frédéric Clerc^{1*} and Raymond Vincent¹

1. INRS - French National Institute for Health and Safety, 1 rue du Morvan, Vandoeuvre les Nancy 54500, France

* Tel: +33-3-83-50-20-57; Fax: +33-3-83-50-20-97; e-mail: frederic.clerc@inrs.fr

ABSTRACT

This study was aimed at quantifying the impact of sampling duration and the number of measurements taken on the quality of assessing occupational exposure to toluene. To this end, a measurement database was built, based on four campaigns carried out in an industrial printing facility. Five Homogeneous Exposure Groups (HEG) were set up and between 120 and 290 individual measurements lasting from 2 to 8 hours were collected for each of them. These measurements were performed with the objective of comparing them to the 8-hour Occupational Exposure Limit (8h-OEL). The resulting data was used to define a reference exposure profile per HEG: the “gold standard”. This exposure profile corresponds to a log normal distribution of measurements from which compliance/non-compliance with the 8h-OEL decision is derived. To simulate the possible sampling strategies used by industrial hygienists, six scenarios were defined, each containing a different number of measurements: two, three, four, six, nine and twelve measurements performed per HEG, over different working days and different seasons of the year. The measurement values per scenario were simulated by sampling from the real measurement per HEG. For each scenario, 1,000 simulated exposure profiles and corresponding simulated compliance decisions were computed. They were compared to the gold standard compliance decision using statistical indicators. Three methods were used for computing the simulated compliance decision: 1) the 95th percentile must be lower than the 8h-OEL; 2) the exceedance fraction with respect to the 8h-OEL must be lower than 0.1% (as defined by standard CEN 689, Appendix D); 3) the 70% Upper Confidence Limit (70%UCL) of the exceedance fraction with respect to the 8h-OEL must be lower than 5% (as defined by French regulations). The results show that exposure assessment quality increases with both the number of measurements and sampling duration when using the 95th percentile and exposure assessment based on French regulations, whereas it decreases when using the standard. Moreover, guidelines for the efficient evaluation of chemical exposure in the workplace can be drawn up to help professional occupational hygienists. Indeed, boundaries can be recommended regarding the number of measurements and sampling duration necessary to obtain a reliable exposure assessment while minimizing effort devoted to sampling and analysis.

1. INTRODUCTION

Assessing the occupational risks linked to the chemical agents present in working atmospheres is based on two main parameters: the intrinsic danger of the agent in question and the exposure of workers to this agent. An agent can take different exposure routes: ingestion, cutaneous contact and

inhalation. This paper covers only exposure by inhalation of an agent. To protect exposed employees, the 8-hour Occupational Exposure Limit (8h-OEL) values have been defined on the basis of the toxicity specific to each chemical agent. These values are set out in regulations at both European and French levels. Recently, the transposition of several European directives (EU-OSHA, 2000) (EU-OSHA, 2004) (EU-OSHA, 2006) (EU-OSHA, 2009) has resulted in the publication of strict or indicative regulatory Occupational Exposure Limits (OELs) for over one hundred chemical agents. There are two types of 8h-OEL: short-term exposure limits, which are intended to protect workers from the acute toxicity of a chemical agent, and long-term exposure limits TLV-TWA 8h aimed at protecting workers from the effects of long-term toxicity.

To compare the exposure of workers to an 8h-OEL, a sampling strategy must be used (Leidel et al., 1977) (Rappaport et al., 1994)(AFNOR, 1995). The principle consists in defining homogeneous exposure groups (HEGs), namely groups of workers carrying out similar tasks and who are considered to be similarly exposed to a given pollutant. Samples are taken from among the workers of the different HEGs formed. The different concentration measurements make it possible to build an exposure profile. In order to assess the exposure profile with respect to the 8h-OEL, various methods can be applied, such as directly comparing the measurements with the 8h-OEL or using simple indicators like averages. The common approach is to assume that a worker should not be exposed to high concentrations for more than 5% of their working time, taking into account any uncertainty on the measurements. To do this, statistical methods are widely used, with the hypothesis of a log-normal distribution of the measurements. The log normal distribution is defined by two parameters: the geometric mean (GM) and the geometric standard deviation (GSD). Based on the log normal distribution, indicators intended to quantify the exposure of the workers of the HEG can be calculated. In particular, the 8h-OEL exceedance fraction -which must not exceed 5% -and the 95th percentile of the distribution are often used. For each of these statistical indicators, confidence intervals can be derived, and the compliance assessment can be based on the Upper Confidence Limits (UCL) of these intervals (Milz & Mulhausen, 2006) (British Occupational Hygiene Society, 2011).

Assessing the exposure level per HEG is a complex issue as the measurements are subject to considerable variability depending on the emissions generated by the process, the tasks performed by the operator, the season or the type of protective systems employed (Rappaport, 1991). Moreover, sampling implies spatial and temporal variability within an HEG depending on the individual and their exact activity within the HEG (Rappaport et al., 1994)(Symanski et al., 2006). This variability leads to a high degree of uncertainty in the exposure assessment, resulting in the need to take the most representative air samples of the real working conditions and tasks carried out by the operator throughout their shift. This means taking samples whose duration is appropriate to the work and in sufficient numbers to ensure compliance with 8h-OEL. The problem of optimising sampling strategies has been examined from different viewpoints. In particular, two studies tackled the problem of optimizing sampling strategy by estimating the appropriate balance between the number of measurement days versus the number of workers (Lampa et al., 2006) (Chen et al., 2009).

This study concerns the relevance of the choice of the duration and numbers of samples required to ensure the reliability of the exposure assessment. Different scenarios (combinations of number of samples / durations) are explored. For each scenario, simulations are performed and an average “compliance” or “non-compliance” decision is compared to a reference “gold standard” decision.

Three methods are used for establishing the “compliance” or “non-compliance” decision. The first, compares the 95th percentile of the probability distribution to the 8h-OEL, while the second uses the CEN 689 – Appendix D standard in which the 99.9th percentile is compared to the 8h-OEL (i.e. exceedance fraction is compared to a threshold of 0.1%). Finally, the third uses French regulatory texts in which the 70% UCL of the exceedance fraction is compared to a threshold of 5%. The study is based on a case studied in the industrial printing sector where workers are exposed to the toluene contained in the inks.

2. METHODS

2.1. Industrial printing

The exposure measurements were performed in an industrial printing plant using the rotogravure technique. Two workshops were investigated: the first housed three identical production lines, while the second housed only one specialised line set aside for pre-print tests. Each machine uses eight cylinders 3.60 m in width for double-sided, four-colour printing. The machines operate 24 hours a day, with three teams working shifts on the production lines and only one team working on the test line. The average printing time is five hours a shift, depending on the changes made to the print runs or on the problems encountered. Over 150,000,000 kilograms of paper are consumed every year, while between 120,000 and 130,000 kilograms of ink are consumed per week. The ink comprises 50% toluene, which is used as a thinner. The toluene vapours are recovered (95%) by ventilation and a recycling unit comprising a system for trapping toluene vapours on active charcoal. The majority of this toluene is resold, with a small portion being used for cleaning printing line components.

Toluene is a colourless, flammable liquid whose toxicological properties are well known. Acute toxicity is low and chronic toxicity is moderate. Toluene is an R2 category chemical agent classed as toxic to reproduction by the EU (EC regulation no.1272/2008), the 8h-OEL having been lowered to 76.8mg/m³ and the 15min-OEL to 384 mg/m³ on 1 January 2012 (Journal Officiel de la République Française, 2012). However, the results of the present study were obtained before the application of this new 8h-OEL. Thus, the 8h-OEL used in this study was the French one in force before January 2012: 192mg/m³.

2.2. Homogeneous Exposure Groups

The three production lines are strictly identical and none is dedicated to any special process, hence the five HEGs were formed in accordance with this hypothesis. The **winder** supplies the printing machine with paper. **Driver 1** is stationed at the general control panel of the machine, in a ventilated booth of about 20m². **Driver 2** is also stationed in this booth. Their role is to check the quality of the print runs and intervene on the printing machine if needed. The **receiver** is in charge of palletising the printed material. The last HEG concerns two multi-skilled people: the pre-print **test drivers** who perform all the tasks mentioned above and which depend on the ad hoc requirements generated by the tests to be carried out. There were 9 workers per HEG (3 lines * 3 shifts), but the function of test driver was ensured by 6 workers (2 workers * 3 shifts). The night shift was not sampled. The workers within the HEG are considered strictly equivalent; no information is available on which worker was sampled. Nevertheless, to ensure maximum variability, sampling was performed as much as possible on different workers.

2.3. Sampling strategies

Two types of samples were taken simultaneously: a passive sample, the GABIE badge (Delcourt and Sandino, 2001), and an active sample: pump and charcoal tube. Throughout their working day, each worker of each HEG was systematically equipped with a badge and a pump, the tube of which was changed every two hours. The time required to change the tubes was not taken into account. Thus, for one shift, three active samples lasting two hours each and one passive sample for a full eight-hour shift were taken. None of them were non-detects; a total of fifteen 2-h active samples and 4 passive samples analyses were missing for non documented reasons. For these situations, the corresponding full-shift measurement was removed. Due to the sequential nature of tube replacement, two types of composite measurements could be performed: two four-hour time periods and one six-hour time period. Thus four types of measurements are studied in the rest of this paper: measurements performed on full-shift passive sampling and the measurements performed on 2-h, 4-h and 6-h active sampling (Figure 1). Note that the two 4h active samples are not strictly independent since the half-way 2h sample is used in both.

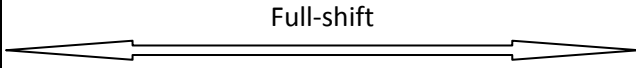
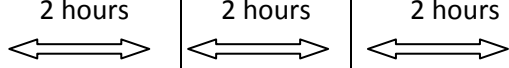

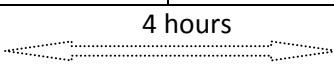
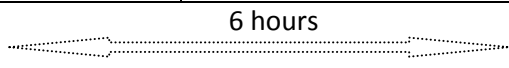
Data type	Duration			Average sampling time	Device
Single measurement				6h57	GABIE passive sampler
3 sequential measurements				2h06	Active sampler
Composite measurement				4h20	Active sampler
Composite measurement				4h09	Active sampler
Composite measurement				6h16	Active sampler

Figure 1. Organisation of measurement strategies: single and composite

The average obtained from three 2-hour measurements represents the mean value obtained in the situation where the IH monitored exposure for 2-h, whereas the 6-h measurement represents an almost full-shift. In the remainder of this paper, the series of measurements are identified as:

- Full-shift – Shift measurement with passive sampler;
- 2-h – Average value of three consecutive 2-h measurements with active samplers (start, half-way, end of shift);
- 4-h – Average value of two consecutive composite 2-h active samples (start to half-way of shift and half-way to end of shift)
- 6-h – Shift measurement obtained with the three consecutive 2-h active samples

The overall campaign consisted of four in-situ interventions lasting 4 days each, organised over different periods over the years 2000 and 2001 (spring 2000 – 57 full-shift samples, autumn 2000 – 55 full-shift samples, winter 2001 – 64 full-shift samples, summer 2001– full-shift 69 samples) in order to integrate seasonal variability.

2.4. Sampling and analytical technique

The sampling devices (badge or tube) were worn by the workers in the immediate vicinity of the upper airways.

The active sampling device is composed of a tube containing two charcoal stages of 100 mg and 50 mg (SKC®, ref. 226-0) connected by a flexible tube to a sampling pump ensuring a controlled flow rate of 200 cm³/min (Gilian®- LFS 113). The workplace air is drawn through both charcoal stages. The first captures the toluene vapours while the second is used as a control of the saturation of the first. The flow rate of the pump is measured at the start and end of sampling by a soap-bubble electronic flow meter (Gilian-Gilibrator®). Determining the quantity of toluene fixed to the GABIE badge or to the charcoal tubes was done by gas chromatography and internal calibration under the following conditions: (1) Desorption of the toluene vapours using 5 ml of carbon sulphide and (2) analysis of the solution by gas chromatography (CPWAX 52 CB capillary column, L = 50 m, or SPB1 column, L = 60 m, at a temperature of 90 °C) with flame ionisation detection (FID), and internal calibration with n-butyl acetate (NIOSH, 2003).

The GABIE badge takes the form of a plastic casing about 4 cm in diameter containing 550 mg of active charcoal. The surface area of radial diffusion is 7 cm². The calculation of employee exposure to toluene was based on the quantity of toluene captured, the toluene diffusion velocity (36.6 cm³/min) and the duration of badge exposure. It was demonstrated that under the same experimental conditions, measurement using the GABIE badge is comparable to that obtained by a tube and pump (Delcourt and Sandino, 2001).

2.5. Exposure profile and method for compliance decision

The decision to be made based on the series of measurements is either “compliance with 8h-OEL” (the null hypothesis Ho) or “non-compliance with 8h-OEL” (alternative hypothesis Ha). Three methods of obtaining a decision are explored, depending on three descriptors with the assumption of a log-normal distribution of measurements per HEG:

1. **The comparison of the 95th percentile category $X_{0.95}$ with the 8h-OEL.** The 95th percentile of the measurements is the concentration value such that 95% of the measurements are below this value (Figure 2, right). Its exact value is classed in five categories depending on the 8h-OEL of toluene, as shown in Table 1 (Mulhausen et al., 2006)

category	definition	Decision
Category 0	$X_{0.95} < 1.92 \text{ mg/m}^3$ (i.e. 1% of the 8h-OEL),	probable compliance with the 8h-OEL: reject Ha
Category 1	$X_{0.95} < 19.2 \text{ mg/m}^3$ (i.e. 10% of the 8h-OEL),	
Category 2	$X_{0.95} < 96 \text{ mg/m}^3$ (i.e. 50% of the 8h-OEL),	
Category 3	$X_{0.95} < 192 \text{ mg/m}^3$ (the 8h-OEL)	
Category 4	$X_{0.95} > 192 \text{ mg/m}^3$	probable exceedance of the 8h-OEL

Table1. $X_{0.95}$ categories, definition and associated decision

2. **The comparison of the 99.9th percentile with the 8h-OEL.** This is inspired by the CEN689 Appendix D procedure. If $X_{0.999} \leq 8\text{h-OEL}$ then the 8h-OEL has probably been

complied with: reject H_a . If $X_{0.999} > 8h\text{-OEL}$ but $X_{0.95} \leq 8h\text{-OEL}$ then no decision can be made: uncertainty. If $X_{0.95} > 8h\text{-OEL}$ then the 8h-OEL has probably been exceeded: H_a is most likely true. In this study, the uncertain and excess categories are lumped together: if $X_{0.999}$ is above the 8h-OEL, then the 8h-OEL is considered exceeded: H_a is most likely true. The exact vocabulary of Appendix D of standard CEN689 (NF, 1995) refers to a threshold of 0.1% and 5% concerning the exceedance fraction. This descriptor can be determined by computation, from the parameters of the log-normal distribution (Figure 2, left) or directly from the series of measurements if the number of values available is large enough.

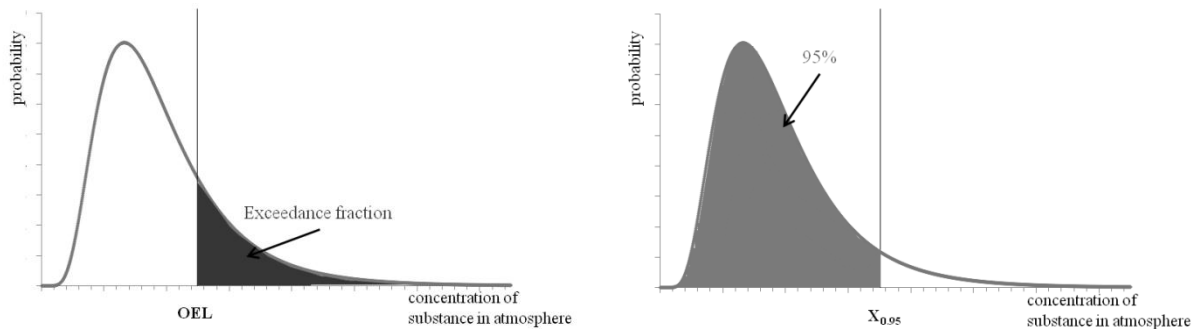


Figure 2. Log-normal distributions. On the left, graphical representation of the exceedance fraction. On the right, graphical representation of the 95th percentile $X_{0.95}$

3. **The French regulatory compliance assessment.** As described in the French regulations, sampling strategy and statistical analysis of the exposure measurement results are used for assessing exposure. It is a two stage procedure.

Stage 1: a minimum of three measurements per HEG are taken during the initial intervention in an enterprise. If all the measurements are below 10% of the 8h-OEL, then the 8h-OEL is probably complied with: reject H_a . If at least one of the measurements is higher than the 8h-OEL then the 8h-OEL is probably exceeded: H_a is most likely true.

Stage 2: If there is no conclusion from stage 1, two other interventions including at least three measurements per HEG are taken in the year at different periods. Every time a measurement exceeds the 8h-OEL, then the 8h-OEL is probably exceeded. When at least 9 measurements have been collected, the 70% upper confidence limit (70%UCL) of the exceedance fraction is calculated. If 70%UCL is lower than 5% (0,05), the 8h-OEL is probably respected: H_a is rejected. Otherwise H_a is most likely true: corrective actions have to be performed (Journal Officiel de la République Française, 2009) (Ogden & Lavoué, 2011).

2.6. Simulation of campaign organisation scenarios

For each sampling duration (2-h, 4-h, 6-h, and Full-shift), different scenarios of different numbers of measurements corresponding to different foreseeable situations are formulated.

Scenario 1: Two measurements per HEG are made the same day, season indifferent.

Scenario 2: Three measurements per HEG are made on three consecutive days, season indifferent.

Scenario 3: Four measurements per HEG are made, two measurements per day on two consecutive days, season indifferent.

Scenario 4: Six measurements per HEG are made, one measurement per day on three consecutive days over two consecutive seasons.

Scenarios 5 and 6: Nine and 12 measurements per HEG are made, one measurement per day on three consecutive days over three and four consecutive seasons. This approach corresponds to the regulatory approach in France.

These scenarios represent the different strategies which could be used by industrial hygienists, depending on the resources available for the exposure assessment.

Once 1,000 runs per scenario have been performed, the corresponding exposure profiles are calculated, averaged and compared to the gold standard exposure profiles. To make this comparison, three metrics are computed, based on the descriptors. These metrics represent the average quality of the “compliant” / “non compliant” decision per scenario, depending on the numerical procedure used:

- **The error when using the 95th percentile category comparison.** For each run, the category of the 95th percentile is computed and a decision is made: category 4 ($X_{0.95} > \text{OEL}$) leads to non-compliance, other categories lead to compliance. Each decision is compared to the gold standard decision. For example, if the gold standard category is 2 ($X_{0.95} < 50\% \text{OEL}$), then a simulation leading to a category 1 ($X_{0.95} < 10\% \text{OEL}$) is correct because it is the same decision: compliance. The percentage of runs resulting in correct and incorrect decisions is then computed for each scenario.
- **The error when using the 99.9th percentile (CEN689 Appendix D standard).** For each run, the decision with regards to the $X_{0.999}$ is computed and compared to the gold standard decision. Following the same principle, the percentage of runs resulting in correct and incorrect decisions is then computed for each scenario.
- **The error when using the French regulatory compliance assessment.** For each run, the simulated decision of compliance or non-compliance is computed and compared to the gold standard decision. When the simulation methodology does not exactly correspond to the specifications (scenarios 1 to 4), the decision from stage 2 is established on the basis of the number of measurements taken, while keeping the conditions of immediate decision from stage 1 active. Again following the same principle, the percentage of runs resulting in correct and incorrect decisions is then computed for each scenario.

3. RESULTS

3.1. Gold standard exposure profile

In Table 2, the exposure profiles and decisions based on all available measurements for each HEG are calculated using the AltrexChimie software tool (INRS, 2010).

	n	Range (mg/m ³)	Arithmetic mean (mg/m ³)	Shapiro-Wilk's test p-value	Geometric mean (mg/m ³)	Geometric standard deviation	value (mg/m ³)	95% CI (mg/m ³) X _{0,95}	cat.	decision	Exceedance fraction	X _{0,999} (CEN689 compliance decision)	70% UCL on exceedance fraction	French regulatory compliance decision	% of meas. > 8h-OEL	
winder	2-h	147	8.06 - 148.82	41.14	23.73%	35.01	1.79	91.49	80.00 - 107.24	2	C	0.2%	U	0,27%	C	0%
	4-h	98	10.78 - 102.69	41.46	20.33%	36.76	1.66	84.67	73.63 - 100.76	2	C	0.06%	C	0,10%	C	0%
	6-h	49	14.21 - 83.47	41.54	15.76%	37.48	1.61	81.62	68.39 - 104.04	2	C	0.03%	C	0,07%	C	0%
	Full shift	50	18.84-88.45	48.05	14.72%	45.30	1.50	86.62	74.84 - 107.02	2	C	0.02%	C	0,05%	C	0%
driver 1	2-h	143	9.46 - 366.48	81.71	20.33%	70.81	1.73	174.89	153.90 - 203.57	3	C	3.75%	U	4,39%	NC*	2%
	4-h	95	20.35 - 198.33	82.1	85.26%	74.56	1.57	156.08	137.66 - 182.58	3	C	1.95%	U	2,48%	NC*	2%
	6-h	47	39.49 - 177	82.42	81.63%	77.32	1.43	139.95	121.98 - 169.31	3	C	0.68%	U	1,07%	C	0%
	Full shift	48	38.55-205.57	91.39	68.85%	87.46	1.43	156.83	134.83-185.52	3	C	1.40%	U	1,75%	NC*	2%
driver 2	2-h	131	19.86 - 568.13	97.16	0.41%	78.08	1.85	214.68	184.99 - 256.66	4	NC	7.6%	NC	8,63%	NC	8%
	4-h	84	24.27 - 555.54	99.67	0.03%	82.1	1.77	209.33	176.89 - 258.66	4	NC	7.22%	NC	8,48%	NC	12%
	6-h	41	39.72 - 381.63	99.07	0.08%	84.89	1.67	196.75	160.00 - 264.28	4	NC	5.94%	NC	7,67%	NC	7%
	Full shift	46	44.47-423.54	115.13	0.29%	101.72	1.64	230.00	186.76-294.69	4	NC	10.03%	NC	12,16%	NC	11%
receiver	2-h	221	2.68 - 192.94	36.36	93.29%	27.66	1.98	80.68	71.14 - 93.26	2	C	0.17%	U	0,22%	NC*	1%
	4-h	145	10.02 - 145.8	36.28	0.01%	30.14	1.78	77.78	68.09 - 91.12	2	C	0.07%	C	0,11%	C	0%
	6-h	71	11.83 - 105.4	36.48	22.18%	31.91	1.66	73.48	62.55 - 90.59	2	C	0.02%	C	0,04%	C	0%
	Full shift	71	17.22-125.53	49.93	3.40%	44.18	1.69	104.77	87.16-128.01	3	C	0.26%	U	0,41%	C	0%
tests driver	2-h	90	4.13 - 254.4	72.11	0.87%	55.73	1.75	205.22	163.47 - 272.79	4	NC	6.21%	NC	7,33%	NC	2%
	4-h	60	12.78 - 199.39	71.29	21.53%	63.16	1.67	147	123.26 - 185.59	3	C	1.67%	U	2,30%	NC*	2%
	6-h	30	20.58 - 142.94	71.16	28.41%	65.78	1.52	131.36	108.07 - 176.85	3	C	0.62%	U	1,12%	C	0%
	Full shift	30	24.12-155.90	78.54	58.09%	74.38	1.50	144.81	117.91-189.59	3	C	0.96%	U	1,67%	C	0%

Table2. Exposure profile parameters. n is the number of samples; “X_{0,95}” is the 95th percentile computed with a log normal hypothesis; 95%CI on the 95th percentile is the 95% confidence interval on the 95th percentile; X_{0,95} category is the AIHA-defined category; X_{0,95} decision is C for compliance and NC for non-compliance; “exceedance fraction” is the exceedance fraction calculated on the basis of a log-normal hypothesis; X_{0,999} (CEN689 decision) is C for compliance, U for uncertainty and NC for non-compliance; the 70% Upper Confidence Limit (UCL) on the exceedance fraction is used for the French regulatory compliance decision, it must not be above 5%; C is for compliance, NC is for non-compliance and NC* is non-compliance because at least one measurement exceeds 8h-OEL; “% excess measurements” is the percentage of measurements exceeding the 8h-OEL.

For the same HEG, the decision of compliance or non-compliance differs depending on the sampling duration. Indeed, the measurement variability (measured through GSD) always decreases with the duration of samples whereas GM always increases with the duration of samples. According to a 1% threshold for the p-value of the Shapiro-Wilk's test, the hypothesis of a log-normal distribution is rejected in all cases for the "driver 2" HEG (this is due to the particular fact that 11% of measurements exceed the 8h-OEL). The log normal distributions of 8-h measurements per HEG are presented in Figure 3.

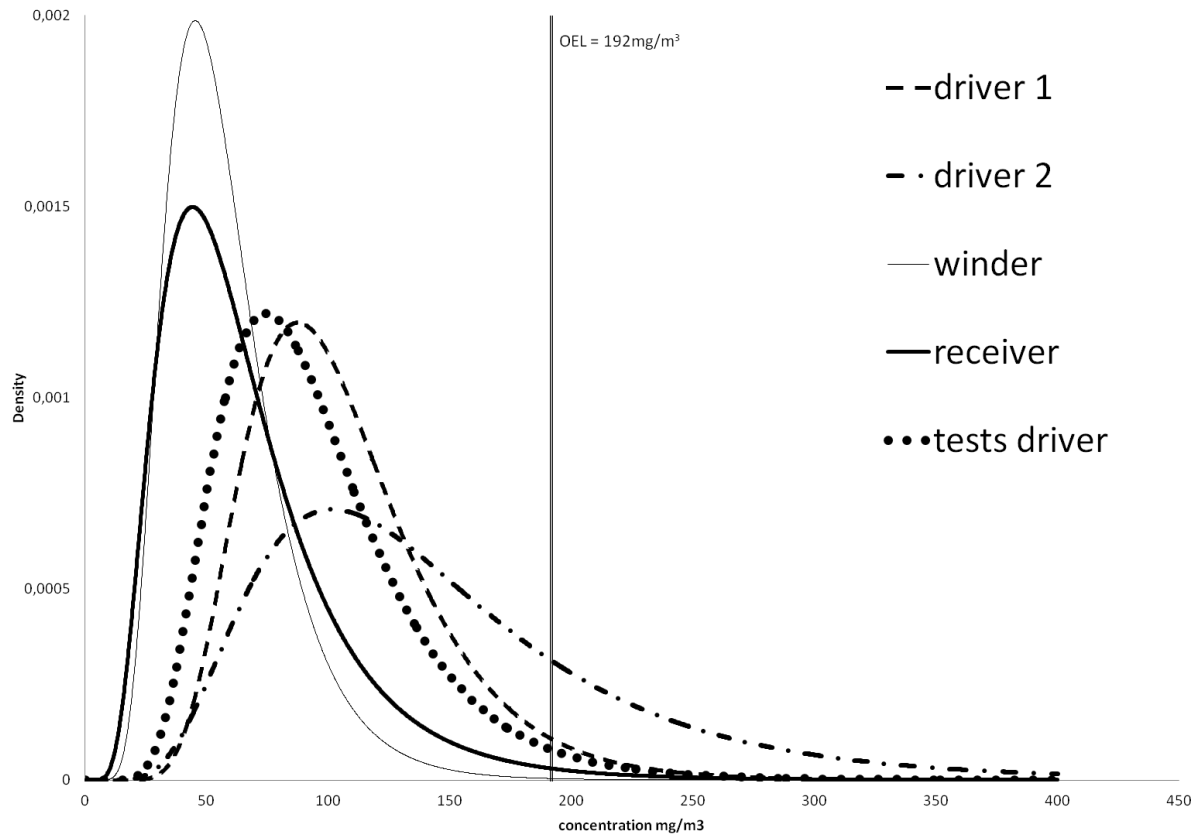


Figure 3. Log-normal distributions of all HEGs: gold standard distributions obtained with the full-shift measurement series used to establish the gold standard exposure profiles.

It is assumed that the series of full-shift measurements collected is fully representative of the usual measurable toluene concentration in the workplace: around 50 measurements per HEG were performed and variability was very low in all cases. Consequently, the gold standard decision is based on the exact percentage of measurements exceeding the 8h-OEL, which must not exceed the threshold of 5%. The conclusion is that:

- The exposure of the workers of the "winder", "driver 1", "receiver" and "test driver" HEGs is probably lower than the 8h-OEL (reject H_a);
- The exposure of the workers of the "driver 2" HEG is probably higher than the 8h-OEL for toluene (cannot reject H_a).

3.2. Simulations

The simulations were performed using algorithms developed with Java 1.7. The probabilistic library used for log-normal computations was JScience 5.0.

When the compliance decision is based on the 95th percentile category, the higher the number of measurements, the better the accuracy. Whatever the duration, the proportion of correct estimations gradually increases to converge towards a plateau from 8-10 measurements onwards. When 10 or more measurements are performed, the full-shift measurements yield the best results, giving a classification error rate of less than 20%. The 4-h and 6-h measurements differed little, resulting in an error rate of less than 30%. The 2-h measurements clearly resulted in a poor error rate, above 40% (Table 3).

Full-shift	47%	37%	39%	33%	26%	24%	22%	20%	19%	17%	16%
6h	50%	43%	38%	35%	32%	31%	29%	27%	27%	26%	26%
4-h	58%	50%	48%	42%	36%	34%	32%	30%	29%	28%	27%
2-h	63%	55%	57%	52%	47%	46%	44%	43%	43%	42%	42%
Number of measurements	2	3	4	5	6	7	8	9	10	11	12

Table3. Average proportion of classification errors of the $X_{0.95}$ with respect to sampling duration and number of measurements. The columns for 5, 7, 8, 10 and 11 measurements are extrapolated linearly from adjacent values. Values above 50% indicate that the proportion of $X_{0.95}$ estimations is more likely to be in the wrong category.

In the simulation with 9 full-shift measurements (20% error rate), the false negative and false positive decision rates per HEG are reported in Table4. A so-called false positive is reported when the correct decision is “compliance” but the computed decision is “non-compliance”. A so-called false negative is reported when the correct decision is “non-compliance” but the computed decision is “compliance”. The false negative is therefore more problematic (Hewett, 2009).

HEG	winder	driver 1	driver 2	receiver	test driver
Gold standard decision	compliance	compliance	non-compliance	compliance	compliance
% false positive	29%	9%	N/A	26.4%	2.1%
% false negative	N/A	N/A	33.4%	N/A	N/A

Table4. Average false positive and false negative rates when the decision is made on the 95th percentile per HEG for scenario 5 (1000 simulated datasets of 9 measurements of 8-h).

When the compliance decision is based on the threshold of 0.1% of the exceedance fraction; the mean percentage of errors increases as the number of measurements increases. Taking 9 full-shift

measurements leads to an error rate of 60%. This also applies to nine 6-h and 4-h measurements, which result in rates of about 60%, while the order of magnitude is 70% for 2-h measurements (Table 5).

Full-shift	41%	50%	42%	49%	56%	57%	59%	60%	61%	61%	62%
6h	38%	41%	39%	45%	52%	52%	53%	54%	54%	55%	55%
4-h	45%	50%	46%	53%	60%	61%	62%	63%	64%	64%	65%
2-h	52%	60%	55%	61%	67%	68%	69%	70%	70%	70%	70%
	2	3	4	5	6	7	8	9	10	11	12

Table5. Average percentage of errors made on an 8h-OEL assessment with respect to the criterion of standard CEN689 (exceedance fraction below 0.1%) in terms of sampling duration and measurement number. The columns for 5, 7, 8, 10 and 11 measurements are extrapolated linearly from adjacent values. The results for the 2-h and 4-h composites are averaged.

In the simulation with 9 full-shift measurements (60% error rate), the false negative and false positive decision rates per HEG are reported in Table 6.

HEG	winder	driver 1	driver 2	receiver	test driver
Gold standard decision	compliance	compliance	non-compliance	compliance	compliance
% false positive	22%	88.2%	N/A	74.4%	85.8%
% false negative	N/A	N/A	31.1%	N/A	N/A

Table6. Average false positive and false negative rates when the decision is made on the CEN689 per HEG for scenario 5 (1000 simulated datasets of 9 measurements of 8-h).

When the quality of exposure assessment is based on French regulations (Journal Officiel de la République Française); the percentage of decision errors decreases as the number of measurements increases. In almost every case, the quality of estimation also benefits from the increase in sampling duration. This is the decision mode that most limits the quantity of errors made compared to the gold standard. Indeed, taking nine full-shift or 6-h samples results in a decision error rate of about 15%, while the order of magnitude is 20% for 4-h samples and 40% for 2-h samples (Table 7).

Full-shift	36%	40%	24%	24%	24%	22%	20%	17%	16%	15%	13%
6h	34%	33%	27%	24%	21%	17%	14%	11%	10%	9%	8%
4-h	41%	41%	36%	34%	32%	29%	26%	23%	21%	20%	18%

2-h	46%	51%	43%	44%	45%	44%	42%	40%	40%	39%	38%
	2	3	4	5	6	7	8	9	10	11	12

Table7. Average percentage of exposure assessment errors as defined by the French regulations in terms of sampling duration and measurement number. The columns for 5, 7, 8, 10 and 11 measurements are extrapolated linearly from adjacent values. The results for 2-h and 4-h composites are averaged.

In the simulation with 9 full-shift measurements (17% error rate), the false negative and false positive decision rates per HEG are reported in Table 8.

HEG	winder	driver 1	driver 2	receiver	test driver
Gold standard decision	compliance	compliance	non-compliance	compliance	compliance
% false positive	0%	32.9%	N/A	7.8%	22.4%
% false negative	N/A	N/A	22.5%	N/A	N/A

Table8. Average false positive and false negative rates per HEG for scenario 5 (1000 simulated datasets of 9 measurements of 8-h).

4. Discussion

The six scenarios explored in this paper were defined empirically. However, the error values from scenario number 2 diverge slightly from the global trend: the number of errors made is higher. This scenario uses a strategy based on 3 measurements per day over 3 consecutive days; whereas other scenarios always include at least two measurements made the same day on two different workers. Although this was not verified, the hypothesis is that the amount of variability accounted for by scenario no.2 is higher than in other scenarios, leading to higher GSD and more errors. In other words, scenario no.2 minimizes autocorrelation (Symanski & Rappaport, 1994)(Wackernagel et al. , 1997). In (Lampa et al., 2006) (Chen et al., 2009), it is shown that sampling a large number of workers is preferred to sampling a large number of days, which is different from our observation: in the study presented in this paper, sampling a large number of days seem to increase the error rate.

The “uncertainty zone” is defined in standard CEN689, Appendix D, as the exceedance fraction between 0.1% and 5% (equivalent to percentiles $X_{0.999}$ and $X_{0.95}$). For this study, the choice was made to strictly comply with the terms of the standard, thereby considering that since the “uncertainty zone” is different from “compliance”, it must be “non-compliance”. The other option consisting in considering the uncertainty zone as “compliance” was not studied but its definition is analogous with the 95th percentile point estimate. The results observed when computing the error when using $X_{0.999}$ (standard CEN689, Appendix D), highlight a large number of errors because most of the simulations led to the uncertainty zone. Concerning the results of table 4 in particular, the exceedance fraction for 3 out of 5 HEGs places them in the “uncertainty zone” (driver1, test driver and receiver). In this case, the “true” decision is “uncertainty”, no matter how many measurements are collected. Since

“uncertainty” is considered as “non-compliance” in this study, the percentage of errors increases with the number of measurements. This highlights the major weakness of CEN-689.

Considering the impact of sampling duration on the assessment of French regulatory compliance (Table7), there is a clear difference between a 2-h sampling duration and a longer duration. The simulations based on the measurements stemming from the 6-h samples performed better, on average, than those stemming from full-shift samples (whereas the gold standard decision is based on full-shift samples). This is due to the immediate decision mode from stage 1 that uses the fact that a single measurement exceeding the 8h-OEL results in a non-compliance decision. Indeed, only 3 values stemming from 6-h samples actually exceeded the 8h-OEL, while 6 values stemming from full-shift samples exceeded the 8h-OEL.

The dataset used in this study was compiled under real conditions, but it is in fact almost a textbook case since the workers’ occupation is strictly defined and the GSDs are very low. The study results do not reflect cases where the occupations of workers from exposure groups are more heterogeneous. Under such conditions, where variability is assumed to be higher, the uncertainty of the compliance decision would also be higher, thus the conclusions would be affected. Moreover, the data available for this study did not allow simulating the situation where 3 measurements are below 10% of the 8h-OEL.

Nevertheless, for a rigorous, full-shift, 8h-OEL compliance assessment, the CEN689-Appendix D procedure is too restrictive due to the use of a very low 8h-OEL compliance threshold: 0.1%. The “95th percentile” and “French regulatory compliance assessment” decision-making procedures allow making similar conclusions.

- Three or fewer samples are insufficiently reliable to yield a correct judgement.
- Samples of 2-h or less are insufficiently reliable to yield a correct judgement.
- It does not appear indispensable to take over 10 measurements as the gain in the quality of judgement obtained is low.

Thus, taking between 4 and 9 6-h or full-shift samples provides a relatively good judgement of exposure. French regulations recommend assessing occupational exposure by following a sampling strategy leading to taking 9 full-shift measurements. It has been noted that in this case study, the decision can turn out to be false in 17% of cases. In the case of the HEG defined as non compliant, the rate of false negative decisions reaches 22.5%, while the average rate of false positive decisions for HEGs defined as compliant ranges from 0% to 22.9%. Nevertheless, by following the same sampling strategy, the use of Standard CEN689 results in 60% errors or uncertainties. The use of a criterion based on the 95th percentile will be erroneous in 20% of cases. Finally, the strategy recommended by French regulations, while not exempt of faults, minimised the number of decision errors in this case study.

5. Conclusion

This paper described the impact of sampling duration and number of measurements on the quality of estimating worker exposure. The collection of a large number of workplace air samples in an industrial printing plant allowed formulating a “gold standard” which was used as an absolute

reference of occupational exposure for five HEGs with low variability. Four HEGs were considered compliant with regards to an 8h-OEL, but the fifth was not. Simulations were carried out to assess average exposure when different sampling number and duration strategies were applied.

With regards to the simulations performed and sampling time, the results show that the quality of exposure estimation is similar for long-term samples (more than 6 hours), regardless of the number of measurements and the type of calculation used for the exposure decision. In contrast, estimating exposure from 2-h samples for comparison with the 8h-OEL is unreliable, and this sampling duration should be abandoned for this specific purpose. In the context of French regulations (9 measurements), 17% of decision errors were made on the basis of long-duration measurements, while 40% of errors were made on the basis of 2-h measurements.

Concerning the number of samples to be taken, the results show that three measurements are insufficient to judge full-shift occupational exposure in a context of low variability. From four measurements onwards, the results gradually improved until reaching a plateau, the inflection taking place between eight and 10 measurements. Even when the resources invested were excellent, with 12 full-shift measurements, the decision error rate in terms of the French regulation is lower but cannot be zero (13% minimum).

For exposure assessment based on a full-shift, when the HEGs are clearly defined with low variability as presented in this study, the strategy set out in the French texts constitutes a compromise providing reasonably reliable exposure judgement while keeping a reasonable number of samples. For a rougher but also faster and less costly exposure assessment under the same conditions, it appears that taking four long-duration measurements spread over two consecutive days is the minimum necessary in order to obtain non-prohibitive error rates (less than 40%). In this case, an in-depth examination of the different indicators, like working conditions and measurement series could provide additional information making it possible to judge whether any over- or underestimation of exposure has occurred. In this context, the judgement of the industrial hygienist is available to compensate for any sampling errors brought about by the sampling strategy. Techniques that combine expert judgement and existing measurements have been developed and their use should increase (Hewett et al., 2006) (Sottas et al., 2009).

6. References

AFNOR (1995). NF EN 689 - Conseils pour l'évaluation de l'exposition aux agents chimiques aux fins de comparaison avec des valeurs limites et stratégie de mesurage - Atmosphères des lieux de travail. Paris.

British Occupational Hygiene Society. (2011). Testing Compliance with Occupational Exposure Limits for Airborne Substances.

Burstyn I, Kromhout H, Kauppinen T, Heikkilä P, Boffetta P (2000). Statistical Modelling of the Determinants of Historical Exposure to Bitumen and Polycyclic Aromatic Hydrocarbons Among Paving Workers. *Ann. Occup. Hyg.*, 44: 43-56.

Chen C-C, Chuang C-L, Wu K-Y, Chan C-C (2009). Sampling Strategies for Occupational Exposure Assessment under Generalized Linear Model. *Ann. Occup. Hyg.*, 53: 509-521.

Delcourt J, Sandino J-P (2001). Performance Assessment of a Passive Sampler in Industrial Atmospheres. *International Archives of Occupational and Environmental Health*, 74: 49-54.

Hewett P. (2009) A 21st century exposure assessment strategy – Minimizing sampling strategy false negative and false positive decisions, *The Synergist* (jan).

Hewett P, Logan P, Mulhausen J, Ramachandran G, Banerjee S. (2006). Rating Exposure Control Using Bayesian Decision Analysis. *Journal of Occupational and Environmental Hygiene*, 10: 568-581.

INRS (2010). *AltrexChimie*. Retrieved from <http://www.inrs.fr/accueil/produits/mediatheque/doc/outils.html?refINRS=outil13>

Journal Officiel de la République Française. (2012). Décret n° 2012-746 du 9 mai 2012 fixant des valeurs limites d'exposition professionnelle contraignantes pour certains agents chimiques.

Journal Officiel de la République Française. (2009). Arrêté du 15 décembre 2009 relatif aux contrôles techniques des valeurs limites d'exposition professionnelle sur les lieux de travail.

Kromhout H, Symanski E, Rappaport S. (1993). A comprehensive evaluation of within and between worker components of occupational exposure to chemical agents. *Ann. Occup. Hyg.*, 37:253-270

Lampa E, Nillson L, Liljelind I, Bergdahl I (2006). Optimizing Occupational Exposure Measurement Strategies When Estimating the Log-Scale Arithmetic Mean Value—An Example from the Reinforced Plastics Industry. *Ann. Occup. Hyg.*, 50: 371-377.

Milz S, Mulhausen J (2006). Descriptive Statistics, Inferential Statistics and Goodness-of-Fit. In Ignacio JS and Bullock WH, *A Strategy for Assessing and Managing Occupational Exposures* (3rd Edition). AIHA, 341-374

Mulhausen J, Damiano J, Pullen EL (2006). Defining and Judging Exposure Profiles. In Ignacio JS, and Bullock WH, *A Strategy for Assessing and Managing Occupational Exposures* (3rd Edition). AIHA, 47-72.

OSHA. (2000). Indicative Occupational Exposure Limits. 2000/39/EC.

OSHA. (2004). Protection of workers from the risks related to exposure to carcinogens or mutagens at work. 2004/37/EC.

OSHA. (2006). Indicative Occupational Exposure Limit Values. 2006/15/EC.

OSHA. (2009). Indicative Occupational Exposure Limit Values. 2009/161/EU.

Rappaport S (1991). Assessment of Long-Term Exposures to Toxic Substances in Air. *The Annals of Occupational Hygiene*, 35(1).

Rappaport S, Lyles R, Kupper L (1994). An Exposure Assessment Strategy Accounting for Within- and Between- Worker Sources of Variability. *Ann. Occup. Hyg.*, 39 : 469-495.

Sottas PE, Lavoué J, Bruzzi R, Vernez D, Charrière N, Droz PO. (2009). An empirical hierarchical Bayesian unification of occupational exposure assessment methods. *Statistics in Medicine*, 1: 75-93.

Symanski E, Maberti S, Chan W (2006). A Meta-Analytic Approach for Characterizing the Within-Worker and Between-Worker Sources of Variability in Occupational Exposure. *Ann. Occup. Hyg.*, 50: 343-357.

Symanski, E., & Rappaport, S. (1994). An Investigation of the Dependence of Exposure Variability on the Interval between Measurements. *Ann. Occ. Hyg.*, 38(4), pp. 361-372.

Wackernagel H, Lajaunie C, Thiéry L, Vincent R, Grzebyk M (1997). Applying geostatistics to exposure monitoring data in industrial hygiene. In Soares A, Gomez-Hernandez J, Froidevaux R, *geoENV I - Geostatistics for Environmental Applications*, Kluwer, 463-476.