Institut de recherche
Robert-Sauvé en santé
et en sécurité du travail

A Comparative Study of the IMIS (OSHA) and LIMS (IRSST) Occupational Exposure Databases

Philippe Sarazin
France Labrèche
Jacques Lesage
Jérôme Lavoué

STUDIES AND RESEARCH PROJECTS

R-1032-en

CHEMICAL AND BIOLOGICAL HAZARD PREVENTION RESEARCH

OUR RESEARCH
is working for you !

The Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST), established in Québec since 1980, is a scientific research organization well-known for the quality of its work and the expertise of its personnel.

Mission
To contribute, through research, to the prevention of industrial accidents and occupational diseases and to the rehabilitation of affected workers;

To disseminate knowledge and serve as a scientific reference centre and expert;

To provide the laboratory services and expertise required to support the public occupational health and safety network.

Funded by the Commission des normes, de l'équité, de la santé et de la sécurité du travail, the IRSST has a board of directors made up of an equal number of employer and worker representatives.

To find out more
Visit our Web site for complete up-to-date information about the IRSST. All our publications can be downloaded at no charge.
www.irsst.qc.ca

To obtain the latest information on the research carried out or funded by the IRSST, subscribe to our publications:

- *Prévention au travail,* the free magazine published jointly by the IRSST and the CNESST (preventionautravail.com)
- InfoIRSST, the Institute's electronic newsletter

Legal Deposit
Bibliothèque et Archives nationales du Québec
2021
ISBN : 978-2-89797-177-9

IRSST – Communications and Knowledge Transfer Division
505 De Maisonneuve Blvd. West
Montréal, Québec
H3A 3C2
Phone: 514 288-1551
publications@irsst.qc.ca
www.irsst.qc.ca

A Comparative Study of the IMIS (OSHA) and LIMS (IRSST) Occupational Exposure Databases

Philippe Sarazin[1], France Labrèche[1],
Jacques Lesage[2], Jérôme Lavoué[3]

[1] IRSST
[2] Université du Québec à Montréal
[3] Université de Montréal

STUDIES AND
RESEARCH PROJECTS

R-1032-en

This study was funded by the IRSST. The conclusions and recommendations are solely those of authors.
This publication is a translation of the French original; only the original version (R-1032) is authoritative.

PEER REVIEW
In compliance with IRSST policy, the research results published in this document have been peer-reviewed.

ACKNOWLEDGEMENTS

The researchers would like to thank Guylaine Beauchamp and Gilles Bensimon of the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) for giving us access to data in the LIMS and for clarifying aspects of its content. We would also like to express our thanks to Patrice Duguay of the IRSST for providing information about industry classification in Québec. Lastly, we are grateful to Dan Vatnik, programmer, Mawoulé Sant'Anna, researcher, and Solène Cadiou, intern, who helped to prepare the data from the IMIS and LIMS and develop the analyses used in the study.

SUMMARY

Good knowledge of levels of occupational exposure to chemical contaminants is crucial in any program to control, prevent and manage associated risks. Québec's occupational health and safety research institute, the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST), thus administers a *Laboratory Information Management System* (LIMS). This electronic database contains all analytical results on workplace exposure as measured by occupational health teams in Québec since 1985. Though the LIMS data are useful for developing exposure profiles, the scarcity of information on sampling circumstances and objectives limits their interpretation. In the USA, the Occupational Safety and Health Administration (OSHA) has maintained the *Integrated Management Information System* (IMIS) since 1979. This database, which contains analytical results from samples collected by OSHA inspectors since that time, includes more complete information on sampling circumstances than the LIMS. In addition, unlike the LIMS, which only reports sample concentrations, most IMIS results are calculated personal exposure measurements directly comparable to occupational exposure limits.

Given the similarity of U.S. and Québec socioeconomic activities, the general objective of this research project was to compare the LIMS data with the exposure data collected by OSHA to determine if the U.S. data can serve as a source of information on occupational exposure conditions in Québec.

Covered by the comparison were all chemical agents analyzed over a common period in the two databases. There were two main parts to the comparison. First, industries associated with exposure in the IMIS and LIMS were compared. Agent-industry pairs for which at least 10 detected values were available were identified in one database and the proportion of pairs found in the other database as well was determined. Second, exposure levels were compared. In a preliminary descriptive analysis, median levels of each chemical agent were compared, regardless of period or industry. In a second analysis, average levels reported in the IMIS and the LMIS were compared by statistical modeling taking into consideration measurement year, industry and sampling time. Results for all agents in two major chemical families, metals and solvents, were compared.

The extracts from the IMIS (352,442 records) and the LIMS (286,083 records) covered the years 1985 to 2011 and 49 common substances: 21 solvents, 15 metals, 5 gases, 4 isocyanates, 2 acids, crystalline silica and styrene. Metals data were more numerous in the IMIS (234,387 records compared to 86,054 records), whereas solvent data were more numerous in the LIMS (247,367 records compared to 71,690 records). Both databases included considerable data on lead, toluene, iron and manganese. For the 49 agents taken together, the proportion of non-detects and the proportion of values exceeding ACGIH threshold limit values (TLV®) were very similar in the two databases.

Most of the records in the extracts from the IMIS as well as the LIMS were for the manufacturing sector, with more than 70% of the measurements from the two industry groups designated as top priorities by Québec's workers compensation board, the Commission des normes, de l'équité, de la santé et de la sécurité du travail (CNESST).[1] In general, the breakdown of the measurements by industry was similar in the two databases. Comparison of agent-industry pairs for which there

[1] See Table 5 for the list of industries in the top priority groups.

were detected values was possible for 36 agents. Of the pairs identified in the IMIS, it was possible to match 61.4% with the LIMS, whereas 62.8% of the pairs identified in the LIMS could be matched with IMIS pairs. In other words, the IMIS and LIMS provide profiles of exposure within North American industry that are in general compatible but complementary.

Direct comparison of exposure levels was limited by the incomplete compatibility of the industrial classification systems used by the IMIS and LIMS and by the lack of information on sampling time in the LIMS prior to 1994. As a result, a descriptive comparison was made of 169,388 IMIS and 367,486 LIMS records, whereas for the modeling no more than 100,000 records in all were used from the two databases combined. The descriptive analyses and the modeling demonstrated that metal exposure levels, short-term as well as long-term, were lower in the IMIS than the LIMS by a factor of about two. This was true for most metals studied. For solvents, the descriptive analyses demonstrated similar short-term exposure levels in the two databases, despite substantial differences for certain agents. With respect to long-term exposure, the analyses suggest slightly higher levels in the IMIS. Substantial differences for certain solvents studied were also noted.

Despite the lack of reference data to check to what extent the IMIS and LIMS measurements are representative of occupational exposure in Québec, this study, which suggests a consistent overall portrait in terms of industries covered and exposure levels in the two databases, is reassuring in this respect. Given the scarcity of measurement data available, the results of this study provide strong support for using the IMIS and LIMS together to assess occupational exposure in Québec.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACGIH: American Conference of Governmental Industrial Hygienists

AOHS: Act respecting occupational health and safety

ART: Advanced REACH Tool

ATABAS: Danish occupational exposure database

CAEQ: *Classification des activités économiques du Québec* (Québec economic activity classification)

CANSIC: Canadian standard industrial classification

CAREX: Canadian carcinogenic exposure assessment project (CARcinogen EXposure)

CEHD: Chemical Exposure Health Data

CNESST: *Commission des normes, de l'équité, de la santé et de la sécurité du travail* (Québec's workers' compensation board)

COLCHIC: French occupational exposure database

CWED: Canadian Workplace Exposure Database

dEF: Difference in exceedance fraction

EXPO: Norwegian workplace exposure database

FIBREX: Database on exposure to organic or artificial inorganic fibres in France

GM: Geometric mean

IARC: International Agency for Research on Cancer

IMIS: Integrated Management Information System

INRS: *Institut national de recherche et de sécurité* (French National Research and Safety Institute for the Prevention of Occupational Accidents and Diseases)

IRSST: *Institut de recherche Robert-Sauvé en santé et en sécurité du travail* (Québec occupational health and safety research institute)

LIMS: Laboratory Information Management System

LOD: Limit of detection

MdEF: Median of differences in exceedance fractions

MEGA: German workplace exposure database

MrM: Median of ratios of medians

MSHA: Mine Safety and Health Administration

NAICS: North American Industry Classification System

ND: Non-detect

NEDB: National Exposure Database

NOES: National Occupational Exposure Survey

OEDB: Occupational exposure database

OEL: Occupational exposure limit

OR: Odds ratio

OSHA: Occupational Safety and Health Administration

PAH: Polycyclic aromatic hydrocarbon

PEL Permissible exposure limit

rM: Ratio of medians

RSPSAT: *Réseau de santé publique en santé au travail* (Québec's public occupational-health network)

SIC: Standard Industrial Classification

SIREP: Italian information system for recording occupational exposure to carcinogens

SOLVEX: French solvent exposure database

STEL: Short-term exposure limit

TLV®: Threshold Limit Value set by the American Conference of Governmental Industrial Hygienists (ACGIH)

TWA: Time-weighted average

TWAEV: Time-weighted average exposure value

USSIC: U.S. standard industrial classification

1. INTRODUCTION

1.1 Knowing exposure can improve occupational health

The assessment of occupational exposure to chemical agents plays a key role in any occupational health risk management and control program. The availability of data on the intensity, duration and frequency of exposure to contaminants makes it possible to develop effective exposure monitoring programs and to support the implementation of prevention policies geared to specific industries or occupations. Such data can also be used for epidemiological studies and for the development of models to predict occupational exposure.

Traditional epidemiological approaches are based on semiquantitative expert assessments that cannot be used to plot the exposure-response curves required for risk management (Teschke et al., 2002). The major investment that industrial hygiene measurements require has fostered the development of approaches that make it possible to reduce the sampling effort required to assess a situation, provided information is available from other sources (Hewett, Logan, Mulhausen, Ramachandran and Banerjee, 2006; Sottas et al., 2009; Tielemans et al., 2011). Lastly, setting occupational health intervention priorities depends on estimations of populations exposed, occupational disease burden and associated costs, but reliable results cannot be expected without sources of valid exposure data for the entire population (Kauppinen et al., 2000; Labrèche et al., 2012; Van Tongeren et al., 2012).

1.2 Occupational exposure databases (OEDB)

National occupational exposure databases, repositories for hygiene measurements made by government agencies as part of their prevention and control activities, would seem an important potential source of populational information. Most of these databases were started in the 1980s, but it wasn't until the early 2000s that epidemiological and public health studies reporting their use started to multiply. Internationally, use was made of databases from France (Clerc, Bertrand and Vincent, 2014; Kauffer and Vincent, 2007; Mater, Paris and Lavoué, 2016), Germany (Koppisch, Schinkel, Gabriel, Fransman and Tielemans, 2012; Pesch et al., 2015), Italy (Scarselli, Binazzi and Di Marzio, 2011; Scarselli, Binazzi, Marzio, Marinaccio and Iavicoli, 2012; Scarselli, Corfiati and Marzio, 2016; Scarselli, Di Marzio, Marinaccio and Iavicoli, 2013; Scarselli, Montaruli and Marinaccio, 2007) and the United States (Cowan et al., 2015; Henn et al., 2011; Sarazin, Burstyn, Kincl and Lavoué, 2016). The French database in particular was used to create a number of prediction tools for exposure to volatile organic compounds, solvents and asbestos (Clerc et al., 2014; INRS, 2015a, 2015b). The SYNERGY project was the very first international multicentre case-control study that quantitatively assessed exposure based on pooled data for five carcinogens from large OEDBs (Olsson et al., 2011; Peters et al., 2012). In Canada, the Canadian Workplace Exposure Database (CWED) comprises data mainly from Ontario and British Columbia. Started in 2008, it was the data source for the CAREX Canada project, which was in turn used to estimate the number of workers exposed to carcinogens in Québec (Hall, Peters, Demers and Davies, 2014; Labrèche et al., 2012).

1.3 Québec's LIMS (IRSST)

In Québec, the *Laboratory Information Management System* (LIMS) is managed by the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST), Québec's occupational health and safety research institute. An electronic database, the LIMS contains all analytical results for samples collected since 1985 by teams of the public occupational-health network (industrial hygienists and technicians of the regional health and social services agencies and the integrated health and social services centres) and by other members of Québec's prevention network (inspectors of the workmen's compensation board, that is, the CNESST, the Commission des normes, de l'équité, de la santé et de la sécurité du travail, and officials of joint sector-based health and safety associations). Research conducted by Lavoué et al. (2012) and funded by the IRSST made it possible to document in the LIMS a total of 557,000 records corresponding to laboratory analyses of samples of workplace air performed between 1985 and 2008.

Though there is little information on the circumstances of each sampling, the LIMS data are nonetheless used to develop exposure profiles because they are the only historical computerized data currently available in Québec. Several partial summaries of the content of the database were compiled by IRSST researchers to study exposure to particular chemical agents (Ostiguy, Cordeiro, Bensimon and Baril, 2011; Ostiguy, Morin, Bensimon and Baril, 2012). However, as discussed in these different summaries, interpretation of the data is greatly limited by the lack of contextual information on the circumstances associated with the analytical results—which must accordingly be interpreted with great care. For example, we do not know the type of sample (personal, ambient air or exposure source), the sampling strategy (task evaluation, calculation of weighted exposure) or the occupation (job title) or task assessed. It is also not possible to identify the samples used to calculate a full-shift time-weighted average exposure (two successive samples, for example, one taken in the morning and the other in the afternoon).

1.4 OSHA's IMIS

The Integrated Management Information System (IMIS), a database maintained by the U.S. Occupational Safety and Health Administration (OSHA), is the only other data source in North America of size comparable to Québec's LIMS. Accessible to the public thanks to the U.S. Freedom of Information Act, the IMIS was created in 1979 and today contains over 1.5 million records of samples collected by OSHA inspectors to verify workplace compliance with permissible exposure limits (PEL). These measurements are accompanied by more complete contextual information than in Québec's LIMS, including information about the company where the inspection was conducted (industry), sampling date, reason for inspection, type of exposure and job title. The exposure results derive from sampling procedures using standardized analytical techniques with rigorous control and quality assurance protocols. In addition, most of the exposure measurements were taken in the worker's breathing zone and are thus more representative of the exposure to which the worker was subject than measurements taken at fixed sampling stations (that is, in the ambient air). The IMIS data cover all OSHA-regulated contaminants, that is, over a thousand different chemical agents (Lavoué, Friesen and Burstyn, 2013). Recent uses of the IMIS data include analyses of exposure to lead (Henn et al., 2011; Okun, Cooper, Bailer, Bena and Stayner, 2004), crystalline silica (Linch, Miller, Althouse, Groce and Hale, 1998; Yassin, Yebesi and Tingle, 2005), formaldehyde (Lavoué, Vincent and Gerin, 2008; Melville and Lippmann, 2001) and beryllium (Hamm and Burstyn, 2011). These studies made it possible to get a broad picture of occupational exposure as well as to estimate the number of workers exposed in certain economic sectors and to show how exposure to these contaminants has changed over time. In addition, a recent publication describes an overall analysis of IMIS data across more than 77 chemical agents designed to identify possible bias in exposure level findings (Sarazin, Burstyn, Kincl, Friesen and Lavoué, 2018; Sarazin et al., 2016). Considering the wealth of contextual information accompanying the exposure results, and the fact that most of the results are directly interpretable in terms of exposure limit values, the IMIS offers excellent potential for the development of applications to prevent occupational disease in North America.

In sum, quantitative data stored in OEDBs are an important source of information on workplace exposure to chemical agents. In Québec, the findings of laboratory analyses recorded in the LIMS have been used to identify research needs and to set prevention priorities. However, the reliability of exposure assessments based on LIMS data has been repeatedly called into question.

2. RESEARCH OBJECTIVES

The general objective of this research project was to compare exposure data collected by OSHA with LIMS data to determine the latter's usefulness as a source of information on occupational exposure conditions in Québec so industries where such exposure is likely can be targeted and we can determine what we need to do to advance our knowledge.

Specific objectives were as follows:

1) To determine to what extent chemical agents were found present in the same industries in both the IMIS (OSHA) and the LIMS (IRSST)

2) To assess the equivalence of reported concentrations in the IMIS and the LIMS for a set of chemical agents in industries common to the two databases

3. METHODOLOGY

The content and type of information in the LIMS and IMIS are first briefly described, followed by a presentation of the approaches used to reach the specific research objectives. These include 1) preparation of extracts from the IMIS and LIMS; 2) comparison of industries associated with measured exposures; 3) comparison of reported concentrations for periods and industries common to the two databases.

3.1 The databases

3.1.1 The Québec LIMS

Chemical agents, industries, time periods and contextual information available in the LIMS were identified thanks to earlier research conducted by Lavoué et al. (2012). In sum, the LIMS extract made it possible to document a total of 557,000 test results for air samples collected in 13,370 establishments between 1985 and 2008, excluding process samples, controls, surface contamination analyses and second tube sections. Of the 173 chemical agents in the LIMS extract described by Lavoué et al. (2012), 63 were associated with more than 1,000 exposure measurements collected in breathing zones or at fixed stations and they accounted for 93% of the total number of records. These chemical agents were mainly classed as metals, solvents, gases, dusts, isocyanates and polycyclic aromatic hydrocarbons (PAHs).

The data extract from the LIMS used for the present research project covers the period from 1985 to 2014. It includes 593,002 test results for 57 substances sampled in 13,132 establishments.

3.1.2 The U.S. IMIS

An electronic extract from the IMIS covering the period from 1979 to 2012 was obtained from OSHA by Jérôme Lavoué's research team thanks to the U.S. Freedom of Information Act. In sum, the IMIS extract contains 851,987 records of 132,280 workplace inspections conducted between 1979 and 2012, the data covering 1,050 industrial activity codes of the 1987 U.S. *Standard Industrial Classification* (SIC) system. Given that the IMIS data derives from inspections to check compliance with standards, most of the data comprise 8-hour time-weighted averages or short-term averages for the OSHA-specified reference period. Of the 1,169 different chemical agents in the database, 65 were the subject of more than 1,000 exposure measurements collected in breathing zones or at fixed sampling stations and constituting 91% of the total number of IMIS records. These chemical agents were from the same classes as those present in the LIMS, that is, metals, solvents, gases, dusts, isocyanates and PAHs. Lastly, just over 65% of the data are from inspections conducted in response to an employee complaint or a request from an inspector, the remainder being from planned inspections.

In addition to numerical values for exposure levels, the IMIS records include contextual information about the company and the inspection: identification of chemical agent sampled, sampling date, company name, job title, industry (four-figure SIC code), sample type (breathing zone or fixed station), type of or reason for inspection (not programmed, that is, follow-up, complaint or compliance officer referral; or programmed, that is, scheduled or monitoring), type of exposure (in the absence of exact sampling times, results are categorized as time-weighted average exposure values [TWAEV], short-term exposure values [STEV] or non detected results [ND]), union status, scope of inspection (full or partial inspection of establishment) and location (city where inspection took place).

3.2 Preparation of IMIS and LIMS extracts

Selection of chemical agents

Chemical agents were selected from among those that were most frequent (>500 measurements) and appear in both databases. In addition, it was necessary to develop a table of concordance between the fields of information describing the chemical agents in the two databases. In fact, for some compounds, such as metals, dusts or compounds with isomers, there were differences in speciation or physicochemical characterization (chromium compounds vs chromium VI, for example), while in other cases compounds had to be grouped to make the data in the two databases comparable (copper fume and copper dust were grouped together in the IMIS, for example).

Assignment of industry codes

To be able to carry out certain comparative analyses between the IMIS and LIMS, the industry classification systems used by the two databases had to be standardized. Every record in the LIMS is associated with an economic activity code assigned to the establishment inspected by the CNESST based on Québec's 1984 industrial classification, the Classification des activités économiques du Québec (CAEQ) (Bureau de la statistique du Québec, 1984). The IMIS, on the other hand, uses the 1987 U.S. Standard Industrial Classification (SIC) (OSHA, 2014). Note that starting in 1998, the code of the 1997 version of the North American Industry Classification System (NAICS) is also indicated in the IMIS (Statistique Canada, 2002). Lastly, for another IRSST study, NAICS codes were assigned to some LIMS establishments between 2000 and 2014 using statistics from the CNESST's Dépôt de données central et régional (DDCR) and LIMS files, processed by the IRSST (2014; 2015).

For the comparative analyses in this report requiring consideration of industry, the IMIS and LIMS were linked in three ways:

- A U.S. SIC code (1987 version, herein the USSIC) was assigned to the LIMS data. This was a two-stage process: first the CAEQ code was switched to a Canadian SIC code (1980 version, herein CANSIC); then a CANSIC/USSIC concordance table, obtained from Statistics Canada was applied.

- A CANSIC code was assigned to the IMIS data. This was done using the CANSIC/USSIC concordance table,

- NAICS codes available in both databases were used.

These three types of links were created because no single approach offered a perfect matchup of the data in the two databases. In the first case, as the CANSIC/USSIC concordance table did not always yield just one concordance (a code might correspond to several different codes in the target system), a USSIC code could not be found for a significant portion of the LIMS data. In the second case, a CANSIC code could not be found for a significant portion of the IMIS data. Lastly, though both databases use NAICS codes, these codes were only assigned in a subgroup of establishments. As it is plausible that different subsets of the databases would be eliminated if records without codes were removed, use of the three approaches is a form of sensitivity analysis.

Measurement reference periods

Sampling times are indicated differently in the two databases. In the IMIS, measurements can be subdivided into two subgroups according to reference period: a short-term result or a shift-long time-weighted average result. The IMIS-LIMS comparisons of concentrations were thus stratified in accordance with this dichotomy. In the IMIS, reference periods are categorized according to the relevant limit values: there are 8-hour time weighted averages (TWA), short-term exposure limits (STEL) and ceilings or peak limits. In a number of recent IMIS analyses, STELs, peaks and ceilings were grouped under the label “short-term” results (Lavoué et al., 2013; Lavoué et al., 2008; Sarazin et al., 2016), an approach used in the present research as well. In the LIMS, the exact sampling time is available for every record. These sampling times were used by selecting a 30-minute cut point to a priori separate short-term and long-term data. A number of sensitivity analyses were, however, performed, to validate this criterion.

3.3 Comparison of lists of industries with measured exposures in the IMIS and LIMS

The purpose of this comparison was to determine the agreement between the IMIS and LIMS for the period from 1985 to 2011 on industries where particular chemical agents are present, regardless of the level of exposure. The goal was to find out to what extent the IMIS and LIMS offer similar responses to the following question: What substances are found in measurable quantities in particular industries.

The unit of comparison for this analysis is an agent-industry pair. The principle of comparison is as follows: identify agent-industry pairs with at least 10 detected values in one database and check if these pairs can also be found in the other database. This is a complex operation because of the problem of linking the IMIS and the LIMS via industry codes. For example, CANSIC codes can be linked to one and only one USSIC code in the case of only 41% of the 548 different CANSIC codes in the LIMS (28% of LIMS records). Conversely, USSIC codes can be linked to a single CANSIC code in the case of only 54% of the 868 different USSIC codes in the IMIS (53% of IMIS records).

To deal with this issue, the CANSIC-coded LIMS was used as the point of departure. For each agent-CANSIC code pair identified in the LIMS (≥ 10 detected values), the CANSIC-USSIC concordance table was used to obtain one or more agent-USSIC code pairs (depending on the number of USSIC target codes). If one or more agent-USSIC pairs were also identified (≥ 10 detected values) in the IMIS, the initial LIMS pair was considered to have an IMIS link. This procedure made it possible to get around the problem of the concordance table for the two classification systems showing more than one link. Next, this procedure was applied in the other direction, using the USSIC-CANSIC concordance table to find, in the LIMS, pairs identified in the IMIS.

For NAICS classifications, available for 22% and 37% of IMIS and LIMS records respectively, the comparison was performed simply by identifying the proportion of agent-industry pairs shared by the two databases or exclusive to one.

Retained for our analysis were only those substances in each database for which there were at least 10 detected results and data across at least six industry codes.

3.4 Comparison of reported concentrations in industries common to both databases

The purpose of the analysis was to determine the match between levels reported in the IMIS and LIMS during the reference period for chemical agents and industries found in both databases. The analysis had two parts: comparison of average exposure levels in each database; and calculation of the proportion of measurements above a given limit (exceedance fraction, EF). This latter technique, used recently with IMIS data (Hamm and Burstyn, 2011; Lee, Lavoué, Spinelli and Burstyn, 2015), made it possible to avoid the methodological challenges of estimating distribution parameters when the data includes non-detects (NDs), that is, values below detection limits (Helsel, 2005). In fact, 41% and 40% respectively of the measurements recorded in the IMIS and LIMS are below detection limits. The limit selected for exceedance fraction calculation was based on the threshold limit values (TLV®) recommended by the ACGIH (ACGIH, 2017).

Comparing average exposure levels required considerable preprocessing of the data in both databases, because neither provides the detection limit for results reported as NDs. Two approaches were thus used to estimate limits of detection (LOD) for each substance: an empirical approach involving examination of quantile-quantile (QQ) plots; and a bibliographic approach involving consultation of historical reference works used for analytical methods in the United States (for the IMIS) and in Québec (for the LIMS). A detailed description of these two approaches as well as a comparison of the results obtained is presented in a report available in an online appendix.[2] Lastly, LODs were determined for the IMIS and LIMS using the bibliographic approach. Our conclusions about the LODs derived from these two methods are summarized below:

- Historical (bibliographic) LODs obtained from OSHA manuals in the US and IRSST manuals in Québec are similar.
- Historical LODs tend to be higher than empirical LODs by a factor of 3 for the LIMS and a factor of 10 to 100 for the IMIS.
- The empirical LODs of the IMIS are decidedly lower than the empirical LODs of the LIMS.
- Historical LODs decrease with time, specific to each class of chemical agent.

Furthermore, the IMIS posed an additional challenge: as the variable identifying a result as an ND value is the same as that identifying sampling time, it is not possible to determine sampling time in the case of ND results. Sampling time, however, together with sampling volume, is a key LOD determinant. A decision tree approach was accordingly used to allow prediction of sampling times in the IMIS (short- or long-term data) based on other variables in the database. This approach is

[2] http://expostats.ca/jlavoue/LOQv2.html.

described in detailed in Philippe Sarazin's doctoral dissertation (Sarazin, 2016). This made it possible to match each IMIS and LIMS record reported as an ND value with an LOD.

Lastly, the analyses described below were restricted to the two main classes of chemical agents (metals and solvents) that contain the largest number of agents common to the two databases (36 out of 49) and account for 80% and 82% of the IMIS and LIMS data respectively. The results are also systematically separated for solvents and metals, as notable differences were found between these two classes of agents in the preliminary descriptive studies.

3.4.1 Descriptive comparison

Data preparation

Separate descriptive comparison analyses were performed for the short-term and long-term data. Data sets were limited to the period between 1994 and 2011 for both databases because sampling times are not given in the LIMS for measurements recorded before 1994. Measurements reported as NDs were replaced by the value of half of the LOD. Analyses were performed for chemical agents for which there were more than 100 measurements in each of the databases.

A) Ratio of medians

This approach consisted in calculating, for each selected chemical agent, the ratio of the median of exposure concentrations in the IMIS to the median of exposure concentrations in the LIMS, yielding a ratio of medians (rM) for each agent. An indicator, called the median of the ratios of medians (MrM), was then used to get a picture of the value of this ratio across agents:

$$rM_i = \frac{med(\text{IMIS values for agent } i)}{med(\text{LIMS values for agent } i)} \quad (1)$$

$$MrM = med(rM) \quad (2)$$

Thus, if exposure concentrations are on the whole higher in the IMIS than in the LIMS, the MrM will be greater than 1. If they are on the whole lower in the IMIS than in the LIMS, the MrM will be lower than 1.

B) Difference in TLV® exceedance fraction (dEF)

This approach consisted in calculating, for each selected chemical agent, the difference between the TLV® exceedance fraction (EF) in the IMIS and the TLV® EF in the LIMS, yielding a difference in exceedance fraction (dEF) for each agent. An indicator, called the median of the differences in exceedance fractions (MdEF) was then used to get a picture of the value of this difference across agents:

$$dEF_i = (\%\ of\ values > TLV\ in\ the\ IMIS\ for\ agent\ i) - (\%\ of\ values > TLV\ in\ the\ LIMS\ for\ agent\ i) \quad (3)$$

$$MdEF = med(dEF) \quad (4)$$

Hence, if TLV® exceedance fractions are on the whole higher in the IMIS than in the LIMS, the MdEF will be greater than 0; if they are higher in the LIMS than in the IMIS, then the MdEF will be lower than 0.

Stratification of analyses

The indicators MrM and MdEF were calculated according to different strata of the following variables, taking sampling time (long- or short-term data) into account:

- The IMIS and LIMS datasets for all periods
- Data associated with industries common to both databases, according to the industry classification system (CANSIC, USSIC, NAICS)
- Period (1994-2002, 2003-2011)
- Cut point selected to distinguish short-term from long-term data in the LIMS (60 minutes and 240 minutes were tested)

For each strata, analyses were performed for chemical agents for which there were more than 100 measurements in each of the databases.

3.4.2 Comparison by statistical modelling

Statistical modelling, which makes it possible to study the relation between chemical contaminant exposure levels and the factors associated with these exposure levels, is a powerful tool for interpreting measurements in databases such as the IMIS and LIMS. In industrial hygiene, multivariate statistical models are now the standard approach for this type of analysis (Friesen et al., 2012; Lavoué et al., 2013; Lee et al., 2015; Sarazin et al., 2016).

The objective of the analysis was to identify systematic differences between the IMIS and LIMS, taking into account sampling time, industry and trends across time. There was one particular challenge: finding an approach that would make it possible to combine the results for all chemical agents and get an overall picture of the differences in exposure levels recorded in the two databases. Such an approach was developed by Philippe Sarazin in his doctoral research (Sarazin, 2016; Sarazin et al., 2016) and is of particular interest when studying large multi-industry sources of data such as occupational exposure databases (OEDB). This approach (described in detail below) was used for this research: first, a model was fitted individually to each chemical; secondly, a meta-analysis was used to get an overall picture of the differences between the IMIS and LIMS across all the chemical agents.

For the comparative analysis by statistical modelling, the dataset was limited to long-term measurements in industries common to both the IMIS and LIMS for the period 1994 to 2011, because there were too few short-term measurements to obtain robust results. In fact, short-term data account for only 13% and 16% of the records in the IMIS and LIMS respectively. Also, the cut point for long-term data had to be lowered (from 240 to 120 min) for the comparison by modeling, because a cut point of 240 min considerably limited the sample size. Data from the two databases were pooled and a variable identifying the source of the data was created. The analysis was performed using two modeling approaches: Tobit models and logistic regression models.

Tobit models and logistic regression models

With Tobit modelling, ND values can be taken into account without replacing them arbitrarily by a single value (Lavoué et al., 2013; Lavoué, Gerin and Vincent, 2011; Lubin et al., 2004; Persson and Rootzen, 1977). One chemical agent selection criterion was that NDs could not constitute more that 60% of the results. A logarithmic transformation of the measurements was performed before the modeling to consider the lognormal distribution of the occupational exposure measurements (AQHSST, 2004).

With logistic regression modeling, binary response variable values can be explained from a combination of continuous or binary explanatory variables (Hosmer and Lemeshow, 2000). In the present study, this approach was used to estimate the probability of exposure exceeding or not exceeding a predetermined threshold (ACGIH TLV®). This approach made it possible to include all results reported as below the detection limit (NDs) without requiring an LOD value by coding them as below the TLV.®

For each chemical agent, the variables common to both databases were included in the Tobit and logistic regression analyses: data source, industry and sampling year. The analyses focused in particular on estimating the influence of the data source variable, indicator of an overall systematic difference between the IMIS and LIMS independent of the other variables of the model. Any interactions noted between data source and the other variables would rather be indicative of different effects for the latter in the IMIS and LIMS (e.g., differences between the databases in trends across time). The data were thus modelled for each agent including the three common variables as well as the interaction between source and sampling year. The interaction between source and industry was not included in the models because the number of measurements was not sufficient for all source-industry pairs.

Grouping of economic activity variables

Type of industry was included in our analyses by modelling to control for potential bias in estimating the effects of other model variables. The variables describing the industry (CANSIC, USSIC, NAICS) include hundreds of categories organized in a hierarchical structure. CANSIC, for example, has a 4-digit coding system, with detailed industries at the 4-digit level, industry groups at the 3- and 2-digit levels and divisions at the highest or 1-digit level. To ensure a sufficient number of measurements per industry for any given chemical agent, industry categories with few measurements were aggregated as follows: in case of <30 measurements in a 4-digit industry category, the most specific digit was dropped, reclassifying the measurements in the broader 3-digit category. This process was repeated until there were ≥30 measurements in the category, or the code was reduced to one digit. Finally, if a 1-digit code was not associated with >30 measurements, the data were put into an “other” category. This approach has been used to manage this type of variable by other researchers (Lavoué et al., 2011; Lee et al., 2015; Sarazin et al., 2016).

IMIS/LIMS ratio of geometric means and odds ratio (OR) of exceeding the TLV®

To determine an average difference between the exposure levels recorded in the IMIS and those recorded in the LIMS, the following were calculated for each chemical agent: 1) using the Tobit models, the predicted IMIS/LIMS ratio of geometric means (GM); 2) using the logistic models, the IMIS/LIMS odds ratio (OR) of exceeding the TLV®, that is, the odds of exposure being above the TLV® in the IMIS divided by the odds of its being above the TLV® in the LIMS (odds being the quotient of the probability of being above the TLV® and the probability of not being above it).

IMIS/LIMS ratios were calculated for scenarios of the following variables:

- Industry classification system (CANSIC, USSIC or NAICS)
- Year (1997: start of common period; 2008: end of common period)
- Cut point selected to separate short-term from long-term data in the LIMS (30 minutes and 120 minutes were tested)

For each scenario studied, analyses were performed for all chemical agents with more than 100 long-term measurements in both databases.

Meta-analysis for summary of results across agents

Exposure levels recorded in the IMIS were compared with those in the LIMS for each chemical agent separately using the Tobit and logistic regression models. For each chemical, an IMIS/LIMS ratio was thus obtained for each scenario studied (an IMIS/LIMS ratio was obtained for the following scenario, for example: classification system = CANSIC, year = 1997 and cut point for long-term data in the LIMS = 30 minutes).

Meta-analysis is a statistical method that makes it possible to combine the results of many independent studies into one common result (Borenstein, Hedges, Higgins and Rothstein, 2010; Sarazin et al., 2016). For our analysis, each chemical agent corresponds to a “study.” A random-effects model meta-analysis with DerSimonian-Laird estimator (Borenstein et al., 2010) was thus used to pool the results of all the chemical agents and obtain an overall picture of the IMIS/LIMS ratio for the scenario studied. The estimated overall IMIS/LIMS ratio is thus the weighted mean of the individual estimates across the chemical agents. The weighting depends on sample size, standard deviation of the agent ratio and the difference between the specific ratio of the agent and the overall average ratio.

Software

The analyses were performed with the R 3.1.3 statistical software (R Development Core Team, Vienna, Austria) using the package *survival* for the Tobit models (Therneau and Lumley, 2016) and the package *metafor* for the meta-analysis (Viechtbauer, 2014).

4. RESULTS

4.1 Detailed descriptive analysis of IMIS and LIMS content

4.1.1 IMIS and LIMS extracts

The extract from the US IMIS used for this study was obtained and preprocessed for earlier research and is described in Section 3 on methodology.

The extract from the LIMS obtained from the IRSST covers the period from 1985 to 2014. It contains 593,002 analytical results for 57 substances collected from 13,132 establishments. Uninterpretable results (n = 4,113) were deleted, as were 78,668 records without an establishment number. Existing CAEQ codes were transformed to CANSIC codes using a concordance table provided by Statistics Canada. The two systems are very similar, but judgment was required in the case of certain links that offer more than one possibility (about 20,000 records). It was also possible to assign CANSIC codes for about 10,000 records that initially were without a CAEQ code by using the code assigned to other records from the same establishment. At the end of the CANSIC code assignment procedure, 4-digit codes had been assigned to 440,238 records and 3-digit codes to another 42,667. There was an NAICS code for 194,296 records.

4.1.2 Shared content

A total of 49 chemical agents were linked in the two databases and associated with a large number of measurements (see Appendix A for a complete list of the agents). These substances were the subject of 423,630 records in the IMIS and 435,930 in the LIMS. They were classified into seven classes: metals (n=15 chemical agents), solvents (n=21), acids (n=2), gases (n=5), isocyanates (n=4), crystalline silica and styrene. Lastly, 386,083 records in the LIMS and 352,442 in the IMIS belonged to the common period 1985 to 2011.

4.1.3 Description

Figure 1 gives a breakdown of the data in the IMIS and LIMS by year. Though on average the two databases have a similar number of annual measurements, the measurements in the IMIS are concentrated before 1995, with a marked decrease in more recent years. For the LIMS, on the other hand, the numbers are relatively stable during the period studied. These trends were similar when results where stratified by class of chemical agent.

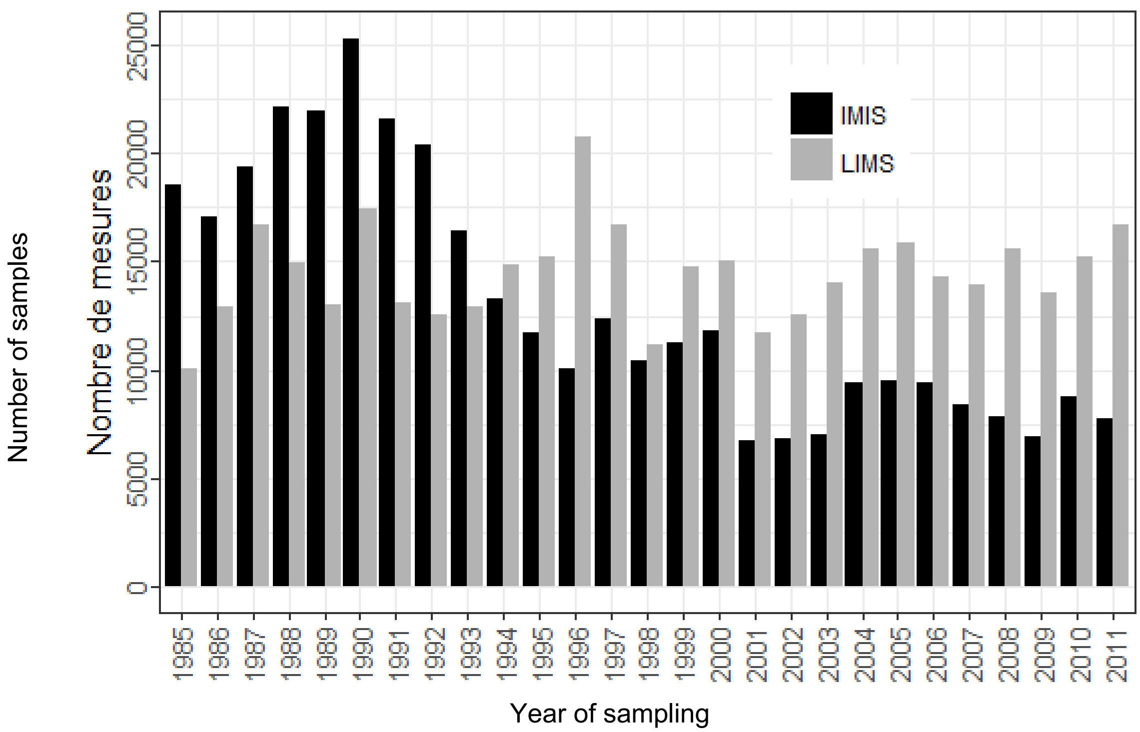

Figure 1. Number of samples per year in the IMIS and LIMS.

Table 1 summarizes the content of the two databases by class of chemical agent. The contrast between solvents and metals is notable: whereas the number of records for the other classes are of the same order of magnitude, there is a clear difference between the IMIS and LIMS for these classes, with three times as many solvent measurements in the LIMS as in the IMIS, and, conversely, three times more metal measurements in the IMIS than in the LIMS.

Table 2 lists the 10 most frequent agents in the two databases. As the table shows, metals predominate in the IMIS and solvents in the LIMS. Lead, toluene, iron and manganese are frequent in both the IMIS and LIMS.

For each of the 49 agents the comparison covers, Table 3 shows the number of samples, the percentage of results reported as non-detects and the ACGIH TLV® exceedance fraction for the reference period (short-term or long-term).

The median percent non-detects was 32% (from 7.9% for trichloroethylene to 91% for antimony) for the IMIS but 48% for the LIMS (from 14% for styrene to 88% for antimony). For both databases, the percent non-detects was systematically higher for metals, gases and acids (medians ranging from 51% to 78% for the LIMS and from 33% to 62% for the IMIS) than for solvents, crystalline silica and styrene (medians ranging from 14% to 44% for the LIMS and from 10% to 23% for the IMIS). The percent non-detects strongly correlated for all agents in both databases, with a Spearman's coefficient of 0.74.

Table 1. Classes of chemical agents in the IMIS and LIMS

Class of agent	Chemical agents	Number of measurements LIMS	Number of measurements IMIS
Solvent	Stoddard solvent, methanol, ethanol, isopropanol, hexane, hexone, benzene, xylene, toluene, methyl chloroform, ethyl acetate, n-butyl acetate, acetone, 2-butanone, tetrachloroethylene, methylene chloride, 2-butoxyethanol, n-butyl alcohol, trichloroethylene, ethylbenzene, phenol	247,367	71,690
Metal	Lead, silver, cadmium, cobalt, nickel, copper, iron , manganese, chromium, zinc, tin, antimony, molybdenum, beryllium, vanadium	86,054	234,387
Crystalline silica	-	10,302	18,838
Gas	Formaldehyde, ammonia, vinyl chloride, ethylene oxide, chlorine	10,838	10,719
Isocyanate	HDI, 4,4'-MDI, 2,6'-TDI, 2,4'-TDI	12,601	7,152
Acid	Sulphuric acid, nitric acid	1,458	2,112
Styrene	-	17,463	7,544

Table 2. Ten most common chemical agents in the IMIS and LIMS

IMIS		LIMS	
Agent	n	Agent	N
Lead	45,530	Toluene	46,799
Iron	24,329	Xylene	34,662
Copper	21,255	Acetone	29,725
Manganese	20,595	2-butanone	25,307
Zinc	20,493	Styrene	17,463
Crystalline silica	18,838	Manganese	14,771
Toluene	17,566	Iron	14,540
Chromium	16,598	Methyl isobutyl ketone	13,995
Nickel	15,283	Isopropanol	13,854
Cobalt	12,787	Lead	12,629

Table 3. Number of records, percent non-detects and ACGIH TLV® exceedance fractions for chemical agents in the IMIS and LIMS

		n	n	ND (%)[a]		TLV® **exceedance fraction (%)**[b]	
Class	**Name**	**IMIS**	**LIMS**	**IMIS**	**LIMS**	**IMIS**	**LIMS**
Solvent	Toluene	17,566	46,799	13	21	22	21
Solvent	Xylene	11,336	34,662	22	42	1.4	1.6
Solvent	Acetone	4,961	29,725	19	34	1.9	0.93
Solvent	2-butanone	4,848	25,307	23	50	3.2	1.1
Solvent	Methyl isobutyl ketone	2,570	13,995	25	49	7.8	4.1
Solvent	Isopropanol	3,292	13,854	20	42	3.7	2.8
Solvent	Stoddard solvent	3,092	9,802	30	44	3.2	3
Solvent	Ethyl acetate	479	9,747	19	59	0.21	0.13
Solvent	Butyl acetate	3,434	8,698	21	44	0.29	0.11
Solvent	Methylene chloride	4,328	8,491	14	31	28	23
Solvent	Ethanol	595	7,756	32	54	0.67	0.46
Solvent	Butanol	608	7,031	36	61	7.1	2.8
Solvent	Hexane	1,463	6,063	23	36	7.8	7.7
Solvent	Methanol	606	5,507	30	42	8.3	12
Solvent	Ethylbenzene	3,001	5,127	28	68	3.4	1.3
Solvent	Benzene	2,106	4,935	61	79	12	8.1
Solvent	Tetrachloroethylene	2,146	2,794	11	44	32	12
Solvent	1,1,1 trichloroethane	2,127	2,759	16	43	4.6	2.2
Solvent	Trichloroethylene	1,082	2,530	7.9	31	61	40
Solvent	2-butoxyethanol	1,289	1,179	36	45	0.62	4.8
Solvent	Phenol	761	606	42	58	0.26	0.83
Metal	Manganese	20,595	14,771	20	19	52	70
Metal	Iron	24,329	14,540	9.4	15	8.3	9.5
Metal	Lead	45,530	12,629	49	48	22	33
Metal	Chromium	16,598	10,146	49	64	1.7	4.4
Metal	Nickel	15,283	9,122	63	65	3.1	8.2
Metal	Copper	21,255	7,805	24	42	3.7	2.6
Metal	Zinc	20,493	5,813	22	17	3.8	7.4

		n	n	ND (%)[a]		TLV® **exceedance fraction (%)**[b]	
Metal	Beryllium	12,508	4,279	85	37	11	28
Metal	Cobalt	12,787	3,128	80	51	5.3	23
Metal	Cadmium	10,777	2,902	69	77	7	14
Metal	Tin	349	277	62	86	0.57	1.8
Metal	Antimony	11,549	238	91	88	0.56	5.9
Metal	Silver	751	196	35	40	34	20
Metal	Vanadium	10,641	133	87	61	1.9	5.3
Metal	Molybdenum	10,942	75	85	76	0.082	0
Gas	Formaldehyde	8,224	8,780	26	33	19	26
Gas	Ammonia	984	1,288	28	33	9.7	12
Gas	Vinyl chloride	370	360	71	83	6.2	0.83
Gas	Ethylene oxide	748	301	33	67	14	24
Gas	Chlorine	393	109	48	54	13	0.92
Isocyanates	HDI	1,410	6,284	57	46	12	20
Isocyanates	4,4'-MDI	3,649	3,677	59	80	19	5.1
Isocyanates	2,4'-TDI	1,420	1,342	62	61	8.6	7.6
Isocyanates	2,6'-TDI	673	1,298	77	54	6.1	11
Acid/Base	Sulphuric acid	1,329	915	56	76	11	16
Acid/Base	Nitric acid	783	543	58	79	2	5.2
Crystalline silica	-	18,838	10,302	24	26	71	64
Styrene	-	7,544	17,463	9.5	14	59	56

a. Percentage of values reported as non-detects (ND, below LOD)

b. Percentage of values above ACGIH threshold limit value (TLV®). The calculations in the table are designed to avoid the pitfalls of processing ND results when comparing exposure levels in the two databases. They indicate the number of results that exceed a target threshold but they cannot be directly linked to occupational health risk (see methodological considerations under Discussion below).

The median ACGIH TLV® exceedance fraction was 7.0% (from 0.1% for molybdenum to 71% for crystalline silica) for the IMIS and 7.4% for the LIMS (from 0% for molybdenum to 64% for crystalline silica). In both databases, TLV® exceedance fractions were higher for styrene and silica (around 50-60%) compared to other agents (median <10%). For all agents in both databases, the percentage of values exceeding the TLV ® was strongly correlated, with a Spearman coefficient of 0.79.

The IMIS data cover 868 four-digit USSIC industry codes, with a median of 54 measurements per code (from 1 to 13,892 measurements). The LIMS data cover 548 four-digit CANSIC industry codes, with a median of 64 measurements per code (1 to 28,549 measurements). Table 4 lists the 10 most frequent industry codes in each of the databases. This table shows a strong concentration of IMIS measurements in metal industries (foundries, metallurgy, motors and metal parts). In the LIMS, the industries with the most measurements included metal industries as well as plastics, furniture, painting and printing industries.

Figure 2 compares a breakdown of measurements in the IMIS and LIMS by industry (CNESST industry classification system) to a breakdown of the Québec labour force (1986 Canadian census).

Table 4. Ten most frequent industries in the IMIS and LIMS

USSIC/ CANSIC Code	n	Industry
IMIS		
3321	13892	Gray and ductile iron foundries
3366	10388	Copper foundries
3714	8654	Motor vehicle parts and accessories
3443	8215	Fabricated plate work (boiler shops)
3441	8028	Fabricated structural metal
3471	6845	Plating and polishing
3341	6591	Secondary nonferrous metals
3325	6553	Steel foundries
3089	6167	Plastic products
3499	6148	Fabricated metal products
LIMS (in French)		
1699	28549	*Autres industries de produits en matière plastique* (other plastic products industries n.e.c[3])
3751	21652	*Industrie des peintures et vernis* (paint and varnish industry)
2611	17435	*Industrie des meubles de maison en bois* (wooden household furniture industry)
2542	17303	*Industrie des armoires et placards de cuisine et des coiffeuses de salle de bain en bois* (wooden kitchen cabinet and bathroom vanities industry)
2819	8314	*Autres industries d'impression commerciale* (other commercial printing industries)
3041	7918	*Industrie du revêtement sur commande de produits en métal* (custom coating of metal products industry)
3099	7578	*Autres industries de produits en métal*, NCA) (other metal fabricating industries, n.e.c.[3])
3281	6797	*Industrie de la construction et de la réparation d'embarcations* (boatbuilding and repair industry)
3081	6624	*Ateliers d'usinage* (machine shop industry)
2941	6064	*Fonderies de fer* (iron foundries)

Figure 2 shows a key contrast between the census statistics and those of the two databases, with most of the labour force in secondary industries and the service sector whereas measurements are concentrated in the primary industries and manufacturing. The construction industry is an exception in that the proportion of measurements in the IMIS from this sector almost equals the proportion of the labour force working in this industry. Overall, there is no correlation between the

[3] n.e.c = not elsewhere classified

percentage of the labour force working in an industry and the percentage that industry represents of the records in the IMIS (Spearman's correlation coefficient = 0.17) or the LIMS (Spearman's correlation coefficient = 0.05). On the other hand, the breakdown of measurements across the different industries was quite similar in the IMIS and LIMS (Spearman's correlation coefficient = 0.64). Figure 2 highlights a few key differences: more measurements in the IMIS for commercial services, construction, the textile industry, electrical products manufacturing and primary metal manufacturing; and more measurements in the LIMS for health and social services, teaching, printing and plastics.

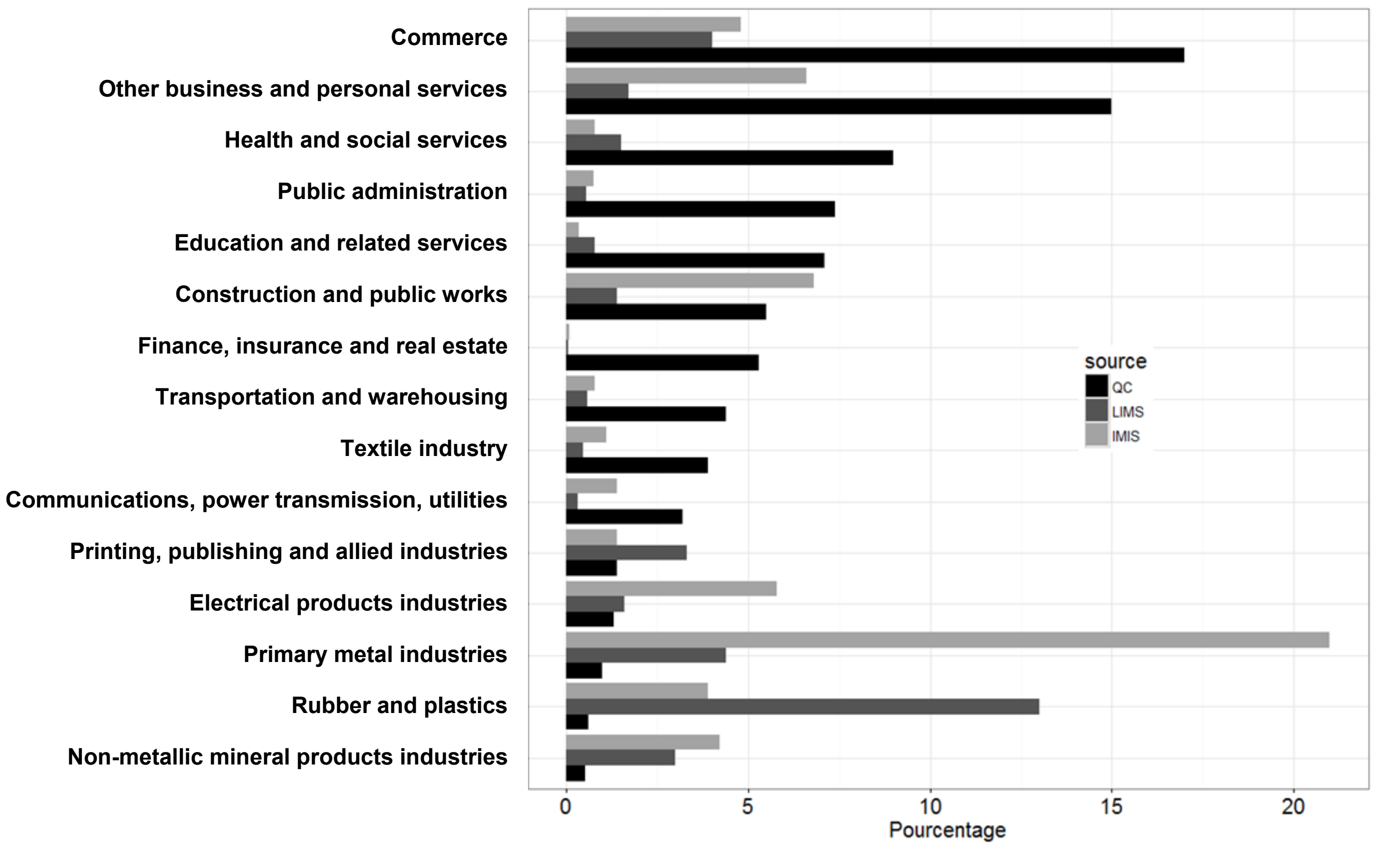

Figure 2. Breakdown of IMIS and LIMS data and Québec's 1986 labour force by industry (CNESST classification system), only industries representing ≤ 3% of any one data source included (15 of 32).

Table 5 gives a breakdown of the IMIS and LIMS data and the Québec labour force by the CNESST's target priority groups. These groups are associated with intervention priorities of Québec's public occupational-health network, the Réseau de santé publique en santé au travail. Establishments whose economic activities are listed in priority group 1 or 2 are subject to all prevention mechanisms provided for in the Act respecting occupational health and safety: prevention program, health program, health and safety committee (HSC) and prevention representative. These first two priority groups are associated with 69% of the LIMS data, though they employed only 14% of the 1986 labour force. The breakdown of the IMIS data is very similar, with differences mainly in priority groups 3, 4 and 5.

Table 5. Breakdown of IMIS and LIMS data and the 1986 Québec labour force by CNESST priority group

CNESST priority group	Labour force (%)[(a)]	Measurements (%)		Industry
		LIMS[b]	IMIS[c]	
1	9.4	28	27	construction and public works; chemical industry; forestry and sawmills; mines, quarries and oil wells; metal fabricating industries
2	4.8	41	35	wood industry (excluding sawmills); rubber and plastic products industries; transportation equipment industries; primary metal industries; non-metallic mineral products industries
3	17	12	4.6	public administration; food and beverage industries; furniture and fixtures industries; paper and allied industries; transportation and warehousing
4	22	7.3	17	commerce; leather industries; machinery industries (except electrical machinery); tobacco products industry; textile industries
5	21	6.8	13	other business and personal services; communications, power transmission and other public utilities; printing, publishing and allied industries; petroleum and coal products industries; electrical products industries
6	26	4.1	3.2	agriculture; knitting mills and clothing industries; education and related services; finance, insurance and real estate industries; health and social services; hunting and fishing; miscellaneous manufacturing industries

a. Percentage of Québec labour force according to 1986 Canadian census
b. Percentage of LIMS records
c. Percentage of IMIS records

4.2 Comparison of lists of industries with measured exposures in the IMIS and LIMS

The purpose of this analysis was to answer the following questions: do the two databases report the same industries as demonstrating air contamination?

A total of 2,898 IMIS agent-USSIC pairs (89% of detected values in the IMIS, n=180,429) and 1,948 LIMS agent-CANSIC pairs (96% of detected values in the LIMS, n=188,458) met the inclusion criteria for an agent (that is, at least six pairs with at least 10 detected results). Seventy percent of the initial agents (36 of 49) were included in the comparison: it was possible to link 62.8% of the agent-CANSIC pairs of the LIMS with an agent-USSIC pair in the IMIS using the concordance table. Working the other way, it was possible to link 61.4% of the IMIS agent-USSIC pairs with an agent-CANSIC pair in the LIMS.

These results were stratified by class of chemical agent, first for the LIMS agent-CANSIC pairs linked with the IMIS (Table 6), and then for the IMIS agent-USSIC pairs linked with the LIMS (Table 7). These tables also show results stratified by the ratio of the number of LIMS agent-CANSIC pairs/number of IMIS agent-USSIC pairs, to take into account that one of the two databases contains far fewer pairs than the other—hence a smaller proportion of linked pairs is expected. Figure 3 shows results for each agent (complete concordance tables in Appendix B).

Table 6. Proportion of LIMS agent-CANSIC pairs also found in the IMIS

Stratum	Number of valid LIMS agent-CANSIC pairs	LIMS agent-CANSIC pairs linked to the IMIS	
		Number	Proportion (%)
Global	1948	1224	62.8
Ratio of number of LIMS agent-CANSIC pairs/number of IMIS agent-USSIC pairs			
< 0.7	815	580	71.2
0.7 – 1.3	939	596	63.5
> 1.3	194	48	24.7
Chemical family			
Solvent	1,158	673	58.1
Metal	522	377	72.2
Gas	74	54	73.0
Isocyanates	61	32	52.5
Dusts	83	55	66.3
Styrene	50	33	66.0

Table 7. Proportion of IMIS agent-USSIC pairs also found in the LIMS

Stratum	Number of valid IMIS agent-USSIC pairs	IMIS agent-USSIC pairs linked to the LIMS	
		Number	Proportion (%)
Global	2,898	1,779	61.4
Ratio of number of LIMS agent-CANSIC pairs/number of IMIS agent-USSIC pairs			
< 0.7	1,856	1,024	55.2
0.7 – 1.3	985	716	72.7
> 1.3	57	39	68.4
Chemical family			
Solvent	1,161	806	69.4
Metal	1,327	746	56.2
Gas	151	78	51.7
Isocyanates	63	32	50.8
Dusts	124	78	62.9
Styrene	72	39	54.2

The results stratified by ratio of number of LIMS agent-CANSIC pairs/number of IMIS agent-USSIC pairs demonstrate the expected trend, with larger percentages of linked agent-industry pairs in the database with the smaller number of pairs. Analysis by class of agent did not reveal any strong trends inconsistent with stratification by the preceding ratio. By agent, the difference between the proportion of IMIS agent-USSIC pairs linked to the LIMS and the proportion of LIMS agent-CANSIC pairs linked to the IMIS ranged from -37.1% (cadmium) to 59.4% (ethyl acetate), with a median of -1.5% and an interquartile range of -16.6 to 15.3% (Appendix B).

The same analysis was repeated but this time with the criterion of 10 results above the ACGIH TLV® instead of 10 detected results (Appendix C). Overall, 496 LIMS agent-CANSIC pairs and 620 IMIS agent-USSIC pairs met the inclusion criteria for an agent. Only 11 of the initial agents were included. Over half (56.3%) of the LIMS agent-CANSIC pairs in the LIMS could be linked to the IMIS, whereas 61.1% of the IMIS agent-USSIC pairs could be linked to the LIMS.

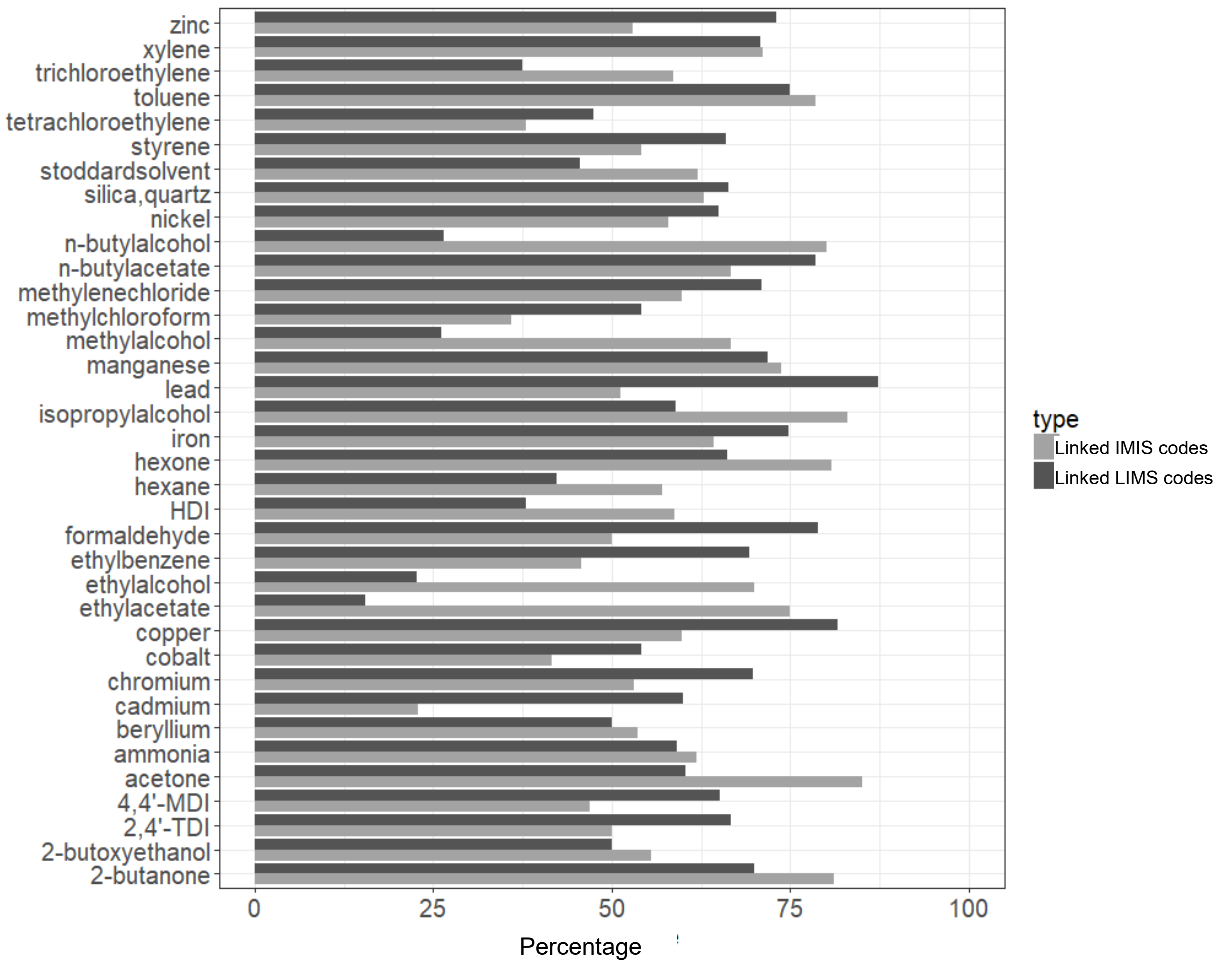

Figure 3. Percentages of LIMS agent-industry pairs linked to the IMIS and of IMIS agent-industry pairs linked to the LIMS, stratified by chemical agent.

4.3 Comparison of reported concentrations in industries common to both databases

4.3.1 Descriptive comparison

The extracts from the IMIS and LIMS include 169,388 and 267,486 records for the selected period of 1994–2011. Overall, long-term data accounted for 87% of the values in the IMIS and 84% in the LIMS. The percent of values below the LOD was 42%, 41%, 40% and 44% respectively for the IMIS long-term, IMIS short-term, LIMS long-term and LIMS short-term data sets.

A) Medians of IMIS/LIMS ratios of medians (MrM)

For the long-term data, 32 chemical agents were selected (n > 100 in the two databases) for the descriptive comparative analyses of the medians of the ratios of medians (MrM), 20 solvents and 12 metals, accounting respectively for 78% and 74% of the long-term results in the IMIS and LIMS for the period 1994-2011. For the short-term data, 22 chemical agents were selected, 14 solvents and 8 metals), accounting respectively for 63% and 55% of the short-term results in the IMIS and LIMS for the period 1994-2011.

Table 8 shows the calculated MrMs as well as the number of chemical agents and measurements for each stratified analysis. Measured solvent levels, short-term as well as long-term results, were on the whole similar in the IMIS and LIMS. Similar results were also noted for the analyses restricted to common industries according to the CANSIC classification or to specific time periods (1994-2002 and 2003-2011). Differences between the IMIS and LIMS showed up in the long-term results when the cut point for short-term measurements in the LIMS was increased to 240 minutes.

Measured metal levels were on the whole higher in the LIMS than in the IMIS, long-term as well as short-term results. Long-term results were also similar when the analyses were restricted to common industries (CANSIC, USSIC and NAICS) or to specific time periods (1994-2002, 2003-2011). The differences were not as marked when the LIMS cut point for short-term measurements was increased to 240 minutes.

Table 8. Medians of ratios of medians (MrM IMIS/LIMS) calculated according to industry classification, period and LIMS short-term cut point, for solvents and metals

	Short-term measurements[a]				Long-term measurements[b]			
	Number of agents	Number of measurements		MrM (IQR[d])	Number of agents	Number of measurements		MrM (IQR[d])
		IMIS	LIMS			IMIS	LIMS	
Solvents								
Total[c]	14	8,506	22,235	1.0 (0.4;2.0)	20	24,333	109,468	1.0 (0.7;1.6)
Common industries								
CANSIC[e]	9	2,995	12,628	1.1 (0.7;1.2)	14	8,797	62,418	0.8 (0.6;1.0)
USSIC[f]	7	2,095	4,650	0.8 (0.7;1.0)	10	6002	24,553	0.7 (0.5;0.8)
NAICS[g]	3	816	1,242	0.7 (0.4;0.9)	10	3,568	12,464	0.7 (0.5;1.5)
Period								
1994-2002	12	4,358	13,396	1.1 (0.6;1.8)	18	13,343	62,421	1.0 (0.8;1.5)
2003-2011	9	3,319	7,094	0.9 (0.3;1.5)	16	10,353	40,956	0.9 (0.5;1.8)
LIMS short-term cut point								
60 min	15	8,614	64,671	0.9 (0.5;1.7)	19	24,003	68,603	1.4 (0.9;2.2)
240 min	15	8,614	108,421	1.3 (0.9;3.3)	18	23,356	13,715	2.1 (1.3;3.9)
Metals								
Total[c]	8	5,701	1,683	0.3 (0.2;0.6)	12	91,980	56,927	0.4 (0.3;0.4)
Common industries								
CANSIC	2	2,247	348	0.3 (0.2;0.4)	10	35,455	33,099	0.4 (0.3;0.5)
USSIC	-	-	-	-	10	17,486	12,142	0.4 (0.3;0.6)
NAICS	-	-	-	-	10	8,854	5,901	0.4 (0.3;0.5)
Period								
1994-2002	2	2,951	428	0.1 (0.1;0.2)	10	50,313	26,940	0.3 (0.2;0.4)
2003-2011	2	1,743	250	0.5 (0.4;0.5)	11	39,193	29,810	0.4 (0.4;0.6)
LIMS short-term cut point								
60 min	9	5,843	34,006	0.6 (0.5;0.8)	12	91,980	52,315	0.4 (0.3;0.5)
240 min	9	5,843	49,844	0.6 (0.3;0.9)	10	87,567	7,526	0.7 (0.4;0.9)

[a] Short-term measurements: for the IMIS, STEL, peak or ceiling limits; for the LIMS, ≤ short-term cut point indicated
[b] Long-term measurements: for the IMIS,TWA; for the LIMS > short-term cut point indicated
[c] All industries in each database, period 1994-2011, LIMS short-term cut point = 30 minutes
[d] IQR, interquartile range
[e] CANSIC, Canadian standard industrial classification
[f] USSIC, US standard industrial classification
[g] NAICS, North American Industry Classification System

B) Medians of differences in TLV® *exceedance fractions (MdEF, IMIS – LIMS)*

For the long-term measurements, 33 agents were selected (n > 100 in the two databases) for the descriptive comparative analyses of the medians of the differences in TLV® exceedance fractions (MdEF) (21 solvents and 12 metals), accounting for 80% and 87% of the long-term data in the IMIS and LIMS respectively in the period 1994-2011. For the short-term measurements, 22 agents were selected (14 solvents and 8 metals), accounting for 64% and 56% of the short-term data in the IMIS and LIMS respectively for the period 1994-2011.

Table 9 shows the calculated MdEFs as well as the number of chemical agents and measurements for each stratified analysis. TLV® exceedance fractions for the solvents were on the whole similar in the IMIS and LIMS, for the long-term as well as the short-term data. Similar results were also obtained in analyses restricted to common industries (CANSIC, USSIC and NAICS) or specific time periods (1994-2002, 2003-2011).

For the metals, the TLV® exceedance fractions were on the whole higher in the LIMS than in the IMIS, for both the long-term and the short-term data. Similar results were also obtained in the analyses restricted to common industries (CANSIC, USSIC and NAICS) or specific time periods (1994-2002, 2003-2011). The differences were not as marked when the LIMS cut point for short-term measurements was increased to 240 minutes.

Table 9. Medians of differences in TLV® exceedance fractions (IMIS–LIMS MdEF) calculated according to industry classification, period and LIMS short-term cut point, for solvents and metals

	Short-term measurements[a]				Long-term measurments[b]			
	Number of agents	Number of measurements IMIS	Number of measurements LIMS	MdEF (IQR[d]) (%)	Number of agents	Number of measurements IMIS	Number of measurements LIMS	MdEF (IQR[d]) (%)
Solvents								
Total[c]	14	8,639	22,445	-1 (-7;2)	21	24,680	137,196	1 (0;2)
Common industries								
CANSIC[e]	9	3,063	12,747	-1 (-4;0)	14	8,975	77,213	0.5 (0;3)
USSIC[f]	7	2,120	4,681	0 (-2;2)	10	6,048	29,147	0 (0;1)
NAICS[g]	4	972	1,480	-0.5 (-4;3)	10	3,702	18,139	0 (0;1)
Period								
1994-2002	12	4,373	13,396	-2 (-5;1)	20	13,694	63,852	0.5 (-2;2)
2003-2011	9	3,335	7,280	2 (-6;2)	16	10,401	65,529	0 (0;1)
LIMS short-term cut point								
60 min	15	8,752	57,866	1 (-1;3)	20	24,349	95,804	1 (1;2)
240 min	15	8,752	113,692	2 (1;6)	19	23,702	31,363	2 (1;4)
Metals								
Total[c]	8	5,832	1,683	-21 (-24;-12)	12	92,874	56,927	-5 (-9;-2)
Common industries								
CANSIC	2	2,253	348	-11 (-11;-10)	10	35,685	33,209	-5 (-14;0)
USSIC	0	0	0		10	17,638	12,157	-2 (-7;0)
NAICS	0	0	0		10	8,926	5,934	-4 (-10;0)
Period								
1994-2002	2	2,951	428	-20 (-20;-20)	10	50,871	26,940	-5 (-17;-3)
2003-2011	2	1,749	250	-24 (-39;-10)	11	39,529	29,810	-4 (-10;-1)
LIMS short-term cut point								
60 min	10	6,768	6,421	-16 (-22;-12)	12	92,874	52,315	-4 (-8;-1)
240 min	10	6,768	50,915	-7 (-11;-5)	10	88,461	7,526	1 (-3;3)

[a] Short-term measurements: for the IMIS, STEL, peak or ceiling limits; for the LIMS, ≤ short-term cut point indicated

[b] Long-term measurements: for the LIMS,> short-term cut point indicated; for the IMIS, TWA

[c] All industries in each database, period 1994-2011, LIMS short-term cut point = 30 minutes

[d] IQR, interquartile range

[e] CANSIC, Canadian standard industrial classification

[f] USSIC, US standard industrial classification

[g] NAICS, North American Industry Classification System

4.3.2 Comparison by statistical modelling

A) TOBIT models

After the chemical agent selection criteria and the restriction to industries common to both databases were applied, the TOBIT model comparative analyses of long-term measurements (LIMS long-term cut point set at 30 min and 120 min) covered a total of 72,348 CANSIC classification measurements (45% IMIS and 55% LIMS), 32,727 USSIC classification measurements (54% IMIS and 46% LIMS) and 41,048 NAICS classification measurements (51% IMIS and 49% LIMS). The smaller number of measurements in the modeling analyses is due to three methodological choices: conversion of the industry codes to a classification common to both databases (that is, CANSIC, USSIC or NAICS); restriction to industries common to both databases; and restriction to long-term data. The number of solvents included ranged from 8 to 13, depending on industry classification and the LIMS long-term cut point. The number of metals ranged from 4 to 5 depending on these same factors.

To determine average difference in exposure levels between the IMIS and LIMS, predicted IMIS/LIMS geometric mean ratios were calculated by meta-analysis for the years 1997 and 2008 for both LIMS long-term cut point (30 and 120 minutes). The ratios and their 95% confidence intervals were calculated separately for the metals and solvents data for common industries only, according to the CANSIC, USSIC and NAICS industry classifications (Table 10).

For solvents, exposure levels recorded in the LIMS and the IMIS were on the whole similar regardless of industry classification or sampling year (IMIS/LIMS ratios ranged from 0.86 to 1.05 in analyses restricted to LIMS data with sampling time > 30). It is noteworthy that the results were substantially different for solvents when the analyses were restricted to LIMS data where sampling time was > 120 minutes: measured levels were on the whole lower in the LIMS than the IMIS with the 120-minute cut point (IMIS/LIMS ratios ranged from 1.33 to 2.76).

For metals, measured exposure levels were on the whole higher in the LIMS than the IMIS, regardless of industry classification or sampling year (IMIS/LIMS ratios ranged from 0.29 to 0.44 in analyses restricted to LIMS data with a sampling time > 30). We must however mention that the number of metals on which these results are based is small (4 or 5). In fact, quite a few metals were excluded from the analyses due to a high proportion of non-detect values (> 60%). Levels recorded in the LIMS remained higher than those in the IMIS when the analyses were restricted to LIMS data with the 120-minute cut point (IMIS/LIMS ratios ranged from 0.39 to 0.61).

Table 10. IMIS/LIMS ratios for solvents and metals in 1997 and 2008 according to industry classification as predicted by TOBIT modeling

LIMS long-term cut point (min)[a]	Year	CANSIC[b]		USSIC[c]		NAICS[d]	
		Number of agents	IMIS/LIMS ratio[e] (94% CI)	Number of agents	IMIS/LIMS ratio (94% CI)	Number of agents	IMIS/LIMS ratio (94% CI)
Solvents							
30	1997	13	1.00 (0.74;1.35)	10	0.98 (0.69;1.39)	11	1.05 (0.69;1.59)
	2008	13	0.86 (0.63;1.19)	10	0.98 (0.70;1.37)	11	1.02 (0.71;1.45)
120	1997	10	1.47 (0.87;2.47)	9	1.33 (0.71;2.49)	8	2.76 (1.58;4.81)
	2008	10	1.52 (1.17;1.98)	9	1.68 (1.15;2.48)	8	2.13 (1.52;2.98)
Metals							
30	1997	5	0.38 (0.22;0.64)	4	0.39 (0.23;0.67)	4	0.33 (0.21;0.51)
	2008	5	0.44 (0.32;0.60)	4	0.42 (0.32;0.56)	4	0.29 (0.18;0.46)
120	1997	5	0.46 (0.30;0.69)	4	0.48 (0.30;0.76)	5	0.61 (0.41;0.92)
	2008	5	0.50 (0.35;0.71)	4	0.47 (0.35;0.64)	5	0.39 (0.24;0.61)

[a] Long-term measurement: for the LIMS, > short-term cut point indicated; for the IMIS: TWA

[b] CANSIC, Canadian standard industrial classification

[c] USSIC, US standard industrial classification

[d] NAICS, North American Industry Classification System

[e] The overall IMIS/LIMS ratio for a scenario is the weighted mean of individual estimates across agents (obtained by meta-analysis)

B) Logistic regression models

After the chemical agent selection criteria and the restriction to industries common to both databases were applied, the comparative analyses by logistic modelling of long-term measurements (LIMS long-term cut point set at 30 min and 120 min) included a total of 92,897 CANSIC classification measurements (46% IMIS and 54% LIMS), 39,703 USSIC classification measurements (56% IMIS and 44% LIMS) and 52,836 NAICS classification measurements (51% IMIS and 49% LIMS). The number of selected solvents ranged from 6 to 12 depending on the industry classification and the LIMS long-term cut point used. The number of selected metals ranged from 7 to 10 depending on these same factors.

To determine average difference in TLV® exceedance probability between the IMIS and LIMS, predicted IMIS/LIMS odds ratios (OR) for years 1997 and 2008 were calculated by meta-analysis for the two LIMS long-term cut points of 30 and 120 minutes. Ratios and their 95% confidence intervals were calculated separately for metals and solvents for data from common industries only, according to the CANSIC, USSIC and NAICS classifications (Table 11).

For solvents, there was no substantial difference in TLV® exceedance probability between the LIMS and IMIS, regardless of industry classification or sampling year (IMIS/LIMS ORs ranged from 0.77 to 1.47 in analyses restricted to LIMS data with a sampling time > 30). When, for the LIMS, only data with a sampling time > 120 minutes were used, the overall IMIS/LIMS ORs were higher, ranging from 0.93 to 3.15.

For metals, probabilities of TLV® exceedance were systematically higher in the LIMS than the IMIS, regardless of industry classification or sampling year (IMIS/LIMS OR ranged from 0.43 to 0.73 in analyses restricted to LIMS data with sampling time > 30). It must be noted that these results are based on a greater number of metals (7 to 10) than the TOBIT modelling results (4 or 5). The exceedance probabilities also remained higher in the LIMS than the IMIS when the analyses included only LIMS data with a sampling time > 120 minutes, but the ORs were closer to null value (0.58 to 0.90).

Table 11. IMIS/LIMS odds ratios (OR) of TLV® exceedance in 1997 and 2008, according to industry classification as predicated by logistic modeling, for solvents and metals

LIMS long-term cut point (min)[a]	Year	CANSIC[b]		USSIC[c]		NAICS[d]	
		Number of agents	IMIS/LIMS OR[e] (94% CI)	Number of agents	IMIS/LIMS OR (94% CI)	Number of agents	IMIS/LIMS OR (94% CI)
Solvents							
30	1997	12	1.47 (0.91;2.38)	9	1.16 (0.61;2.19)	8	1.00 (0.46;2.16)
	2008	12	0.99 (0.58;1.69)	9	0.88 (0.41;1.87)	8	0.77 (0.61;0.96)
120	1997	7	2.42 (1.34;4.35)	7	3.15 (1.19;8.35)	7	1.15 (0.53;2.49)
	2008	8	1.14 (0.67;1.95)	6	0.93 (0.38;2.29)	6	1.70 (1.18;2.44)
Metals							
30	1997	10	0.43 (0.30;0.62)	8	0.73 (0.43;1.22)	10	0.56 (0.39;0.81)
	2008	10	0.57 (0.42;0.77)	7	0.69 (0.35;1.39)	10	0.49 (0.32;0.75)
120	1997	10	0.58 (0.38;0.87)	8	0.90 (0.49;1.68)	10	0.75 (0.47;1.22)
	2008	10	0.65 (0.49;0.87)	9	0.58 (0.30;1.12)	10	0.65 (0.42;1.01)

[a] Long-term measurements for the LIMS: > long-term cut point indicated; for the IMIS, TWA

[b] CANSIC, Canadian standard industrial classification

[c] USSIC, US standard industrial classification

[d] NAICS, North American Industry Classification System

[e] The overall IMIS/LIMS ratio for a scenario is the weighted mean of individual estimates across agents (obtained by meta-analysis)

5. DISCUSSION

5.1 Preliminary remarks

The LIMS contains the results of analyses performed by the IRSST since the early 1980s in Québec, mainly for stakeholders in the province's public occupational-health network (RSPSAT). The LIMS was on occasion used to identify industries where levels of exposure to chemical contaminants were high (Ostiguy et al., 2011; Ostiguy et al., 2012). In doing so, the authors of the studies always insisted on the limitations of the LIMS data and their studies, mainly because the data were raw and the sampling circumstances were not recorded. The LIMS was recently used to determine the number of Québec workers exposed to carcinogenic agents. It is noteworthy that the researchers report that they considered using the US IMIS as a secondary source but rejected this option because of the suspicions of bias suggested in the literature (Labrèche et al., 2012).

In sum, the LIMS is a database of the results of sampling performed in Québec, but interpretation of these results is complicated because of the lack of contextual information. The IMIS, on the other hand, mainly comprises measurements taken in workers' breathing zones, and sampling times are representative of durations pertinent for the permissible exposure limit values considered. The applicability of the IMIS measurements to Québec is, however, unknown.

Through an exhaustive comparison of the content of the LIMS and IMIS, this research lays the groundwork for joint use of these two sources of information to document occupational exposure in Québec as well as more broadly throughout North America.

5.2 Descriptive analysis

The descriptive analysis (Section 4.1) confirms a shared limitation of the two databases, but one that is without doubt more serious in the LIMS: shortcomings in the contextual information for each measurement. Included with the LIMS sampling data are an industry code (CAEQ1984), the reason for the intervention and the sampling time (after 1994). The lack of information on occupation, sample type (personal exposure or ambient air) and detection limits for non-detect values is the key shortcoming. IMIS records provide a bit more information (company inspected and sample type), and a link to another public database makes it possible to obtain a profile of violations at each establishment (Sarazin et al., 2016). The job that was monitored, however, is identified only as a nonstandarized character string. With respect to contextual information, the French COLCHIC and German MEGA databases are the gold standard, though they are far from containing all information deemed important by a number of groups of experts (Lippmann, 1995; Rajan, Alesbury, Carton and Gerin, 1997; Stamm, 2001; Vincent and Jeandel, 2001).

Regarding the number of records, though the current study covers only chemical agents common to the LIMS and IMIS, the initial extracts from these databases, described in other publications (Lavoué et al., 2013; Lavoué et al., 2012), each include some 500,000 measurements (to about 2010). This is remarkable, given the huge difference in the active populations of Québec (4 million in 2017) and the United States (160 million in 2016). The wealth of the Québec LIMS data is thus considerable, given the size of the active Québec population, even compared to the COLCHIC database (about 1 million measurements for an active population in 2013 of 40 million) and the MEGA database (about 2 million measurements for an active population in 2014 of 42 million)

(Mater et al., 2016; Stamm, 2001). With respect to Canada, Figure 4 compares the number of records in the Canadian Workplace Exposure Database (CWED), recently created by a British Columbia team, and the number in Québec's LIMS (Hall et al., 2014; Lavoué, Sauvé and Sarazin, 2014). Here again, the Québec data is remarkable for its richness compared to the data collected in the rest of Canada. These numbers clearly show the importance of making the best possible use of the Québec occupational exposure data.

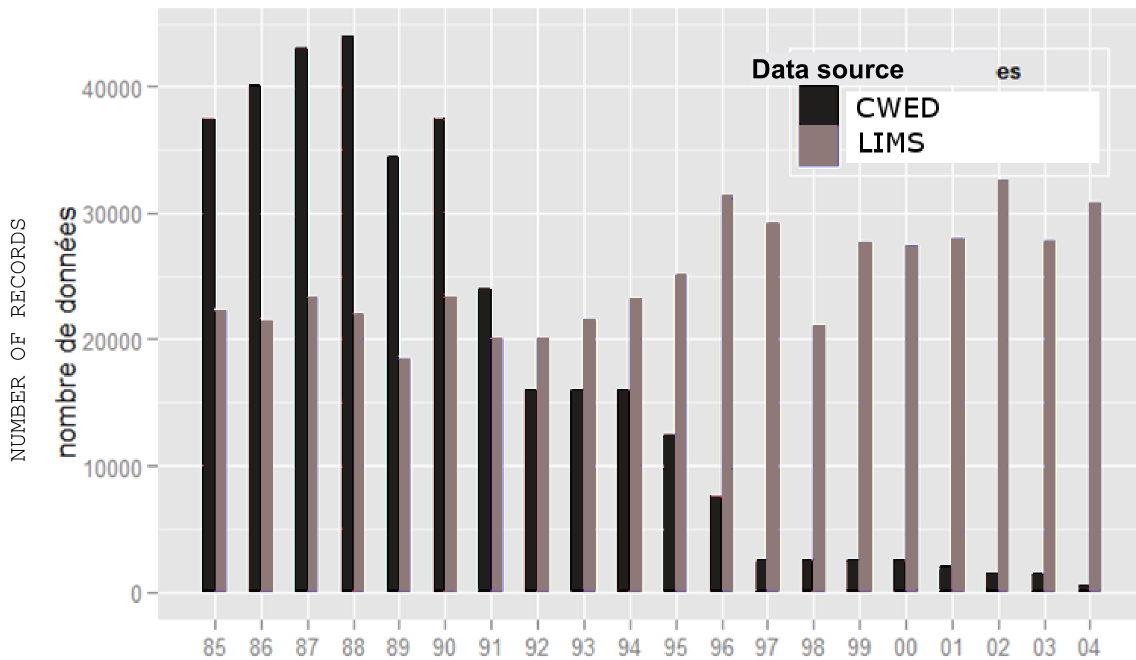

Figure 4. Number of records per year in the CWED and LIMS (1985-2004).

In terms of the breakdown of measurements by agent, our analysis shows a marked difference between the IMIS and LIMS, with measurements of solvents predominant in the LIMS while metals are much more present in the IMIS. Despite these differences, the chemical agents most frequently encountered are similar. Though published information on this topic is fragmentary, it seems that the French and German databases are similar to the LIMS in terms of the breakdown by agent (Mater et al., 2016; Stamm, 2001; Vincent and Jeandel, 2001). In all these databases, the most frequent agents are lead, toluene, manganese and crystalline silica in its various forms. The differences or similarities between the databases may be explained by differences in industrial fabric or in government action priorities or by relatively similar lists of occupational exposure limit values in the different countries. In all cases, a key feature of large exposure databases must be acknowledged: their content reflects the study of known risk factors, which makes them less helpful in detecting new problems.

5.3 Industry comparisons

The IMIS and LIMS cover a multitude of industries (4-digit codes: 500 in the LIMS and 900 in the IMIS). The difference in the number of industry codes is not significant, as it simply reflects the greater complexity of the U.S. classification: the 1987 SIC classification system (USSIC) includes 1,504 four-digit industries, whereas the Canadian SIC-E 1980 classification system (CANSIC) has 860.

The measurements are mainly concentrated in the manufacturing sectors, though there are more workers in tertiary sector industries, a disparity consistent with the findings of Mater et al. (2016) with respect to COLCHIC. Along the same lines, 70% of the records in the IMIS and LIMS are for the CNESST's two top priority sectors, mainly manufacturing (primary and secondary sectors), which employ 14% of the active population. This is because chemical products are far less frequent in the tertiary sector compared to their use in specific manufacturing processes. Apart from the overall similarities between the IMIS and LIMS regarding the two top priority sectors, the breakdown of the measurements in the two databases is quite different, with little correlation in the number of measurements per CNESST industry category. These differences may reflect differences in government priorities or in industrial fabric.

As mentioned, the fact that the measurements recorded in occupational exposure databases are for agents whose health effects are known limits their usefulness in the detection of new problems. These databases can nonetheless be helpful in identifying industries where these agents are present. To compare the IMIS and LIMS in this role and evaluate the possibility of their use to complement one another, the research team tried to determine the extent to which contaminants were identified in the same industries.

Overall, regardless of the direction of linking (IMIS to LIMS or LIMS to IMIS), about two-thirds of the industries for which measurements were recorded were identified by both databases. Two clear trends were noted in addition to this overall finding. First, an imbalance in the proportion of links appeared when the number of agent-industry pairs in one bank was smaller than in the other: fewer pairs in one bank meant a higher proportion of links, which, purely numerically, seems logical. On the other hand, the proportion of links was greater from the LIMS for metals and from the IMIS for solvents. This trend probably stems from the imbalance between the two databases in the number of records for these two chemical families. The median of variability between the individual agents in the comparison was close to zero, and the extreme values again corresponded to a major difference in the number of measurements in the IMIS and LIMS.

To the authors' knowledge, this analysis is the first of its kind to assess the respective capacities of two occupational exposure databases (OEDB) to identify exposure situations. The only study found that undertakes a similar comparison is a 1992 study by Valiante, Richards and Kinsley (1992). These researchers compared lists of USSIC codes identified by three different data sources for silica exposure: a silicosis case registry (code identified by occupation deemed associated with exposure); the IMIS, completed by data from the mining industry collected by the Mine Safety and Health Agency (MSHA) (code identified by at least one detected result in a USSIC code); and the NOES study (Boiano and Hull, 2001; Sieber, Sundin, Frazier and Robinson, 1991), which involved identifying products present in a representative sample of US industries during preliminary assessments by industrial hygienists, without measurements being taken. Of the 204 SIC codes identified by the three data sources, 9% were identified by all three sources and 33% were identified by two or more. The researchers concluded that given the limitations of each of the sources, their joint use is essential when establishing priorities.

It is also important to note that different researchers use different criteria to identify potential exposure circumstances from OEDBs. In a study by Ostiguy et al. (2012) that looked at measurements in the LIMS, for example, an industry was pinpointed if at least 20% of the measurements exceeded 50% of Québec's regulatory value and there were at least 25 measurements for the industry. In the report by Labrèche et al. (2012) that estimated workers exposed to carcinogens, industries with at least two measurements 20% above the regulatory standard were considered as potentially exposing workers. For Valiante et al. (1992), one detected value was sufficient. Clearly, the selection criteria depend on the purpose of the study: for example, determination of possible exposure as opposed to identification of excessive exposure situations. The criterion used for the present study (at least 10 detected values) is appropriate for identification of exposure situations regardless of the intensity of the exposure. It is noteworthy that similar percentages of links between the IMIS and LIMS were noted by restricting the analysis to industries with excessive exposure situations (at least 10 measurements above the ACGIH-recommended TLV®).

Lastly, this analysis showed a connection for most circumstances associated with exposure in the two databases, but the proportion of unlinked situations demonstrates the need to use both databases together to monitor exposure.

5.4 Comparison of exposure levels

The comparison of exposure levels was limited to the two main families, metals and solvents, which account respectively for 82% and 80% of the extracts initially obtained from the LIMS and the IMIS. Separate analyses were conducted for each of the two families, as abundance differs in the two databases (more metals in the IMIS but more solvents in the LIMS) and preliminary analyses showed differences in exposure levels. The results of the quantitative comparisons between the IMIS and the LIMS will thus be discussed separately for metals and solvents.

Sampling time also seems to be an important determinant in exposure levels in both the LIMS and IMIS. Whereas sampling time is divided into two categories in the IMIS, short-term or long-term, based on regulations, in the LIMS, the actual sampling time in minutes is given with the analytical results. There is substantial literature demonstrating that exposure levels are systematically lower for long-term samples than for short-term ones. This was noted with respect to the IMIS data (Lavoué et al., 2013; Lavoué et al., 2008; Sarazin et al., 2016) and COLCHIC (Lavoué et al., 2011; Mater, 2016). It has also been reported with respect to Québec data in an analysis of volatile organic compound measurements from the HYGIÈNE database (Bégin, Gérin, Adib, Fournier and Deguire, 1995) for a pilot project on establishing a Québec exposure measurement database funded by the IRSST (Lavoué et al., 2012). In the present research, all comparisons of long-term data in the IMIS and the LIMS were affected by the long-term data cut point selected for the LIMS. The long-term results are thus discussed in light of the most restrictive cut points used for this analysis, that is, > 240 minutes for the descriptive analysis and > 120 min for the modeling analysis (the 240-minute cut point was eliminated for the modeling analysis because it considerably limited sample size). We want to point out here that though the long-term data in the IMIS are considered representative of a full work shift, joint analysis of the IMIS and the *Chemical Exposure Health Data* (CEHD) database, which contains laboratory analysis results, showed, when sequential measurements were grouped, that sampling time was about 240 minutes for solvents (about 60 minutes per individual sample) and 400 minutes for metals (about 200 minutes per individual sample). The Québec OEDB pilot project mentioned earlier made it possible to study 5,000 sampling results from the LIMS by complementing these results with information from reports written by RSPSAT industrial hygienists. When sequential measurements were grouped together, the median sampling time was 240 minutes, converted to 410 minutes when the reported objective was evaluation of exposure over a full work shift. The impossibility of linking sequential samples in the LIMS and calculating a weighted result corresponding to a full work shift was thus a limitation in studying this factor, and the addition of a variable allowing such grouping is a major recommendation stemming from this work.

Overall, the descriptive and modeling analyses demonstrated that exposure levels were lower for metals in the IMIS than in the LIMS, short-term as well as long-term results, by a factor of about 2. Using the most restrictive time cut points reduced the difference, but the trend remained the same. These overall observations were valid for most of the metals studied. As for the solvents, the descriptive analyses showed similar short-term levels overall in the two databases, though appreciable differences between agents were noted. For the long-term data, application of the most restrictive cut points suggested slightly higher levels in the IMIS. This difference, however, was not observed with certain approaches (MdEF) or for certain strata (industries and years). Overall, substantial differences between the solvents studied were reported. It is noteworthy that the percentage of non-detect results for metals was appreciably higher in the IMIS (62%) than in the LIMS (51%), whereas for solvents it was much higher in the LIMS (44%) than in the IMIS (23%).

We must mention as well that the LIMS data are from both personal and ambient air sampling, the proportion of each unknown. This may have contributed to the differences in exposure levels noted between the two databases, given that the IMIS contains only personal sampling results. In fact, some studies analyzing chemical agent data from the French database COLCHIC systematically conclude that concentrations measured during personal sampling are higher than those measured during sampling of ambient air in the workplace (Kauffer and Vincent, 2007; Lavoué, Vincent and Gerin, 2006; Vincent and Jeandel, 2001), though the reverse can sometimes also occur (in case of air sampling close to an emission source). If it is true that ambient air measurements are systematically lower, then the LIMS exposure levels must underestimate real exposure in the case of agents for which more of the results are from ambient air sampling. This could not be investigated, however, because this information is not available in the LIMS. Lavoué et al. (2012) report 25% of the LIMS data as ambient air measurements in a pilot project to computerize RSPSAT data.

The results of our comparative analyses show differences between exposure levels recorded in the IMIS and in the LIMS to be of the same order of magnitude as those reported in studies of other exposure databases, if not smaller. For example, Olsen, Laursen and Vinzents (1991) compared measurements in the Danish database ATABAS with a solvent exposure dataset from the Danish furniture industry. The results showed higher exposure levels in ATABAS for toluene (median exposure levels higher by a factor of 5 to 10). Vinzents, Carton, Fjeldstad, Rajan and Stamm (1995) compared xylene exposure results for wood-working and spray-painting establishments in five European OEDBs: ATABAS (Denmark), COLCHIC (France), EXPO (Norway), MEGA (Germany) and NEDB (UK). They found reported exposure levels to be higher in the three databases composed of data from inspections checking compliance with environmental standards (ATABAS, EXPO, NEDB) than in the two databases comprising data collected for insurance purposes (geometric mean ratio of exposure levels = 4). Peters et al. (2011) compared measurements of exposure to respirable crystalline silica (RCS) in a Canadian OEDB with those recorded in several European OEDBs: exposure levels were lower in the databases of Northern Europe than in those of the UK and Canada (largest geometric mean ratio observed was 4.5, between the UK and Northern Europe). Lavoué et al. (2011) report similar portraits of formaldehyde exposure in the IMIS and COLCHIC despite potential for very different occupational settings and industry profiles in the two countries. Lastly, Mater (2016) used the meta-analytical approach employed in the present study to compare findings reported in two French OEDBs for a set of chemical agents between 2007 and 2015. This study found exposure levels were higher in 2007 (by a factor of two) in the COCHIC database, whose purpose is prevention, than in the SCOLA database, whose purpose is regulatory compliance, but become comparable in more recent years. We note that results obtained with specific agents seem generally more extreme than those from multi-agent meta-analytical studies such as the study by Mater or the present research.

Sarazin et al. (2016) also investigated, in the IMIS, the association between exposure levels for 77 chemical agents (90% of the content of the database) and variables that reflected characteristics of the inspection and the establishment visited. Overall, the study showed that elements of the process of selecting worksites for inspection influence exposure levels reported in the IMIS. Higher exposure levels were generally measured during follow-up inspections than planned inspections (by a factor of 1.6). A similar difference was noted between exposure levels measured during inspections triggered by an employee complaint and those recorded in planned inspections. Exposure levels measured in establishments with a history of noncompliance were also generally higher (by a factor of 1.5). Lastly, Sarazin et al. (2018) investigated the factors associated with the selective recording of results into the IMIS by cross-referencing with the database of CEHD sampling results for all chemical agents present in both databases. The study showed the overall proportion of CEHD sampling results recorded into the IMIS was 38% and that ND results (particularly those that were part of a panel of agents, a panel of metals for example) were less likely (by a factor of 1.7) to be recorded into the IMIS.

In sum, the exposure information in the two databases was on the whole consistent. The comparison of exposure levels recorded in the IMIS and LIMS showed moderate differences, considering the uncertainty associated with any effort to assess occupational exposure. Our findings open the door for future joint use of the data, which will in turn improve our exposure knowledge and render more robust the information on substances common to the two databases. To support appropriate use of the IMIS and LIMS data, new studies of specific chemical agents should be conducted to explore in detail the differences between the two databases.

5.5 Methodological considerations

Due to a number of considerations, caution is in order in interpreting the results of this research.

A major source of uncertainty stems from the lack of information on the tasks and industrial processes associated with each measurement, documented only by industrial code without indication of occupation. Without such information being taken into account in the calculations, differences noted between average exposure levels could reflect a differential distribution of the measurements for these variables rather that a real difference.

The difficulty linking industries in databases coded using different classification systems was a major obstacle and an additional source of uncertainty: the existence of multiple links in the CANSIC/USSIC concordance table was a major factor in reducing the number of measurements available for exposure level comparisons. In fact, though the initial extracts from the IMIS and LIMS each included about 500,000 records, only 50,000 to 100,000 records from each database were included in the exposure level comparisons depending on the analysis. The difficulty linking industry classes was largely responsible for this: the CANSIC and USSIC systems made it possible, respectively, to link 28% and 53% of the records, and NAICS system data were available for 22% and 37% of IMIS and LIMS records respectively. The comparative analysis of agent-industry pairs was affected to a lesser degree thanks to the development of an approach that made match-ups possible via one or more concordance table links, such that in the end more than 90% of detected values in the extracts could be included in the analysis.

These elements underscore the need, in both the LIMS and the IMIS, for more ancillary information for each record in the form of codes from a variety of occupational and industry classification systems, which would facilitate links with other OEDBs and with demographic information and disease registries. In a pilot project on creating a Québec database from RSPSAT data, Lavoué et al. (2012) used six classification systems (three for occupations and three for industries) making it possible to establish links with Canadian, North American and international sources. This type of initiative, meant to render exposure information systems transferable, was also noted in two recent job exposure matrixes, the CANJEM[4] system developed by the Université de Montréal (coded using three classifications of occupation-industry pairs) and the Matgéné[5] system developed by the Institut national de veille sanitaire in France (coded using two classifications of occupation-industry pairs).

We must also mention that the analyses in the present study are limited by the coding systems used, which were designed not to represent exposure situations but rather industry categories for administrative and statistical purposes, sometimes without any direct connection with occupational exposure. Coding is itself an arduous task and subject to error. In a recent article, Remen et al. (2018) report a rate of job-code agreement between two expert coders of about 60%, similar to that mentioned in other literature. This rate likely overestimates the quality of the coding in OEDBs, where coding is a secondary task not performed by experts. Coding errors in the IMIS and the LIMS have undoubtedly caused underestimation of the real matching of industries identified in the LIMS with those identified in the IMIS.

The lack of information in the LIMS on sampling times for the period prior to 1994 also considerably reduced the number of records available, given that only 64% of the LIMS records and 44% of the IMIS records are from 1994 or later. Though the number of records included in the comparison is small compared to the number in the initial extracts, Ostiguy et al. (2012) and Mater et al. (2016) report the same type of restrictions on their data. These restrictions are motivated by a desire to include only comparable and pertinent records in the analysis, but in the end they preclude extrapolation of the findings to the two databases as a whole, as only about 20 of the 49 agents initially considered could be included in the exposure level comparisons.

Another limitation of this research stems from the large number of agents in the IMIS and LIMS OEDB databases. The inclusion of all records for agents common to both databases is

[4] http://www.canjem.ca

[5] http://exppro.santepubliquefrance.fr/exppro/matgene

unquestionably a strength of the research, because it meant a representative overall portrait of the two information systems as a whole could be drawn. However, in-depth study of results specific to each agent was unrealistic. In comparing the industries covered, exhaustive lists of industries common to both databases or exclusive to one could not be developed to try to explain each difference, as hundreds of agent-industry pairs would have had to be described. Likewise, for the quantitative comparison of exposure levels, no attempts were made to interpret each individual result by agent/industry; only overall trends or trends for families of agents were described.

The management of reported measurements below the level of detection (NDs, non-detects) merits some attention. NDs account for close to half the records in the two databases, though neither documents them adequately: that is, neither database specifies the limit of detection (LOD) expressed as an air concentration for each ND result. A trainee, a research agent and an industrial hygienist spent months using an empirical approach and a bibliographic approach based on historical OSHA and IRSST analytical methods to develop matrices that provided an LOD for every possible combination of agent, analytical method, sampling time, year and database. Comparisons of the different approaches highlighted the limitations of each, and their numerous assumptions, and, though the findings of this project can be shared with the scientific community, the managers of the IMIS and LIMS systems must be encouraged to link all ND results with corresponding LODs. Over the last decade, the handling of ND results has been increasingly addressed by literature in the field of occupational hygiene (Helsel, 2010; Ogden, 2010), and a number of methodological approaches, including the one used in the present research, have been described. All current approaches, however, require knowing the LOD for each ND result.

Levels of exposure to chemical agents in the IMIS were compared to those reported in the LIMS using descriptive approaches and multiple regression statistical models (Burstyn and Teschke, 1999; Hamm and Burstyn, 2011; Lavoué et al., 2008; Lee et al., 2015; Sauvé et al., 2013). The descriptive approaches made it possible to include more chemical agents but did not allow the impact of other factors on exposure levels to be assessed. TOBIT modeling made it possible to analyze quantitative exposure, but LODs had to be assigned to all ND values in the IMIS and the LIMS. With logistic modeling, the problem of assigning LODs could be avoided, but exposure levels had to be dichotomized according to the reference period of the TLV®, entailing loss of information on magnitude of exposure. The strategy used in the present research, which involved comparing the IMIS and LIMS data using several approaches, is of particular interest because it allows the consistency of the results to be verified independent of the methodological problems associated with each approach. In this study, the differences between the databases were similar regardless of the comparative approach used, lending strength to our conclusions. Though trends were in the same direction, determining if the ratios of medians (rM), differences in TLV® exceedance fractions (dER), geometric mean ratios and odds ratios (ORs) reported for the descriptive and modeling approaches reflect quantitative differences of similar magnitude remains complex. These indicators measure different quantities: for example, an IMIS/LIMS odds ratio of 2.0 corresponds to a difference in TLV® exceedance fraction of about 5% for an exceedance fraction in the LIMS equal to 5%.

We must also point out a limitation on interpreting TLV® exceedance fractions. These fractions cannot be formally interpreted as deriving from a situation of non-compliance with ACGIH recommendations. For one thing, the TLV® for the year of each measurement was not used, and for another, the TLV®-TWA was always used regardless of sampling time. Lastly, the ancillary information accompanying the LIMS measurements does not make it possible to know which measurements can be formally compared to an exposure limit value or which can be grouped together to make the comparison. For our research project, exceedance fractions were calculated so exposure levels in the two databases could be compared without the pitfalls of processing NDs. These fractions give an indication of the number of measurements that exceed a limit, but they cannot be directly associated with an occupational health risk.

Lastly, as mentioned, it cannot be assumed by default that the data recorded in the IMIS and LIMS are representative of occupational exposure among Québec workers in general, as not all industries are equally well documented in either database, nor is any one industry necessarily equally well documented in both databases. Questions have also been raised about the reliability of OEDB data. The representativeness, accuracy and reliability of the measurements are decidedly affected by, among other things, the reason for the sampling, the industrial structure of the country or region concerned and the effective occupational health and safety (OHS) standards and government policies. Measurements may have been taken for any of a wide variety of reasons, including an exploratory inspection, a regulatory compliance inspection, a remedial order or safety citation or follow-up on a problem, a complaint or presence of disease. Industries and occupations represented are definitely influenced by how industries are structured in the two databases (climate-related industries such as agriculture and fishing, for example) and probably reflect government priorities more than the real distribution of risk factors. Nonetheless, without a statistical yardstick to use to compare the contents of the LIMS and IMIS, the observed relative comparability of the databases regarding industries covered, agents measured and exposure levels is reassuring.

5.6 Recommendations

For managers of occupational exposure databases (OEDBs):

- Add to the LIMS a variable that allows grouping of samples collected from a workstation over a workday.
- Add applicable limits of detection and quantification for each record.
- Use several industry and occupation classification systems to facilitate linking with other sources of information.
- Add contextual information to facilitate interpretation of measurements as per international recommendations since the 1990s.

For users (researchers and monitoring institutions):

- Systematically consider complementary use of the LIMS and IMIS databases as sources of information on occupational exposure in North America.
- Compare in detail the portraits drawn by the two databases separately before using them jointly in projects focusing on particular circumstances with respect to agents or industries.
- During these comparisons, consider factors demonstrated to be important overall in the analyses conducted for the present research or in other publications (sampling time, for example).

6. CONCLUSION

Many activities essential to the prevention of occupational diseases depend on the availability of reliable sources of information about occupational exposure to substances or chemical. In this respect, national occupational exposure databases seem to have tremendous potential, but their use is still limited by questions about the representativeness of their content coupled with a lack of contextual information for the measurements. The present study demonstrates that the IRSST's LIMS and the IMIS regulatory database in the U.S. in general draw consistent and complementary portraits of occupational exposure to the agents included in the analysis. The results suggest, accordingly, that the two data sources should be used in combination in future exposure assessment projects. The main limitations of the LIMS derive from the lack of contextual information, highlighting the priority that must be given to creating a real Québec occupational exposure database. As suggested in the pilot project conducted by Lavoué et al. (2012), such a database could be created by combining the LIMS analytical results with information about the exposure circumstances available in RSPSAT reports, putting to excellent use a unique resource developed in Québec since the 1980s. Lastly, though the results of the present research are on the whole reassuring, questions remain as to how representative of the general population the LIMS and IMIS data are. Regardless of the questions raised, however, we can definitely expect growing use of the huge amount of data available in these databases. It is thus important to encourage evaluative research designed to improve our understanding of possible bias, keeping in mind that there are very few other sources of information. This study paints a summary portrait that will serve as a useful reference for future studies of specific agents that might want to include in their interpretations knowledge pertinent to a particular substance.

BIBLIOGRAPHY

ACGIH. (2017). *2017 TLVs and BEIs*. Cincinnati, USA: ACGIH.

AQHSST. (2004). *Manuel d'hygiène du travail - Du diagnostic à la maîtrise des facteurs de risque*. Montréal, QC: Modulo-Griffon.

Bégin, D., Gérin, M., Adib, G., Fournier, C. and Deguire, L. (1995). Development of an Occupational Exposure Data Bank on the Territory of a Department of Community Health in Montreal. *Appl Occup Environ Hyg, 10,* 355-360.

Boiano, J. M. and Hull, R. D. (2001). Development of a National Occupational Exposure Survey and Database associated with NIOSH hazard surveillance initiatives. *Appl Occup Environ Hyg, 16*(2), 128-134. doi: 10.1080/104732201460217

Borenstein, M., Hedges, L. V., Higgins, J. P. and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods, 1,* 97-111.

Bureau de la statistique du Québec. (1984). *Classification des activités économiques du Québec 1984*. Montréal, QC: Bureau de la statistique du Québec.

Burstyn, I. and Teschke, K. (1999). Studying the determinants of exposure: A review of methods. *Am Ind Hyg Assoc J, 60*(1), 57-72. doi: 10.1080/00028899908984423

Clerc, F., Bertrand, N. and Vincent, R. (2014). TEXAS: A Tool for EXposure ASsessment-Statistical Models for Estimating Occupational Exposure to Chemical Agents. *Ann Occup Hyg,* 59(3), 277-291. doi: 10.1093/annhyg/meu094

Cowan, D. M., Cheng, T. J., Ground, M., Sahmel, J., Varughese, A. and Madl, A. K. (2015). Analysis of workplace compliance measurements of asbestos by the U.S. Occupational Safety and Health Administration (1984-2011). *Regul Toxicol Pharmacol, 72*(3), 615-629. doi: 10.1016/j.yrtph.2015.05.002

Friesen, M. C., Coble, J. B., Lu, W., Shu, X. O., Ji, B. T., Xue, S., … Vermeulen, R. (2012). Combining a job-exposure matrix with exposure measurements to assess occupational exposure to benzene in a population cohort in shanghai, china. *Ann Occup Hyg, 56*(1), 80-91. doi: 10.1093/annhyg/mer080

Hall, A. L., Peters, C. E., Demers, P. A. and Davies, H. W. (2014). Exposed! Or not? The diminishing record of workplace exposure in Canada. *Can J Public Health, 105*(3), e214-217.

Hamm, M. P. and Burstyn, I. (2011). Estimating occupational beryllium exposure from compliance monitoring data. *Arch Environ Occup Health, 66*(2), 75-86. doi: 10.1080/19338244.2010.511309

Helsel, D. (2005). *Nondetects and Data Analysis: Statistics for Censored Environmental Data*. California, USA: Wiley.

Helsel, D. (2010). Much ado about next to nothing: incorporating nondetects in science. *Ann Occup Hyg, 54*(3), 257-262. doi: 10.1093/annhyg/mep092

Henn, S. A., Sussell, A. L., Li, J., Shire, J. D., Alarcon, W. A. and Tak, S. (2011). Characterization of lead in US workplaces using data from OSHA's integrated management information system. *Am J Ind Med, 54*(5), 356-365. doi: 10.1002/ajim.20926

Hewett, P., Logan, P., Mulhausen, J., Ramachandran, G. and Banerjee, S. (2006). Rating exposure control using Bayesian decision analysis. *J Occup Environ Hyg, 3*(10), 568-581. doi: 10.1080/15459620600914641

Hosmer, D. and Lemeshow, S. (2000). *Applied Logistic Regression,* 2nd Edition. New York, USA.: John Wiley and Sons, Inc.

INRS. (2015a). Base de données Fibrex. Retrieved from http://www.inrs.fr/publications/bdd/fibrex.html.

INRS. (2015b). SOLVEX. Retrieved from http://www.inrs.fr/inrs-pub/inrs01.nsf/IntranetObject-accesParReference/Rubrique9g/$FILE/visu.html.

Kauffer, E. and Vincent, R. (2007). Occupational exposure to mineral fibres: Analysis of results stored on COLCHIC database. *Ann Occup Hyg, 51*(2), 131-142. doi: 10.1093/annhyg/mel063

Kauppinen, T., Toikkanen, J., Pedersen, D., Young, R., Ahrens, W., Boffetta, P., …Kogevinas, M. (2000). Occupational exposure to carcinogens in the European Union. *Occup Environ Med, 57*(1), 10-18.

Koppisch, D., Schinkel, J., Gabriel, S., Fransman, W. and Tielemans, E. (2012). Use of the MEGA exposure database for the validation of the Stoffenmanager model. *Ann Occup Hyg, 56*(4), 426-439. doi: 10.1093/annhyg/mer097

Labrèche, F., Duguay, P., Ostiguy, C., Goyer, N., Boucher, A., Roberge, B. and Baril, M. (2012). *Carcinogenic Substances: Exposure Profile of Québec* (Report No. R-732). Montréal, QC: IRSST.

Lavoué, J., Friesen, M. C. and Burstyn, I. (2013). Workplace measurements by the US Occupational Safety and Health Administration since 1979: Descriptive analysis and potential uses for exposure assessment. *Ann Occup Hyg, 57*(1), 77-97. doi: 10.1093/annhyg/mes055

Lavoué, J., Gérin, M., Bégin, D., Ostiguy, C., Arcand, R. and Adib, G. (2012). *Valorisation des données d'exposition professionnelle mesurées au Québec depuis 1980 par les équipes du Réseau public québécois en santé au travail - Étude préliminaire* (Report No. R-723). Montréal, QC: IRSST.

Lavoué, J., Gerin, M. and Vincent, R. (2011). Comparison of formaldehyde exposure levels in two multi-industry occupational exposure databanks using multimodel inference. *J Occup Environ Hyg, 8*(1), 38-48. doi: 10.1080/15459624.2010.515553

Lavoué, J., Sauvé, J. F. and Sarazin, P. (2014). *Quebexpo: Prospects for the creation of an occupational exposure database in Québec.* Presentation for CARWH, Saskatoon, SK.

Lavoué, J., Vincent, R. and Gerin, M. (2006). Statistical modelling of formaldehyde occupational exposure levels in French industries, 1986-2003. *Ann Occup Hyg, 50*(3), 305-321. doi: 10.1093/annhyg/mei068

Lavoué, J., Vincent, R. and Gerin, M. (2008). Formaldehyde exposure in U.S. industries from OSHA air sampling data. *J Occup Environ Hyg, 5*(9), 575-587. doi: 10.1080/15459620802275023

Lee, D., Lavoué, J., Spinelli, J. and Burstyn, I. (2015). Statistical modeling of occupational exposure to polycyclic aromatic hydrocarbons using OSHA data. *J Occup Environ Hyg* 1-14. doi: 10.1080/15459624.2015.1043049

Linch, K. D., Miller, W. E., Althouse, R. B., Groce, D. W. and Hale, J. M. (1998). Surveillance of respirable crystalline silica dust using OSHA compliance data (1979-1995). *Am J Ind Med, 34*(6), 547-558.

Lippmann, M. (1995). Exposure Data Needs in Risk Assessment and Risk Management: Database Information Needs. *Appl Occup Environ Hyg, 10*(4), 244-250.

Lubin, J. H., Colt, J. S., Camann, D., Davis, S., Cerhan, J. R., Severson, R. K., … Hartge, P. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect, 112*(17), 1691-1696.

Mater, G. (2016). *Caractérisation des mesures d'exposition à des produits chimiques dans les bases de données françaises COLCHIC et SCOLA pour la prévention des maladies professionnelles.* (Doctoral dissertation, Université de Lorraine, Lorraine, France).

Mater, G., Paris, C. and Lavoué, J. (2016). Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA. *Am J Ind Med,* 59(5), 379-391. doi: 10.1002/ajim.22569

Melville, R. and Lippmann, M. (2001). Influence of data elements in OSHA air sampling database on occupational exposure levels. *Appl Occup Environ Hyg, 16,* 884-899.

Ogden, T. L. (2010). Handling results below the level of detection. *Ann Occup Hyg, 54*(3), 255-256. doi: 10.1093/annhyg/mep099

Okun, A., Cooper, G., Bailer, A. J., Bena, J. and Stayner, L. (2004). Trends in occupational lead exposure since the 1978 OSHA lead standard. *Am J Ind Med, 45*(6), 558-572. doi: 10.1002/ajim.20008

Olsen, E., Laursen, B. and Vinzents, P. S. (1991). Bias and random errors in historical data of exposure to organic solvents. *Am Ind Hyg Assoc J, 52*(5), 204-211. doi: 10.1080/15298669191364604

Olsson, A. C., Gustavsson, P., Kromhout, H., Peters, S., Vermeulen, R., Bruske, I., … Straif, K. (2011). Exposure to diesel motor exhaust and lung cancer risk in a pooled analysis from case-control studies in Europe and Canada. *Am J Respir Crit Care Med, 183*(7), 941-948. doi: 10.1164/rccm.201006-0940OC

OSHA. (2014). SIC Manual - detailed information for a specified SIC, Division, or Major Group. Retrieved from https://www.osha.gov/oshstats/index.html

Ostiguy, C., Cordeiro, R., Bensimon, G. and Baril, M. (2011). *Portrait de l'exposition professionnelle au plomb au Québec et niveaux de plombémie de janvier 2001 à décembre 2008* (Report No. R-713). Montréal, QC: IRSST.

Ostiguy, C., Morin, S., Bensimon, G. and Baril, M. (2012). *Résultats des analyses de substances chimiques produites à l'IRSST pour chaque région administrative du ministère de la Santé et des Services sociaux pour la période 2001-2008* (Report No. R-731). Montréal, QC: IRSST.

Persson, T. and Rootzen, H. (1977). Simple and highly efficient estimators for a type I censored normal sample. *Biometrika, 64*(1), 123-128.

Pesch, B., Kendzia, B., Hauptmann, K., Van Gelder, R., Stamm, R., Hahn, J. U., … Bruning, T. (2015). Airborne exposure to inhalable hexavalent chromium in welders and other occupations: Estimates from the German MEGA database. *Int J Hyg Environ Health, 218*(5)*, 500-506*. doi: 10.1016/j.ijheh.2015.04.004

Peters, S., Kromhout, H., Olsson, A. C., Wichmann, H. E., Bruske, I., Consonni, D., … Vermeulen, R. (2012). Occupational exposure to organic dust increases lung cancer risk in the general population. *Thorax, 67*(2), 111-116. doi: 10.1136/thoraxjnl-2011-200716

Peters, S., Vermeulen, R., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., … Kromhout, H. (2011). Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies. *J Environ Monit, 13*(11), 3262-3268. doi: 10.1039/c1em10628g

Rajan, B., Alesbury, R., Carton, B. and Gerin, M. (1997). European proposal for core information for the storage and exchange of workplace exposure measurements on chemical agents. *Appl Occup Environ Hyg, 12*(1)*,* 31-39.

Remen T., Richardson L., Pilorget C., Palmer G., Siemiatycki J. and Lavoue J. (2018). Development of a coding and crosswalk tool for occupations and industries. *Ann. Work Expo. Health*, *62,*(7), 796–807. https://doi.org/10.1093/annweh/wxy052

Sarazin, P. (2016). *Caractérisation des mesures d'exposition recueillies par l'agence fédérale américaine OSHA pour l'estimation des expositions professionnelles en Amérique du Nord.* (Doctoral dissertation, Université de Montréal, Montréal, QC). Retrieved from http://hdl.handle.net/1866/18583

Sarazin, P., Burstyn, I., Kincl, L., Friesen, M. C. and Lavoué, J. (2018). Characterization of the selective recording of workplace exposure measurements into OSHA's IMIS databank. *Ann Work Expo Health, 62*(3), 269-280. doi: 10.1093/annweh/wxy003

Sarazin, P., Burstyn, I., Kincl, L. and Lavoué, J. (2016). Trends in OSHA compliance monitoring data 1979-2011: Statistical modeling of ancillary information across 77 chemicals. *Ann Occup Hyg, 60*(4), 432-452. doi: 10.1093/annhyg/mev092

Sauvé, J. F., Beaudry, C., Begin, D., Dion, C., Gerin, M. and Lavoué, J. (2013). Silica exposure during construction activities: Statistical modeling of task-based measurements from the literature. *Ann Occup Hyg, 57*(4), 432-443. doi: 10.1093/annhyg/mes089

Scarselli, A., Binazzi, A. and Di Marzio, D. (2011). Occupational exposure levels to benzene in Italy: Findings from a national database. *Int Arch Occup Environ Health, 84*(6), 617-625. doi: 10.1007/s00420-011-0616-9

Scarselli, A., Binazzi, A., Marzio, D. D., Marinaccio, A. and Iavicoli, S. (2012). Hexavalent chromium compounds in the workplace: Assessing the extent and magnitude of occupational exposure in Italy. *J Occup Environ Hyg, 9*(6), 398-407. doi: 10.1080/15459624.2012.682216

Scarselli, A., Corfiati, M. and Marzio, D. (2016). Level and extent of occupational exposure to formaldehyde in the Italian industrial sectors. Presentation to the International Society for Environmental Epidemiology, Rome, Italy. Retrieved from https://ehp.niehs.nih.gov/isee/2016-p3-225-3178/

Scarselli, A., Di Marzio, D., Marinaccio, A. and Iavicoli, S. (2013). Assessment of work-related exposure to polycyclic aromatic hydrocarbons in Italy. *Am J Ind Med, 56*(8), 897-906. doi: 10.1002/ajim.22172

Scarselli, A., Montaruli, C. and Marinaccio, A. (2007). The Italian information system on occupational exposure to carcinogens (SIREP): Structure, contents and future perspectives. *Ann Occup Hyg, 51*(5), 471-478. doi: 10.1093/annhyg/mem022

Sieber, W. K., Jr., Sundin, D. S., Frazier, T. M. and Robinson, C. F. (1991). Development, use, and availability of a job exposure matrix based on national occupational hazard survey data. *Am J Ind Med, 20*(2), 163-174.

Sottas, P. E., Lavoué, J., Bruzzi, R., Vernez, D., Charriere, N. and Droz, P. O. (2009). An empirical hierarchical Bayesian unification of occupational exposure assessment methods. *Stat Med, 28*(1), 75-93. doi: 10.1002/sim.3411

Stamm, R. (2001). MEGA-database: One million data since 1972. *Appl Occup Environ Hyg, 16*(2), 159-163. doi: 10.1080/104732201460262

Statistics Canada. (2002). *North American Industry Classification System (NAICS) 1997 – Canada,* Ottawa, ON: Statistics Canada, Standards Division.

Teschke, K., Olshan, A. F., Daniels, J. L., De Roos, A. J., Parks, C. G., Schulz, M. and Vaughan, T. L. (2002). Occupational exposure assessment in case-control studies: Opportunities for improvement. *Occup Environ Med, 59*(9), 575-593; Commentary 594.

Therneau, T. M. and Lumley, T. (2016). survival: Survival Analysis. Retrieved from https://cran.r-project.org/web/packages/survival/index.html

Tielemans, E., Warren, N., Fransman, W., Van Tongeren, M., McNally, K., Tischer, M., … Cherrie, J. W. (2011). Advanced REACH Tool (ART): Overview of version 1.0 and research needs. *Ann Occup Hyg, 55*(9), 949-956. doi: 10.1093/annhyg/mer094

Valiante, D. J., Richards, T. B. and Kinsley, K. B. (1992). Silicosis surveillance in New Jersey: Targeting workplaces using occupational disease and exposure surveillance data. *Am J Ind Med, 21*(4), 517-526.

Van Tongeren, M., Jimenez, A. S., Hutchings, S. J., MacCalman, L., Rushton, L. and Cherrie, J. W. (2012). Occupational cancer in Britain. Exposure assessment methodology. *Br J Cancer, 107 Suppl 1* S18-26. doi: 10.1038/bjc.2012.114

Viechtbauer, W. (2014). Meta-Analysis Package for R. Retrieved from http://cran.r-project.org/web/packages/metafor/index.html

Vincent, R. and Jeandel, B. (2001). COLCHIC-occupational exposure to chemical agents database: Current content and development perspectives. *Appl Occup Environ Hyg, 16*(2), 115-121. doi: 10.1080/104732201460190

Vinzents, P., Carton, B., Fjeldstad, P., Rajan, B. and Stamm, R. (1995). Comparison of Exposure Measurements Stored in European Databases on Occupational Air Pollutants and Definition of Core Information. *Appl Occup Environ Hyg, 10*(4), 351-354.

Yassin, A., Yebesi, F. and Tingle, R. (2005). Occupational exposure to crystalline silica dust in the United States, 1988-2003. *Environ Health Perspect, 113*(3), 255-260.

APPENDIX A: AGENTS ANALYZED FOR IMIS-LIMS COMPARISON

Table 12. Chemical agents analyzed for IMIS-LIMS comparison

	Name in project	Original name in LIMS	Original name in IMIS
1	2-Butanone	Méthyl éthyle cétone	2-Butanone
2	2-Butoxyethanol	Butoxy-2 éthanol	2-Butoxyethanol
3	2,4'-TDI	TDI-2,4	2,4'-TDI
4	2,6'-TDI	TDI-2,6	2,6'-TDI
5	4,4'-MDI	MDI	4,4'-MDI
6	Acetone	Acétone	Acetone
7	Ammonia	Ammoniac	Ammonia
8	Antimony	Antimoine	Antimony
9	Benzene	Benzène	Benzene
10	Beryllium	Béryllium	Beryllium
11	Cadmium	Cadmium (en Cd)	Cadmium dust Cadmium fume Cadmium (Twa)
12	Chlorine	Chlore	Chlorine
13	Chromium	Chrome (en Cr)	Chromium
14	Cobalt	Cobalt (en Co)	Cobalt
15	Copper	Cuivre (en Cu)	Copper dusts Copper fume
16	Ethyl acetate	Acétate d'éthyle	Ethyl acetate
17	Ethyl alcohol	Alcool éthylique	Ethyl alcohol
18	Ethylbenzene	Éthylbenzène	Ethyl benzene
19	Ethylene oxide	Oxyde d'éthylène	Ethylene oxide
20	Formaldehyde	Formaldéhyde	Formaldehyde
21	HDI	HDI	HDI
22	Hexane	Hexane normal	Hexane
23	Hexone	Méthyl isobutyl cétone	Hexone
24	Iron	Fer (en Fe)	Iron oxide fume
25	Isopropyl alcohol	Alcool isopropylique	Isopropyl alcohol
26	Lead	Plomb (en Pb)	Lead, inorganic
27	Manganese	Manganèse (en Mn)	Manganese fume
28	Methyl alcohol	Alcool méthylique	Methyl alcohol

	Name in project	Original name in LIMS	Original name in IMIS
29	Methyl chloroform	Trichloro-1,1,1 éthane	Methyl Chloroform
30	Methylene chloride	Chlorure de méthylène	Methylene Chloride
31	Molybdenum	Molybdène	Molybdenum
32	n-Butylacetate	Acétate de butyle normal	N-Butyl Acetate
33	n-Butyl alcohol	Alcool butylique normal	N-Butyl alcohol
34	Nickel	Nickel (en Ni)	Nickel
35	Nitric acid	Acide nitrique	Nitric acid
36	Phenol	Phénol	Phenol
37	Silica, quartz	Quartz, silice cristalline	Silica, quartz
38	Silver	Argent (en Ag)	Silver
39	Stoddard solvent	Solvant Stoddard	Stoddard Solvent
40	Styrene	Styrène (monomère)	Styrene
41	Sulfuric acid	Acide sulfurique	Sulfuric acid
42	Tetrachloroethylene	Perchloroéthylène	Tetrachloroethylene
43	Tin	Étain	Tin
44	Toluene	Toluène	Toluene
45	Trichloroethylene	Trichloroéthylène	Trichloroethylene
46	Vanadium	Vanadium (en V)	Vanadium Fume
47	Vinyl chloride	Chlorure de vinyle (monomère)	Vinyl chloride
48	Xylene	Xylènes (isomères o,m,p)	Xylene
49	Zinc	Zinc (en Zn)	Zinc oxide fume

APPENDIX B: RESULTS OF COMPARISON OF INDUSTRIES WITH RECORDED MEASUREMENTS IN THE IMIS AND LIMS, BY AGENT

Table 13. Quantity of LIMS agent-industry pairs also in the IMIS, stratified by chemical agent (detected values)

Agent	Number of valid LIMS agent-CANSIC pairs	LIMS agent-CANSIC pairs linked to the IMIS	
		Number	Percentage (%)
Toluene	176	132	75.0
Xylene	127	90	70.9
Acetone	106	64	60.4
2-butanone	100	70	70.0
Isopropanol	83	49	59.0
Stoddard solvent	68	31	45.6
Methyl isobutyl ketone	65	43	66.2
Methylene chloride	62	44	71.0
Hexane	45	19	42.2
Ethyl acetate	45	7	15.6
Ethanol	44	10	22.7
Butyl acetate	42	33	78.6
Methanol	42	11	26.2
Butanol	34	9	26.5
Trichloroethylene	32	12	37.5
Ethylbenzene	26	18	69.2
1,1,1 trichloroethane	24	13	54.2
Tetrachloroethylene	19	9	47.4
2-butoxyethanol	18	9	50.0
Iron	99	74	74.7
Manganese	96	69	71.9
Zinc	63	46	73.0
Copper	60	49	81.7
Lead	55	48	87.3
Chromium	43	30	69.8
Nickel	40	26	65.0
Beryllium	32	16	50.0
Cobalt	24	13	54.2

Agent	Number of valid LIMS agent-CANSIC pairs	LIMS agent-CANSIC pairs linked to the IMIS	
		Number	Percentage (%)
Cadmium	10	6	60.0
Formaldehyde	52	41	78.8
Ammonium	22	13	59.1
HDI	29	11	37.9
4,4'-MDI	23	15	65.2
2,4'-TDI	9	6	66.7
Crystalline silica	83	55	66.3
Styrene	50	33	66.0

Table 14. Quantity of IMIS agent-industry pairs also in the LIMS, stratified by chemical agent (detected values)

Agent	Number of valid IMIS agent-USSIC pairs	IMIS agent-USSIC pairs linked to the LIMS	
		Number	Percentage (%)
Toluene	242	190	78.5
Xylene	163	116	71.2
2-butanone	95	77	81.1
Methylene chloride	92	55	59.8
Acetone	80	68	85.0
Isopropanol	65	54	83.1
Ethylbenzene	59	27	45.8
Stoddard solvent	58	36	62.1
Butyl acetate	54	36	66.7
Methyl isobutyl ketone	52	42	80.8
1,1,1 trichloroethane	50	18	36.0
Hexane	35	20	57.1
Tetrachloroethylene	29	11	37.9
Trichloroethylene	29	17	58.6
2-butoxyethanol	18	10	55.6
Methanol	12	8	66.7
Ethanol	10	7	70.0
Butanol	10	8	80.0
Ethyl acetate	8	6	75.0
Lead	203	104	51.2
Iron	202	130	64.4
Copper	182	109	59.9
Zinc	174	92	52.9
Manganese	164	121	73.8
Chromium	126	67	53.2
Nickel	100	58	58.0
Cadmium	70	16	22.9
Cobalt	65	27	41.5
Beryllium	41	22	53.7
Formaldehyde	130	65	50.0
Ammonium	21	13	61.9
4,4'-MDI	32	15	46.9
HDI	17	10	58.8
2,4'-TDI	14	7	50.0
Crystalline silica	124	78	62.9
Styrene	72	39	54.2

APPENDIX C: RESULTS OF COMPARISON OF INDUSTRIES WITH VALUES > TLV® IN THE IMIS AND LIMS, BY AGENT

Table 15. Quantity of LIMS agent-industry pairs also in the IMIS, stratified by chemical agent (values >TLV®)

Chemical agent	Number of valid LIMS agent-CANSIC pairs	LIMS agent-CANSIC pairs linked to the IMIS	
		Number	Percentage (%)
Toluene	81	49	60.5
Methylene chloride	33	18	54.5
Trichloroethylene	20	6	30.0
Manganese	84	57	67.9
Zinc	63	23	36.5
Lead	39	29	74.4
Iron	26	17	65.4
Beryllium	17	7	41.2
Formaldehyde	27	13	48.1
Styrene	31	16	51.6

Table 16. Quantity of IMIS agent-industry pairs also in the LIMS, stratified by chemical agent (values >TLV®)

Chemical agent	Number of valid IMIS agent-USSIC pairs	IMIS agent-USSIC pairs linked to the LIMS	
		Number	Percentage (%)
Toluene	81	61	75.3
Methylene chloride	31	18	58.1
Trichloroethylene	18	8	44.4
Manganese	113	88	77.9
Lead	103	51	49.5
Iron	37	20	54.1
Beryllium	25	9	36.0
Zinc	23	19	82.6
Formaldehyde	36	14	38.9
Styrene	36	20	56.0