



HAL
open science

Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA

Gautier Mater, Christophe Paris, Jérôme Lavoué

► **To cite this version:**

Gautier Mater, Christophe Paris, Jérôme Lavoué. Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA. *American Journal of Industrial Medicine*, 2016, 59 (5), pp.379-391. 10.1002/ajim.22569 . hal-01414267

HAL Id: hal-01414267

<https://hal.science/hal-01414267>

Submitted on 12 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA

Authors:

Gautier Mater, MS,¹ Christophe Paris, PHD,² and Jérôme Lavoué, PHD³

¹Institut National de Recherche et de Sécurité (INRS), Department of Pollutants Metrology, Vandœuvre-lès-Nancy, France

²Centre Hospitalier Régional Universitaire de Nancy (CHRU), Occupational Diseases Department - INGRES EA7298, Vandœuvre-lès-Nancy, France

³Université de Montréal, Department of environmental and occupational health, Montréal, Québec, Canada

Manuscript word count:

Main manuscript: 4,615 words

Author Contributions Statement:

The 3 authors contributed significantly to the conception of the study and the drafting of the manuscript. They approved the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA

Abstract

Background

Several countries have built databases of occupational hygiene measurements. In France, COLCHIC and SCOLA co-exist, started in 1987 and 2007 respectively.

Methods

A descriptive comparison of the content of the two databases was carried out during the period 1987-2012, including variables, workplaces and agents, as well as exposure levels.

Results

COLCHIC and SCOLA contain respectively 841,682 (670 chemicals) and 152,486 records (70). They cover similar industries and occupations, and contain the same ancillary information. Across 17 common agents with >500 samples, the ratio of the median concentration in COLCHIC to the median concentration in SCOLA was 3.45 [1.03-14.3] during 2007-2012. This pattern remained when stratified by industry, task, and occupation, but was attenuated when restricted to similar sampling duration.

Conclusions

COLCHIC and SCOLA represent a considerable source of information, but result from different purposes (prevention, regulatory). Potential differences due to strategies should be evaluated when interpreting data from these databases.

Keywords: exposure database, risk assessment, sampling strategy, workplace exposure, chemical agents

Figure legends:

Figure 1 Distribution of the number of samples in COLCHIC and SCOLA per year

Figure 2 Proportion of the samples (%) in COLCHIC and SCOLA for industrial sectors and comparison with the population by industrial sector in 2012

Appendices as supplementary files for review:

Appendix I: Measurement distribution curves by database for the period 2007 – 2012 part 1, part 2 and part 3

Appendix II: Comparison of exposure levels in classes of activity common to COLCHIC and SCOLA over the period 2007 to 2012 restricted to the sample duration between 240 and 480 minutes

INTRODUCTION

Objective measurements of workers' exposure play an important role in occupational health. First, making a full inventory of chemical agents present in a workplace is a major challenge for risk assessment. This task may be facilitated by the availability of existing measurement data [Bégin et al., 1995]. Moreover, such data enable the implementation of targeted prevention policies by helping prioritize industries and occupations. Lastly, exposure measurements are particularly important as a source of information for carrying out epidemiological studies, for toxicological risk assessment, and to support the development of exposure models.

Aiming at managing large quantities of data generated following the collection of air samples during various occupational health related governmental activities, several large occupational exposure databases have been set up. The oldest contain data dating from the 1970s, such as the Integrated Management Information System (IMIS) from the US Occupational Safety and Health Administration (OSHA) [Stewart et al., 1990] and the MEGA database from the German Institute for Occupational Safety (BIA) [Stamm, 2001]. MEGA and IMIS both include more than 1.6 million data points [Henn et al., 2011; Koppisch et al., 2012]. Others are more recent, such as the NEDB (1986) in the United Kingdom [Burns et al., 1989], SIREP (1996) in Italy [Scarselli et al., 2007], and COLCHIC (1986) [Carton, 1995; Vincent et al., 2001] and SCOLA (2007) in France.

There are many and varied examples of ways in which these databases have been used. For instance, Lavoué et al. [2011] estimated levels of exposure to formaldehyde in various industries from the COLCHIC and IMIS databases. Linch et al. [1998] presented an estimation of the number of workers exposed to crystalline silica. Hamm et al. [2011]

created a job-exposure matrix for beryllium using IMIS. Pesch et al. [2015] evaluated exposure of welders to hexavalent chromium using the data recorded in MEGA. Scarcelli et al. [2012] assessed occupational exposure to hexavalent chromium from the SIREP database in Italy. Coble et al. [2010] created a portrait of chemical agents in the pulp and paper industry using IMIS data.

In France, two occupational exposure databases co-exist. COLCHIC was set up in 1986, using data from prevention and research activities; SCOLA was set up in 2006, using data from compliance activities. COLCHIC and SCOLA come from the same population of workers, but result from different goals, possibly corresponding to different sampling strategies. It is therefore conceivable that they provide different pictures of exposure. No empirical assessment has yet been performed to evaluate this issue.

The objective of this article is to present the first comparative description of the two French databases, COLCHIC and SCOLA, as well as a preliminary assessment of potential differences in exposure levels recorded in these databases.

METHODS

Origin of COLCHIC and SCOLA

The first French database of occupational exposure to chemical products (COLCHIC) was established in 1986 at the request of the Commission of Work Accidents and Occupational Diseases. It is managed by the Institut National de Recherche et de Sécurité (INRS). This database contains results of measurements in the air of workplaces by eight French regional health insurance funds, interregional chemical laboratories, and the laboratories of the INRS within the framework of occupational risk prevention. The objectives of setting up COLCHIC were to centralize measurement data from various laboratories, to harmonize workplace sampling and air analysis methods, and to support chemical risk assessment in France. The data in COLCHIC led to several research projects focusing, for example, on exposure to lead, formaldehyde, and mineral fibers [Carton, 1995; Vincent et al., 2002; Kauffer et al., 2007; Lavoué et al., 2006], as well as on exposure prediction models. [Clerc et al., 2015].

The second French occupational exposure database (SCOLA) results from the French requirement that measurements made during assessment of compliance to regulatory occupational exposure limit values (OELV) should be archived in a national register. The measurements stored in SCOLA are performed by certified laboratories, independent of the establishment controlled. The data are used to define priorities for national prevention programs. SCOLA was recently used to update a Web application about *a priori* levels of asbestos dustiness (<http://scolamiante.inrs.fr>).

Coding of the information in COLCHIC and SCOLA

Ancillary information in COLCHIC can be classified into four major categories: administrative information, sampling conditions, exposure conditions, and analytical conditions. Most of the data are coded using standardized classifications specific to COLCHIC, such as task or ventilation, or using national classifications (industries and occupations as coded by the French national systems: classification of activities (NAF) [INSEE, 2015] and operational directory of occupations and jobs (ROME) [Pôle emploi, 1999], respectively). In 2002, a second version of COLCHIC was developed [Vincent et al., 2001]. A new coding system was established with the modification and/or replacement of existing variables and the creation of new variables in accordance with recommendations of specific work groups [ACGIH-AIHA, 1996; Rajan et al., 1997]. For instance, since 2002, COLCHIC has contained the worker's occupation, and the category of product or process giving rise to the occupational exposure. The original "Individual protection," "collective protection," "global ventilation," "representativeness," and "type of work place" variables were modified, with either new categories being created or existing ones merged or their definitions altered.

SCOLA was built using the same software system as COLCHIC. The coding of ancillary information is very similar to that in COLCHIC. Table I summarizes the main differences between the two systems.

Circumstances leading to records in COLCHIC and SCOLA

The data collected in COLCHIC come from measurement campaigns performed in establishments from the national social security scheme. The choice of targets leading to

the measurements in COLCHIC stems from general prevention programs defined by a period of four years by the national health insurance system, as well as from national sampling surveys such as for beryllium (2005-2007), chromium (2012-2014), or styrene (2013-2016). The eight interregional chemical laboratories are assigned to defined geographical areas, and cover companies insured by the national social security scheme (~55% of companies in France; public services, agriculture, and small retail and artisans are excluded). Three main situations can trigger a visit to a company. First, a visit can be requested by a safety controller or engineer from the regional health insurance fund. The origin of such a request is highly variable: it may come from an establishment, from an employee representative, or from an occupational physician. Second, visits can be made for research purposes (e.g., industry-wide surveys) in the French prevention network. In this context, companies are usually contacted by the researchers. Third, a visit can occur within the framework of specific prevention actions performed by each regional laboratory. The laboratory will then identify and contact companies to be visited. Establishments cannot refuse to be visited, but the visits are scheduled in advance.

The identification of companies in SCOLA is different. Certified laboratories take measurements at the request of companies in order to fulfill their regulatory obligations. In France, the company director is legally responsible for risk assessment. An initial evaluation is performed by company experts or consultants, which results in a document identifying situations with a potential for overexposure. For each of the situations potentially at risk for which an OELV exists, a measurement strategy has to be elaborated in compliance with the regulation. The current law requires a minimum of three and up to

nine samples representative of the workers' exposure in order to verify compliance to an OELV [Le ministre du travail, de la famille, de la solidarité et de la ville, 2009].

Descriptive summary of COLCHIC and SCOLA

The descriptive analysis included data for all agents from 1987 for COLCHIC and from 2007 for SCOLA, up until 2012. Several quality criteria were defined to select data. First, the analysis was restricted to quantitative air measurements performed in the worker's breathing zone (personal sample) or at a fixed point (area sample). Second, units had to be expressed as an air concentration. Measurements with sampling duration less than 60 minutes were identified as "Short term," and all other records as "Long term". Only administrative variables relevant for exposure assessment were included in the dataset. Partial samples (using several sequential samples to evaluate a full day) are treated differently in COLCHIC and SCOLA. In SCOLA, sequential samples are aggregated into a single time weighted average (TWA) value before being recorded. In COLCHIC, sequential samples are entered separately, but information to link them is not well or consistently coded. Using the available information to aggregate partial samples together would require manual processing of each record, which was not possible in the current study.

After data extraction, descriptive statistics were calculated in parallel for both databases. We examined the number of records and their distribution across levels of all variables common to COLCHIC and SCOLA. We also identified the industrial sectors covered by both databases in relation to the French workforce. Finally, we calculated summary statistics of exposure levels for all agents in both COLCHIC and SCOLA with at least 500 records in each database.

Comparison of exposure levels between COLCHIC and SCOLA

The comparison of exposure levels between COLCHIC and SCOLA was performed in three parts.

First, for each of the agents having at least 500 measurements in both databases, empirical cumulative distribution curves (ECDF) were created after restriction to personal samples over the common period 2007 – 2012. Area samples were not included because they can be very different from personal samples, and are infrequent in SCOLA except for asbestos.

Second, to take into account potential confounding due to differences in industries included in the two databases, we identified for each chemical agent a list of industries having at least 20 measurements both in COLCHIC and SCOLA. We chose 20 as a threshold because it corresponded approximately to the number required to cover more than one establishment. We then calculated for each agent and each industry category the proportion of values exceeding the OELV (exceedance), the geometric mean (GM) of recorded concentrations, and the proportion of data lower than the limit of quantification (proportion of non-detects) in COLCHIC and SCOLA. Calculation of the geometric mean when a data was reported as <LOQ was performed using regression on order statistics (ROS) as described in [Helsel, 2012]. In situations where fewer than three detected values were available, GM was estimated by replacing non-detects with $LOQ/2$. For each chemical agent, we evaluated agreement between the two databases across the included industry categories by determining: 1) the mean difference in proportions of data exceeding the OELV, 2) the mean relative difference in GM values, and 3) the mean difference in proportions of data reported below LOQ.

In both COLCHIC and SCOLA, the data on the industrial activity are codified by way of a standardized classification using three variables: industry, occupation, and task. The previous analysis was also performed using the task and occupation classification systems in place of industry.

Third, we compared exposure levels over the common period 2007 – 2012 within establishments present in both COLCHIC and SCOLA. For each chemical agent present in both databases in at least three common establishments, we calculated, for each of these establishments, the ratio of the geometric mean of measurements in COLCHIC ($GM_{COLCHIC}$) to the the geometric mean of measurements in SCOLA (GM_{SCOLA}). Asbestos data were excluded as it was shown that phase-contrast microscopy (used in SCOLA) is not comparable to transmission electron microscopy (used in COLCHIC) [Clerc et al., 2011].

All analyses were performed using the R 3.0.3 statistical software (R Development Core Team, Vienna, Austria).

RESULTS

Descriptive summary of COLCHIC and SCOLA

During the period 1987 – 2012, 841,682 air measurement results from 670 agents were recorded in COLCHIC by eight interregional chemical laboratories, covering 49,145 interventions in 37,569 different companies. From 2007 to 2012, 152,486 air measurement results from 70 agents were recorded in SCOLA by 86 different certified laboratories, covering 37,277 interventions in 7,449 different companies. Eight percent (n=70,937) of COLCHIC extracted data were eliminated before analysis according to quality criteria (3.5%, objective was not to characterise the exposure of workers, 4.5% unit was not expressed as an air concentration). All extracted records were used for analysis in SCOLA. Table II presents the distribution of records in both databases according to categories of the main ancillary variables.

A median of 31,917 measurements per year were recorded in COLCHIC, relatively stable over the entire study period (Figure 1). Measurements appeared in SCOLA in 2007, starting with 3,598 in 2007 and steadily increasing to 52,202 in 2012. While COLCHIC contains a mix of area and personal samples (ratio personal/area 1.44), the compliance related data in SCOLA are much more frequently personal samples (ratio personal/area 2.23, increasing to 34 when excluding data for asbestos).

SCOLA contains a large quantity of asbestos measurements (46% of all data) due to the fact that this contaminant's regulation has warranted measurements for a longer period than other agents. Most samples correspond to monitoring activities for containment and removal of asbestos-containing materials, or activities involving interventions on materials or equipment that may release asbestos fibers.

There are also fewer measurements reported as <LOQ in COLCHIC compared to SCOLA (ratio <LOQ/detected 0.31 for COLCHIC vs. 1.13 for SCOLA).

In terms of sampling duration, median sampling duration for personal samples over the entire period covered in COLCHIC was 35 min. for short term samples, and 172 min for long term samples. For area samples, these numbers were 39 min. for short term samples, and 180 min for long term samples. When restricted to the period common to both databases (2007 to 2012), sampling durations in COLCHIC and SCOLA, for personal samples were 28 and 15 minutes, respectively, for short term samples, and 215 and 377 minutes, respectively, for long term samples. For area samples, these numbers were 41 and 27 minutes, respectively, for short term samples, and 195 and 60 minutes, respectively, for long term samples.

A comparison between the distribution of the total number of workers in France per broad industry sector obtained from the National Institute of Statistics and Economic Studies (INSEE) and the distribution of the number of measurements performed in COLCHIC and SCOLA is presented in Figure 2. The three most frequent industries in COLCHIC are manufacture of metal products, manufacture of rubber and plastic products, and manufacture of chemicals and chemical products. They represent 27% of the data. In SCOLA, specialized construction activities, remediation activities and other waste management services, and manufacture of other non-metallic mineral products are the three most visited industries. They represent 45% of the data. Globally, 69% of industry categories included in COLCHIC or SCOLA are common to the two databases, representing more than 95% of the data collected.

As occupation was recorded only after 2002 in COLCHIC, this variable is “missing” for 62% of records for the total period from 1987 to 2012. Welder, surface treatment worker and printing press operator are the three most frequent occupations in COLCHIC and represent 8% of the data. In SCOLA, the three main occupations are depolluting treatment officer, production unit supervisor in the chemical and energy industry, and machine tool operator. They represent 50% of the measurements. Globally, 70% of occupation categories included in COLCHIC and SCOLA are common to the two databases, representing 46% of records.

The three most frequent tasks coded in COLCHIC, application of air-pressure spraying, machining by removing material, and blending operator and controller, represent 13% of the data. In SCOLA, other carcassing work, other special works, and other finishing and maintenance works represent 35% of the data. Globally, 65% of task categories included in COLCHIC or SCOLA are common to the two databases, representing 90% of the data collected.

Seventeen chemical agents common to COLCHIC and SCOLA and with >500 measurements represent 38% and 88%, respectively, of records in the databases. The three most measured agents in COLCHIC were inhalable particulate fraction (38,304 personal, 23,476 area), toluene (20,274 personal, 10,935 area), acetone (19,790 personal, 8,678 area), and for SCOLA asbestos (45,001 personal, 25,205 area), wood dust (16,055 personal, 485 area), alveolar particulate fraction, (6,697 personal, 173 area). The proportion of personal measurements (Table III) reported below LOQ had a median of 15% across the 17 agents in COLCHIC (ranging from 1%, anhydrous ammonia, to 86%,

cristobalite), and of 58% in SCOLA (ranging from 13%, anhydrous ammonia to 92%, cristobalite).

Comparison of exposure levels between COLCHIC and SCOLA

The empirical cumulative distribution curves of personal measurements by database and for each agent (Table III) are presented in Appendix I. Across the 17 chemicals the ratio of the median concentration in COLCHIC to the median concentration in SCOLA varied between 1.03 (butanone) and 14.3 (n-hexane), with a median of 3.45.

The comparison between personal exposure levels in COLCHIC and SCOLA [2007-2012] in common industries, occupations, and tasks is shown in Table IV. Results are presented in three parts: 1) the probability of exceeding the OELV; 2) the geometric mean; 3) the proportion of non-detected. Agreement between COLCHIC and SCOLA varied across agents, activity class, and metric, but generally indicated higher exposures in COLCHIC compared to SCOLA. Hence, for the industry category, the median of the relative bias of the geometric mean was +38%, the median bias for proportion of non-detects was 3.72, and the median bias for % above the OEL was 0.62%. Because sampling duration was usually higher in SCOLA compared to COLCHIC, and longer samples have often been associated with lower exposure levels, we also performed the comparison restricted to sample durations between 240 and 480 minutes in both databases, corresponding to between half and 100% of an 8-hour shift (this interval was selected as a compromise between representativeness of full-shift and sufficient sample size). The results are presented in appendix II. They show a decrease, albeit a partial one, in the amplitude of the pattern seen in Table IV.

The comparison of exposure levels in establishments common to COLCHIC and SCOLA is presented in Table V. Results are presented by agent in three parts: 1) the number of common establishments; 2) the number of measurements in each database; 3) the median ratio of geometric mean across the common establishments. Nine chemical agents met the inclusion criteria. Exposure levels were similar for wood dust and 1-methoxy-2-propyl acetate, showed higher exposure levels in COLCHIC for quartz, cristobalite, lead, toluene, xylene, and ethylbenzene [1.5-15], and lower exposure levels in COLCHIC for acetone. Fourteen other chemical agents had one or two establishments in common. The median of their median ratio values was 1.6.

DISCUSSION

The originality of this study resides in the comparison of the two French databases COLCHIC and SCOLA, with data coming from the same base population. COLCHIC and SCOLA share a common objective of centralising and archiving exposure data to be used as a source of information to support occupational health prevention efforts. Taken together, nearly a million data points are recorded in COLCHIC and SCOLA, putting them among the largest international occupational exposure databases covering an extended time period, such as IMIS in the United States and MEGA in Germany. As such they represent a remarkable source of information on exposure levels.

Descriptive comparison of COLCHIC and SCOLA

COLCHIC and SCOLA include data measured from the same industrial settings, and share a similar set of extensive ancillary information. Some COLCHIC variables evolved historically through taking into account recommendations formulated by international work groups [ACGIH-AIHA, 1996; Rajan et al., 1997]. This represents an improvement for future data interpretation, and is implemented in SCOLA. However, these changes (fusion, creation of new categories) represent an additional challenge to interpreting historical data mixing the old and new definitions. For instance, for the type of workplace variable, three categories (“Confined space”, “Underground gallery, tunnel” and “Other, non classified types”) were added in 2002 to the four initial categories (“In open air”, “Booth or control room”, “Standard industrial premises” and “Partially open”). When analyzing the relationship between type of workplace and exposure levels, the fact that e.g. “confined space” samples were coded in another category from 1986 to 2002 (maybe

“standard industrial premises”) might affect the results. It is however hard to predict in advance the amplitude of such phenomenon.

The yearly rate of measurements archived in COLCHIC has been approximately stable at around 30,000 since 1987. SCOLA only started recently, and accumulated up to 2012 a less impressive 150,000 data points compared to COLCHIC’s 870,000. However, yearly rates are increasing each year (up to 60,000 in 2012), and should continue to grow as the regulatory requirements to measure exposure and send the results to SCOLA to prove compliance are progressively enforced in France. SCOLA will therefore increasingly become the major source of information in the next decades.

The number of interventions recorded in COLCHIC and SCOLA by industry is not correlated with the number of employees working in these industries, although most industries have been the object of measurements. This observation is linked to the absence of exposure to chemical products in certain industries, and to the fact that measurement campaigns are carried out in priority industries. Regarding industries covered in COLCHIC compared to SCOLA, while the two databases do not share the same distribution of measurements across industries, 95% of measurements are in common categories. Differences in industries and activities covered in the two databases might be related to their different purposes (prevention for COLCHIC and regulation for SCOLA); it renders their joint use valuable since the information collected in one database can offset the absence of information in the other.

Over the period common to both databases, 2007 – 2012, the duration of sampling was significantly shorter in COLCHIC than in SCOLA. While this difference can be partly due to not being able to aggregate sequential samples in this analysis, it is likely also

explained by the two databases' respective objectives. Hence SCOLA is related to regulatory compliance assessment, with strict sampling guidelines. In particular, there are very few short term samples in SCOLA, reflecting the smaller number of short term OELV in the current French regulations. On the other hand, COLCHIC is the result of measurements taken by insurance hygienists making targeted interventions, more likely focused on specific tasks within a day. In addition, there was a constant increase in sampling duration in COLCHIC over the entire study period, likely reflecting evolving exposure assessment practices and analytical sampling constraints. Since sampling duration has often been associated with exposure levels in large databases of exposure measurements [Lavoué et al., 2011; Clerc et al., 2014; Clerc et al; 2015], with longer sampling time associated with lower exposure levels, these differences should be taken into account when interpreting COLCHIC and SCOLA.

Another distinction between COLCHIC and SCOLA is the types of samples, with personal samples approximately twice more frequent than area samples in COLCHIC. In SCOLA, area samples are virtually absent when excluding asbestos (which depends on a particularly regulation).

Comparison of exposure levels between COLCHIC and SCOLA

While COLCHIC and SCOLA jointly represent a considerable source of exposure information, their data were collected for different purposes. COLCHIC data are measured in companies within industries targeted as potentially problematic, while SCOLA measurements come from companies mandating certified laboratories to measure exposures in their facilities for regulatory compliance assessment. It is therefore questionable whether these differences might lead to differences in exposure levels.

The comparison empirical cumulative distribution curves show that the concentrations recorded in COLCHIC are higher than in SCOLA for the majority of the agents included in the comparison, with a median ratio of the 50th percentile of concentrations around 3. On average, 22% of the measurements of the 17 common chemical agents are lower than the LOQ in COLCHIC versus 51% in SCOLA, and 14% are higher than the OELV in COLCHIC versus 5% in SCOLA. Moreover, comparisons of exposure levels within common industry, task, or occupation categories also showed a similar, albeit more variable trend, with higher proportions of measurements above the OEL, higher geometric means, and lower proportions of measurements reported as below the LOQ. In certain circumstances, exposure levels are very similar in COLCHIC and SCOLA (e.g., acetone, cristobalite) while in others they are very different (e.g., toluene, lead). An analysis restricted to longer sampling time in COLCHIC, limited by a smaller number of agents included, did reduce, but did not eliminate the pattern of our original observations. This pattern of higher exposures in COLCHIC was also observed in the comparison of exposure levels within common establishments (the median of the median ratio values $GM_{COLCHIC}/GM_{SCOLA}$ for the nine included agents was 1.9).

Higher exposure levels in COLCHIC is not implausible, since it contains data presumably measured in situations where potential risk was suspected, whereas compliance must be verified in all situations where a contaminant is deemed present in the workplace. Indeed, a majority (67%) of measurements in COLCHIC were undertaken with the reason for sampling being “possible risk of exposure”. It is therefore possible that COLCHIC would reflect the higher tail of exposure distribution compared to what is found in SCOLA. Very few reports have been published comparing large exposure databases to evaluate whether

they provide similar exposure portraits [Vinzents et al., 1994; Clerc et al. 2014; Clerc et al., 2015]. Recently, a descriptive comparison of formaldehyde measurements between MEGA and COLCHIC [Clerc et al. 2015] showed variable differences in median concentrations around a factor from 1.6 to 2 for the five most frequent industries in both databases, interpreted as “comparable exposure levels” by the authors. Similar conclusions were reached by Lavoué et al. [2011] for formaldehyde exposure levels across several industries in the IMIS and COLCHIC databases, with moderately higher levels in COLCHIC (+14%). Peters et al. [2011] assembled and analysed an international database of crystalline silica measurements. In their statistical models, differences between regions corresponded to predicted geometric means within a factor of 4. Our own observations lay within the differences reported in the literature. They warrant caution when jointly interpreting COLCHIC and SCOLA. Multivariate analysis, such as used in [Lavoué et al., 2011] or [Peters et al., 2011] should allow evaluation of differences between the two databases while taking into account other ancillary information. This would improve the quantification of these differences and help select approaches to better combine them.

Limits of the study

Several limits of this study can be noted: firstly, SCOLA is relatively recent. Therefore the comparison was hampered by smaller sample sizes in SCOLA compared to COLCHIC, causing variability in the stratified comparisons. This limitation should be overcome in future updated comparisons, since the number of records entered each year in SCOLA is rapidly growing. Second, the agglomeration of sequential measurements in COLCHIC needs manual intervention, which could not be done for all chemicals agents

included in our analysis. This might have affected the comparison of exposure levels between the two databases, although stratified analyses restricted to higher sampling duration in both databases confirmed the main observations, albeit with decreased amplitude.

CONCLUSION

COLCHIC and SCOLA are two databases of occupational exposure to chemical products which centralize data stemming from the same population of workers in France. The potential uses of these databases for prevention are numerous: improved information about *a priori* exposure levels when assessing risk within a company, strengthened elaboration of preventive policies, increased retrospective exposure information for epidemiology and worker compensation. The fact that COLCHIC and SCOLA share a similar structure, cover similar agents, and may complement each other in terms of workplaces visited, is a strong argument for their joint interpretation. This study also showed differences in exposure levels recorded in COLCHIC and SCOLA, possibly related to the difference in purpose between these two databases (prevention vs compliance). Future comparisons using multivariate analyses [Sauvé et al., 2012], permitting us to simultaneously take into account relevant ancillary variables, should refine this initial portrait and provide insight into how to best interpret the two databases jointly.

ACKNOWLEDGEMENTS

This study was funded as a part of the INRS internal research program.

REFERENCES

AIHA. 1996. Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise. *Appl Occup Environ Hyg* 11: 1294-1311.

Bégin D, Gérin M, Adib G. 1995. Development of an Occupational exposure Data Bank on the Territory of a Department of Community Health in Montréal. *Appl Occup Environ Hyg* 10: 355-360.

Burns D, Beaumont P. 1989. The HSE National Exposure Database (NEDB). *Ann Occup Hyg* 33: 1-14.

Carton B. 1995. COLCHIC Chemical Exposure Database: Information on Lead and Formaldehyde. *Appl Occup Environ Hyg* 10: 345-350.

Clerc F, Bertrand N, Vincent R. 2015. TEXAS: a Tool for EXposure ASsessment—Statistical Models for Estimating Occupational Exposure to Chemical Agents. *Ann Occup Hyg* 59: 277-291.

Clerc F, Eypert Blaison C, Guimon M, Romero-Hariot A, Vincent R. 2011. Campagne de mesures d'exposition aux fibres d'amiante par microscopie électronique à transmission analytique (META) INRS 1: 1-164.

Clerc F, Steinhausen M, Bertrand N, Vincent R, Gabriel S, Van Gelder R. 2015. Comparison of formaldehyde exposure measurements stored in French and German databases. *Gefahrstoffe - Reinhaltung der Luft* 75 (2015) Nr. 4 - April: 119-126.

Clerc F, Vincent R. 2014. Assessment of Occupational Exposure to Chemicals by Air Sampling for Comparison With Limit Values: The Influence of Sampling Strategy. *Ann Occup Hyg* 58: 437-449.

Coble J, Lees P, Matanoski G. 2010. Time Trends in Exposure Measurements from OSHA Compliance Inspections of the Pulp and Paper Industry. *Appl Occup Environ Hyg* 16: 263-270.

Databases A-ATGoOE. 1996. Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise. *Appl Occup Environ Hyg* 11: 1294-1311.

Hamm M, Burstyn I. 2011. Estimating Occupational Beryllium Exposure from Compliance Monitoring Data. *Arch Environ Occup Health* 66: 75-86.

Helsel D. 2012. *Statistics for Censored Environmental Data Using Minitab and R*, 2nd Edition.

Henn SA, Sussell AL, Li J, Shire JD, Alarcon WA, Tak S. 2011. Characterization of lead in US workplaces using data from OSHA's integrated management information system. *Am J Ind Med* 54: 356-365.

INSEE. 2008. French classification of activities - NAF Rev. 2, 2008 (second edition, 2015).

Institut National de Recherche et de Sécurité. MetroPol.

Kauffer E, Vincent R. 2007. Occupational Exposure to Mineral Fibres: Analysis of Results Stored on Colchic Database. *Ann Occup Hyg* 51: 131-142.

Koppisch D, Schinkel J, Gabriel S, Fransman W, Tielemans E. 2012. Use of the MEGA Exposure Database for the Validation of the Stoffenmanager Model. *Ann Occup Hyg* 56: 426-439.

Lavoué J, Gérin M, Vincent R. 2011. Comparison of Formaldehyde Exposure Levels in Two Multi-Industry Occupational Exposure Databases Using Multimodel Inference. *8*: 38-48.

Lavoué J, Vincent R, Gérin M. 2006. Statistical Modelling of Formaldehyde Occupational Exposure Levels in French Industries, 1986–2003. *Ann Occup Hyg* 50: 305-321.

Le ministre du travail, de la famille, de la solidarité et de la ville. 2009. Décret n° 2009-1570 du 15 décembre 2009 relatif au contrôle du risque chimique sur les lieux de travail. *Journal officiel de la république française*. p 21758.

Linch K, Miller W, Althouse R. 1998. Surveillance of Respirable Crystalline Silica Dust Using OSHA Compliance Data (1979-1995). *Am J Ind Med* 34: 547-558.

Pesch B, Kendzia B, Hauptmann K, Van Gelder R, Stamm R, Hahn J-U, Zschiesche W, Behrens T, Weiss T, Siemiatycki J, Lavoué J, Jöckel K-H, Brüning T. 2015. Airborne exposure to inhalable hexavalent chromium in welders and other occupations: Estimates from the German MEGA database. *Int J Hyg Env Health* 218: 500-506.

Peters S, Vermeulen R, Portengen L, Olsson A, Kendzia B, Vincent R, Savary B, Lavoué J, Cavallo D, Cattaneo A, Mirabelli D, Plato N, Fevotte J, Pesch B, Brüning T, Straif K, Kromhout H. 2011. Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies. *J Environ Monit* 13: 3262-3268.

Pôle Emploi. 1999. Répertoire Opérationnel des Métiers et des Emplois.

Rajan B, Alesbury R, Carton B, Gérin M, Litske H, Marquart H, Olsen E, Scheffers T, Stamm R, Woldbaek T. 1997. European Proposal for Core Information for the Storage and Exchange of Workplace Exposure Measurements on Chemical Agents. *Appl Occup Environ Hyg* 12: 31-39.

Sauvé J, Beaudry C, Bégin D, Dion C, Gérin M, Lavoué J. 2012. Silica Exposure During Construction Activities: Statistical Modeling of Task-Based Measurements from the Literature. *Ann Occup Hyg* 57: 432-443.

Scarselli A, Binazzi A, Di Marzio D, Marinaccio A, Iavicoli S. 2012. Hexavalent Chromium Compounds in the Workplace: Assessing the Extent and Magnitude of Occupational Exposure in Italy. *J Occup Environ Hyg* 9: 398-407.

Scarselli A, Montaruli C, Marinaccio A. 2007. The Italian Information System on Occupational Exposure to Carcinogens (SIREP): Structure, Contents and Future Perspectives. *Ann Occup Hyg* 51: 471-478.

Stamm R. 2001. MEGA-Database: One Million Data Since 1972. *Appl Occup Environ Hyg* 16: 159-163.

Stewart P, Rice C. 1990. A source of Exposure Data for Occupational epidemiology Studies. *Appl Occup Environ Hyg* 5: 359-363.

Vincent R, Jeandel B. 2001. COLCHIC - Occupational Exposure to Chemical Agents Database: Current Content and Development Perspectives. *Appl Occup Environ Hyg* 16: 115-121.

Vincent R, Jeandel B. 2002. Occupational exposure to lead. Information issued from the Colchic database. *187*: 63-72.

Vinzents P, Carton B, Fjeldstad P. 1994. Extraction of Core Information and Possibilities for Comparison Between European Databases for Occupational Air Pollution Measurements. *Appl Occup Environ Hyg* 10(4): 351-354.