Université de Montréal

ÉVALUATION DE L'EXPOSITION PROFESSIONNELLE AU FORMALDÉHYDE À PARTIR DE SOURCES DE DONNÉES PRÉEXISTANTES

Par

Jérôme Lavoué

Département de santé environnementale et santé au travail

Faculté de médecine

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de

Philosophiæ Doctor (Ph.D.)

en Santé publique

option Toxicologie de l'environnement

Avril 2006

© Jérôme Lavoué 2006

Faculté des études supérieures

Université de Montréal

WA
5
U58
2006
V.017

Université de Montréal

Direction des bibliothèques

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

ÉVALUATION DE L'EXPOSITION PROFESSIONNELLE AU FORMALDÉHYDE

À PARTIR DE SOURCES DE DONNÉES PRÉEXISTANTES

présentée par

Jérôme Lavoué

a été évaluée par un jury composé des personnes suivantes :

Docteur Gaétan Carrier, président-rapporteur

Docteur Michel Gérin, directeur de recherche

Docteur Guy Perrault, codirecteur de recherche

Docteur Jack Siemiatycki, membre du jury

Docteur Ben Armstrong, examinateur externe

Thèse acceptée le :

RÉSUMÉ

Les données préexistantes sur les niveaux d'exposition professionnelle constituent des sources de données privilégiées pour la surveillance de l'exposition ou l'épidémiologie. Les diverses circonstances dans lesquelles elles sont générées peuvent causer des biais pour l'évaluation de l'exposition des travailleurs. Dans le but de caractériser ces biais, et d'établir des portraits historiques de l'exposition professionnelle au formaldéhyde, récemment classé cancérogène chez l'humain par le Centre international de recherche sur le cancer, nous avons analysé plusieurs sources de données d'exposition à cette substance.

Cinq sources de mesures d'exposition au formaldéhyde ont été exploitées. Dans le secteur des panneaux de bois aggloméré, elles comprenaient les données issues d'une revue de littérature, les données mesurées au Québec de 1984 à 2002 par des hygiénistes du gouvernement et en 2001-2002 par une équipe de recherche. Les mesures de formaldéhyde enregistrées à partir des années 80 dans les banques de données d'exposition professionnelle (BDEP) multisectorielles française (COLCHIC) et états-unienne (IMIS) ont également été obtenues. Les sources ont été analysées par modélisation statistique puis comparées entre elles. Une méthode méta-analytique permettant l'analyse numérique des données de la littérature a été élaborée.

Les modèles statistiques ont expliqué entre 25 et 61% de la variabilité des mesures. Pour les panneaux de bois aggloméré, l'emploi, la zone de travail, le procédé, la saison, l'année et la durée de mesure étaient déterminants. Les mesures du gouvernement étaient plus élevées celles de l'équipe de recherche (facteur 2 à 3), indiquant un biais de stratégie de mesure. Les déterminants mis en évidence dans IMIS et COLCHIC incluaient l'industrie, le poste de travail, l'année (-5 à -10% par an), la ventilation, la saison, la durée et le débit d'échantillonnage associés aux mesures, ainsi que la raison de la visite, cette dernière indiquant un biais de sélection des établissements. Les niveaux mesurés sur de courtes durées étaient en moyenne 2 fois plus élevés que ceux de type valeur d'exposition moyenne pondérée (VEMP). Une corrélation significative (entre 0,2 et 0,8) des mesures prises durant une même campagne d'échantillonnage a été identifiée dans les données québécoises, COLCHIC et IMIS. Les déterminants communs aux différentes sources avaient des influences similaires sur l'exposition alors que les niveaux eux-mêmes différaient par des facteurs de 1,3 à 3. Une réduction importante des niveaux d'exposition durant les deux dernières décennies a été observée (facteur 3 à 10). Les secteurs d'activité associés aux expositions de type VEMP les plus élevées comprenaient les panneaux de bois, les travaux de charpenterie et les analyses biologiques et anatomopathologiques, avec des moyennes géométriques, estimées entre 2001 et 2003 à partir de IMIS et COLCHIC, autour de 0,2-0,3 mg/m^3.

Nos résultats ont montré que des biais sont présents dans les données préexistantes. Ils peuvent cependant être au moins partiellement contournés si l'information entourant les mesures est suffisante pour les identifier et les quantifier, et si les estimations sont corrigées en conséquence. Les portraits rétrospectifs multisectoriels de l'exposition au formaldéhyde dressés dans cette étude pourront être utilisés pour la surveillance de l'exposition et l'épidémiologie.

Mots clés : banques de données d'exposition professionnelle, modèles linéaires mixtes, méta-analyse, stratégie de mesure

ABSTRACT

Existing data on exposure levels represent a choice data source for occupational exposure surveillance or epidemiology. However, the varying circumstances in which they are measured may cause biases when assessing the exposure of workers. Aiming to better characterize these biases and establish a historical picture of occupational exposure to formaldehyde, recently classified as carcinogenic to humans by the International Agency for Research on Cancer, we analyzed and compared several sources of formaldehyde exposure data.

Five datasets of formaldehyde exposure levels were used. In the reconstituted wood panel industry, they included data from a literature review, data measured in Quebec in 1984- 2002 by governmental hygienists, and data measured in Quebec in 2001-2002 by a research team. Formaldehyde levels recorded since the 1980s in the French and U.S. occupational exposure databanks (OEDB) COLCHIC and IMIS were also obtained. The datasets were first analyzed with statistical models then compared with one another. A new meta-analytic method allowing the computational analysis of literature data was proposed.

The statistical models explained between 25 and 61% of the variance of the log-transformed exposure levels. In the wood panel industry, the variables job, work zone, process, season, year, and measurement duration were determinants of formaldehyde exposure levels. In addition, concentrations measured by governmental hygienists were two to threefold as high as those measured by the research team, suggesting a sampling strategy-related bias. Exposure determinants in the multi-industry datasets from IMIS and COLCHIC included industry, task / workstation, year (-5 to -10% per year), ventilation, season, measurement duration and sampling flow, and the reason for sampling, the latter suggesting a bias caused by differential selection of the visited plants. Short-term measurements were on average twice as high as time weighted averaged (TWA) data in IMIS and COLCHIC. Significant correlation (0.2 to 0.8) between measurements taken during the same sampling campaign was observed in the Quebec data, COLCHIC, and IMIS. Exposure determinants common to the different sources had similar influence on exposure but exposure levels differed among sources by factors from 1.3 to 3. An important decrease of formaldehyde exposure levels over time was observed during the last two decades (factor 3 to 10). Industries associated with the highest TWA exposure levels included wood panels, carpentry work, and biological and anatomopathological analyses. The corresponding geometric means, estimated from COLCHIC and IMIS for 2002-2003, were between 0.2 and 0.3 mg/m^3.

Our results confirm that biases are present in existing occupational exposure data. However they also suggest these biases can be controlled if the available ancillary information is sufficient to allow their identification and quantification, and if exposure estimates are corrected accordingly. The historical multi-industry pictures of occupational exposure to formaldehyde presented in our work can be used for exposure surveillance and occupational epidemiology.

Keywords: occupational exposure databank, linear mixed-effect models, meta-analysis, sampling strategy

TABLE DES MATIÈRES

INDEX DES TABLEAUX

CHAPITRE I : INTRODUCTION GÉNÉRALE

CHAPITRE II : ÉTUDE DES DÉTERMINANTS DE L'EXPOSITION AU FORMALDÉHYDE DANS L'INDUSTRIE DES PANNEAUX DE BOIS AGGLOMÉRÉ AU QUÉBEC

CHAPITRE III : UTILISATION DE LA SIMULATION MONTE CARLO POUR RECONSTRUIRE DES NIVEAUX D'EXPOSITION AU FORMALDÉHYDE À PARTIR DE PARAMÈTRES DE SYNTHÈSE RAPPORTÉS DANS LA LITTÉRATURE

CHAPITRE IV : MODÉLISATION STATISTIQUE DES NIVEAUX D'EXPOSITION PROFESSIONNELLE AU FORMALDÉHYDE DANS L'INDUSTRIE FRANÇAISE, 1986-2003

INDEX DES FIGURES

CHAPITRE I : INTRODUCTION GÉNÉRALE

CHAPITRE II : ÉTUDE DES DÉTERMINANTS DE L'EXPOSITION AU FORMALDÉHYDE DANS L'INDUSTRIE DES PANNEAUX DE BOIS AGGLOMÉRÉ AU QUÉBEC

LISTE DES SIGLES ET ABRÉVIATIONS

%	Pourcent
°C	Degré Celsius
ACGIH	American Conference of Governmental Industrial Hygienists
AIC	Akaike Information Criterion
AM	Arithmetic Mean
ASD	Arithmetic Standard Deviation
ATABAS	Banque de données d'exposition professionnelle danoise
atm	Atmosphère
BDEP	Banque de Données d'Exposition Professionnelle
BIC	Bayesian Information Criterion
C2	Mention cancérogène soupçonné dans le Règlement sur la santé et la sécurité au travail
CI	Confidence Interval
CIIT	Chemical Industry Institute of Toxicology
CIRC	Centre International de Recherche sur le Cancer
COLCHIC	Banque de données d'exposition professionnelle française
CSST	Commission de la Santé et de la Sécurité du Travail
DECOS	Dutch Expert Committee on Occupational Standards
DNPH	Dinitrophenylhydrazine
ECDF	Empirical Cumulative Distribution Function
EM	Mention de réduction de l'exposition au minimum dans le Règlement sur la santé et la sécurité au travail
EXPO	Banque de données d'exposition professionnelle norvégienne
g/mol	Gramme par mole
GEE	Generalized Estimating Equations
GM	Geometric Mean
GSD	Geometric Standard Deviation

h.	Heure
HCHO	Formaldéhyde (formule chimique)
IARC	International Agency for Research on Cancer
ICC	Intra-class Correlation Coefficient
IMIS	Banque de données d'exposition professionnelle États-unienne
INRS	Institut National de Recherche et de Sécurité
IRSST	Institut de Recherche Robert-Sauvé en Santé et en Sécurité du Travail
ISIC	International Standard Industrial Classification
L/min.	Litre par minute
LEV	Local Exhaust Ventilation
LOD	Limit of Detection
M	Median
MDF	Medium Density Fibreboard
MEGA	Banque de données d'exposition professionnelle allemande
mg/m3	Milligramme par mètre cube
min.	Minute
ML	Maximum Likelihood
MUF	Melamine-Urea-Formaldéhyde
NACE	Classification d'activités économiques dans la Communauté européenne
NAF	Nomenclature d'Activité Française
NAICS	North American Industrial Classification System
ND	Not detected
NEDB	Banque de données d'exposition professionnelle du Royaume-Uni
NIOSH	National Institute for Occupational Health and Safety
OEDB	Occupational Exposure Databank
OEL	Occupational Exposure Limit
OSB	Oriented-Strand Board
OSHA	Occupational Safety and Health Administration

PB	Particle Board
PEL	Permissible Exposure Limit
PF	Phenol-Formaldehyde
ppm	Partie par millions
R	Range
REML	Restricted Maximum Likelihood
RIE	Relative Index of Exposure
RP	Mention d'interdiction de recirculation de l'air dans le Règlement sur la santé et la sécurité au travail
RSD	Relative Standard Deviation
RSST	Règlement sur la Santé et la Sécurité du Travail
SIC	Standard Industrial Classification
SMEST	Banque de données d'exposition professionnelle québécoise
STEL	Short Term Exposure Limit
TLV	Threshold Limit Value
TWA	Time Weighted Average
U.S. EPA	United States Environmental Protection Agency
UF	Urea-Formaldehyde
VEMP	Valeur d'Exposition Moyenne Pondérée
VLE	Valeur Limite d'Exposition

REMERCIEMENTS

Je tiens à remercier en premier lieu le Dr. Michel Gérin, qui m'a aiguillé vers le domaine de la santé au travail et m'a encouragé et soutenu depuis mon arrivée au Québec et au Département de santé environnementale et santé au travail. Je dois beaucoup de mon curriculum vitæ à son acharnement à me faire participer durant ces années à de multiples projets scientifiques, d'enseignement ou à des conférences internationales.

Je remercie également le Dr Guy Perrault pour ses commentaires toujours encourageants et sa disponibilité.

Merci du fond du cœur à Denis Bégin, qui m'a appris qu'il est toujours possible de faire mieux, de manière plus élégante, et dans un meilleur Français si on consent à faire un effort…

Merci aussi à Charles Beaudry, hygiéniste errant, qui a toujours su remettre sur terre mes pieds d'étudiant idéaliste.

Finalement merci à tous ceux avec lesquels j'ai pu respirer un peu d'air frai pour mieux replonger dans les équations et les concepts : Jean-Claude et Marie-France, mes parents, Sylvia, ma douce (et grande) moitié, Annie, Ross, Asta, Nolwenn, et les copains de Blue Beat Gangster et des Encrystés.

"Ne mesure jamais une deuxième fois l'exposition d'un travailleur,

tu trouverais un résultat différent"

Hygiéniste anonyme.

CHAPITRE I

INTRODUCTION GÉNÉRALE

1. INTRODUCTION GÉNÉRALE

1.1 Mise en contexte

En santé au travail la connaissance a priori des conditions d'exposition des travailleurs représente un atout majeur à plusieurs points de vue. Tout d'abord, au sein de la démarche d'hygiène industrielle, l'étape primordiale d'identification des agents agresseurs est facilitée par la possibilité d'anticiper la présence de ces agents dans le milieu de travail à l'étude (Bégin et coll., 1995). Ensuite, dans un cadre plus large de surveillance en santé au travail, l'existence d'un portrait à l'échelle nationale de l'exposition des travailleurs à un vaste éventail de substances ou de dangers physiques permet de mieux définir les priorités d'intervention et de gestion des risques (Goldman et coll., 1992;Gomez, 1993;LaMontagne et coll., 2002a). De la même façon cette vision globale facilite les prises de décision en vue de l'établissement de normes en santé et sécurité du travail (Botkin et Conway, 1995). Enfin, dans un contexte de recherche scientifique, la possibilité d'associer un type d'emploi ou de tâche professionnelle à une exposition quantifiée à des agents spécifiques représente un outil de choix lors de la réalisation d'études de type épidémiologique ou d'analyse de risque toxicologique, en particulier lors de l'établissement de matrices emploi-exposition (Goldberg et coll., 1993;Kauppinen et coll., 1998;Stewart et Rice, 1990). À cause des coûts importants associés à la mesure directe de l'exposition professionnelle d'une population, et de l'impossibilité de mesurer directement l'exposition passée, les données d'exposition préexistantes constituent souvent la seule source d'information disponible pour atteindre ces objectifs.

La littérature scientifique d'hygiène industrielle et les données rendues disponibles par des organismes gouvernementaux ou paragouvernementaux, collectées dans le cadre d'activités de prévention ou de contrôle du respect des normes, constituent les principales sources de données d'exposition préexistantes accessibles au public. Les monographies du Centre international de recherche sur le cancer (CIRC) illustrent l'utilisation de la littérature scientifique pour établir des bilans globaux d'exposition (voir par exemple (CIRC, 1995)). Les données collectées par les organismes d'État peuvent se présenter sous plusieurs formes. Par exemple, le National Institute for Ocupational Safety and Health (NIOSH), aux États-Unis, a rendu disponibles en ligne les rapports complets d'évaluation de visites industrielles réalisées par les équipes de cet institut. Dans plusieurs pays, les données d'exposition sont stockées sous forme de banques de données informatisées, appelées banques de données d'exposition professionnelle (BDEP), dans lesquelles les concentrations mesurées sont associées à un certain nombre de variables les caractérisant (par exemple : unité de mesure, méthode d'analyse, emploi évalué). On peut citer notamment les bases COLCHIC (Carton et Goberville, 1989) en France, NEDB (Burns et Beaumont, 1989) au Royaume-Uni, MEGA (Stamm, 2000) en Allemagne, IMIS (Stewart et Rice, 1990) aux États-Unis et SMEST[1] au Québec.

[1] http://www.sogique.qc.ca/sysinfo/reponse.asp?pagew=91

Le principal obstacle relié à l'utilisation de données préexistantes vient du fait qu'elles n'ont en général pas été collectées pour représenter la population d'intérêt. Il devient donc critique de connaître dans quelle mesure les données disponibles reflètent adéquatement ou non cette population. La problématique peut être abordée selon deux avenues : la recherche de consensus sur le choix et la standardisation des informations devant accompagner des données d'exposition, et la conduite d'études visant spécifiquement à mettre en évidence la présence ou l'absence des biais dans les données (Olsen et Jensen, 1994;Ulfvarson, 1983).

Suite à une demande de la Commission de la santé et de la sécurité du travail (CSST), une recherche a été initiée en 2001 à l'Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) pour étudier l'impact d'un abaissement de la norme d'exposition au formaldéhyde en milieu de travail au Québec (Goyer et coll., 2004). L'établissement d'un portrait provincial de l'exposition au formaldéhyde constituait un volet majeur de cette recherche. Dans le cadre de cette étude, cinq ensembles de données de mesures d'exposition au formaldéhyde ont été rendues disponibles: dans le secteur de la fabrication des panneaux de bois aggloméré, secteur utilisant la plus grande quantité de formaldéhyde au Québec, une campagne de mesures dans les 12 usines québécoises de ce secteur a été menée en 2001-2002, les données prises historiquement dans ces usines par les équipes d'hygiène des Centres locaux de services communautaires (CLSC) depuis 1984 ont été collectées, et une revue exhaustive de la littérature sur l'exposition au formaldéhyde a été effectuée. Les procédés inclus dans la revue de littérature et les données québécoises comprenaient les panneaux de particule (PB, « Particle board »), les panneaux de fibre (MDF, « Medium density fibreboard ») et les panneaux OSB (« Oriented strand board »). De plus, les données multisectorielles d'exposition au formaldéhyde contenues dans les BDEP française (COLCHIC) entre 1986 et 2003 et états-unienne (IMIS, « Integrated Information Management System ») entre 1979 et 2001 ont pu être obtenues auprès des organismes responsables. Ces cinq sources de données constituaient un terrain privilégié sur lequel élaborer un projet de recherche visant à étudier l'utilisation des sources de données préexistantes pour évaluer l'exposition et à améliorer les connaissances sur l'exposition professionnelle au formaldéhyde.

1.2 Les problématiques reliées à l'utilisation des sources de données préexistantes

Bien que les problématiques des biais et de la qualité des informations entourant les données d'exposition touchent tant l'utilisation de la littérature que l'exploitation des BDEP, elles ont été abordées dans la littérature scientifique de façon distincte pour chacune des deux sources. La présente section est structurée en conséquence.

1.2.1 Utilisation des BDEP pour l'établissement de portraits d'exposition

Utilisations des BDEP rapportées dans la littérature

Les études publiées sur l'utilisation de BDEP ont présenté en général des portraits multisectoriels de l'exposition à un contaminant particulier (4 études portant sur le plomb, 3 sur la silice, 1 sur le formaldéhyde et 1 sur les poussières de bois) (Carton, 1995;Carton et Jeandel, 1993;Coble et coll., 2001;Freeman et Grossman, 1995;Froines et coll., 1990;Froines et coll., 1986;Oudiz et coll., 1983;Vincent et Jeandel, 2002). Les résultats ont été principalement présentés sous forme de moyenne ou de médiane des valeurs par secteur industriel. Teschke et coll. (1999) ont analysé les données d'IMIS sur les poussières de bois au moyen de modèles linéaires mixtes. L'emploi de cette technique a permis d'identifier les variables influentes sur les niveaux d'exposition et de tenir compte de la corrélation des mesures prises lors d'une même visite.

On retrouve également dans la littérature quelques descriptions d'utilisations plus spécifiques des BDEP, notamment en matière de surveillance et d'analyse de risque. Ainsi, Valiante et coll. ont utilisé les niveaux d'exposition à la silice mesurés entre 1980 et 1992 dans l'État du New Jersey et rapportés dans IMIS pour l'identification de milieux de travail à risque pour la silicose (Valiante et coll., 1992). Linch et coll. ont utilisé IMIS pour estimer la proportion des employés exposés à la silice à différentes fractions d'une valeur limite d'exposition (VLE) dans plusieurs secteurs industriels (Linch et coll., 1998). La banque de donnée britannique NEDB a été utilisée au début des années 90 pour développer un modèle d'expert capable de prédire une fourchette d'exposition en fonction de paramètres reliés à la volatilité et à des caractéristiques générales de l'utilisation d'une substance (Bredendiek-Kämper, 2001).

Études sur les informations entourant les mesures dans les BDEP

Bien que plusieurs auteurs ou organisations aient émis des recommandations concernant le degré d'information devant accompagner les mesures d'exposition, celles issues de deux groupes de travail en Europe et aux États-Unis formés au milieu des années 90 représentent sans doute l'effort le plus systématique dans ce domaine (Joint ACGIH-

AIHA Task Group on Occupational Exposure Databases, 1996;Rajan et coll., 1997). L'expérience québécoise avait d'ailleurs contribué à l'élaboration de ces recommandations (Rajan et coll., 1997). Le tableau N°1 présente les grandes catégories d'informations jugées importantes par le groupe européen ainsi qu'une brève description de leur contenu. Les recommandations états-uniennes sont essentiellement similaires.

Tableau N°1 : Recommandations sur les informations devant accompagner les mesures dans les BDEP (Rajan et coll., 1997)

Appellation de la catégorie	**Types d'informations inclus**
Situation	Description de l'entreprise (activité économique, taille, adresse)
Aire de travail	Description du lieu de mesure (département, poste de travail, procédé utilisé)
Activité de l'employé	Description de la profession et des tâches réalisées par l'employé évalué
Produit	Identification du produit / matière première / intermédiaire de production qui est la source d'exposition
Agent chimique	Description de l'agent chimique mesuré (Numéro CAS, noms commerciaux)
Éléments influents sur l'exposition	Informations sur des paramètres déterminants comme la structure de l'exposition (continue, intermittente), les moyens de maîtrise (ventilation locale / générale), ou encore les équipements de protection personnels.
Stratégie de mesure	Information sur les objectifs de la mesure (évaluation du pire des cas pour la conformité aux normes, établissement d'un portrait représentatif, évaluation de moyens de maîtrise de l'exposition)
Procédure de mesurage	Information sur la méthode d'échantillonnage et d'analyse (méthode analytique, durée d'échantillonnage par rapport à la durée de l'exposition, date de l'échantillonnage)
Résultats	Valeur numérique de la mesure
Référence	Référence du rapport d'hygiène correspondant à la mesure

Études sur les biais présents dans les BDEP

Chaque étape conduisant à la mesure puis à l'enregistrement d'un niveau d'exposition dans une BDEP est susceptible d'introduire des biais dans l'interprétation de cette mesure par rapport à la population générale (Olsen et coll., 1991). Ce principe peut être conceptualisé de manière simple et est illustré dans la figure 1. Certains des biais mentionnés dans cette figure peuvent être éliminés si les informations entourant les mesures sont adéquates. Par exemple, si le titre d'emploi est disponible, le biais potentiellement causé par la concentration des mesures dans les emplois les plus exposés peut être évité puisque l'utilisateur de la BDEP connaît alors la population représentée par les mesures. D'autres biais, en revanche, sont plus difficile à identifier et à contourner, en particulier ceux reliés aux stratégies de sélection des employés et des périodes mesurées. Ainsi, dans le but d'optimiser l'utilisation des ressources disponibles pour les mesures, il semble pratique courante en hygiène de concentrer ces dernières sur des périodes ou des circonstances associées à de fortes expositions (Kromhout, 2002). Si l'on peut démontrer que les « pires circonstances » (stratégie « worst case ») sont acceptables vis-à-vis d'une VLE, on peut alors conclure que le milieu de travail dans son ensemble correspond à des conditions d'exposition acceptables. Bien que cette méthode permette dans certains cas la maîtrise des conditions d'exposition dans un milieu de travail, elle ne fournira pas, globalement, une bonne image de l'exposition « moyenne » dans ce milieu de travail. De plus, même si une mesure est identifiée comme ayant été prise dans les « pires conditions », l'ampleur de la différence entre ces « pire conditions » et les conditions normales d'exposition est difficile à évaluer et varie avec le niveau d'expertise du mesureur. Il devient alors nécessaire d'essayer de qualifier et de quantifier ce type de biais à partir d'études plus spécifiques (Olsen et coll., 1997).

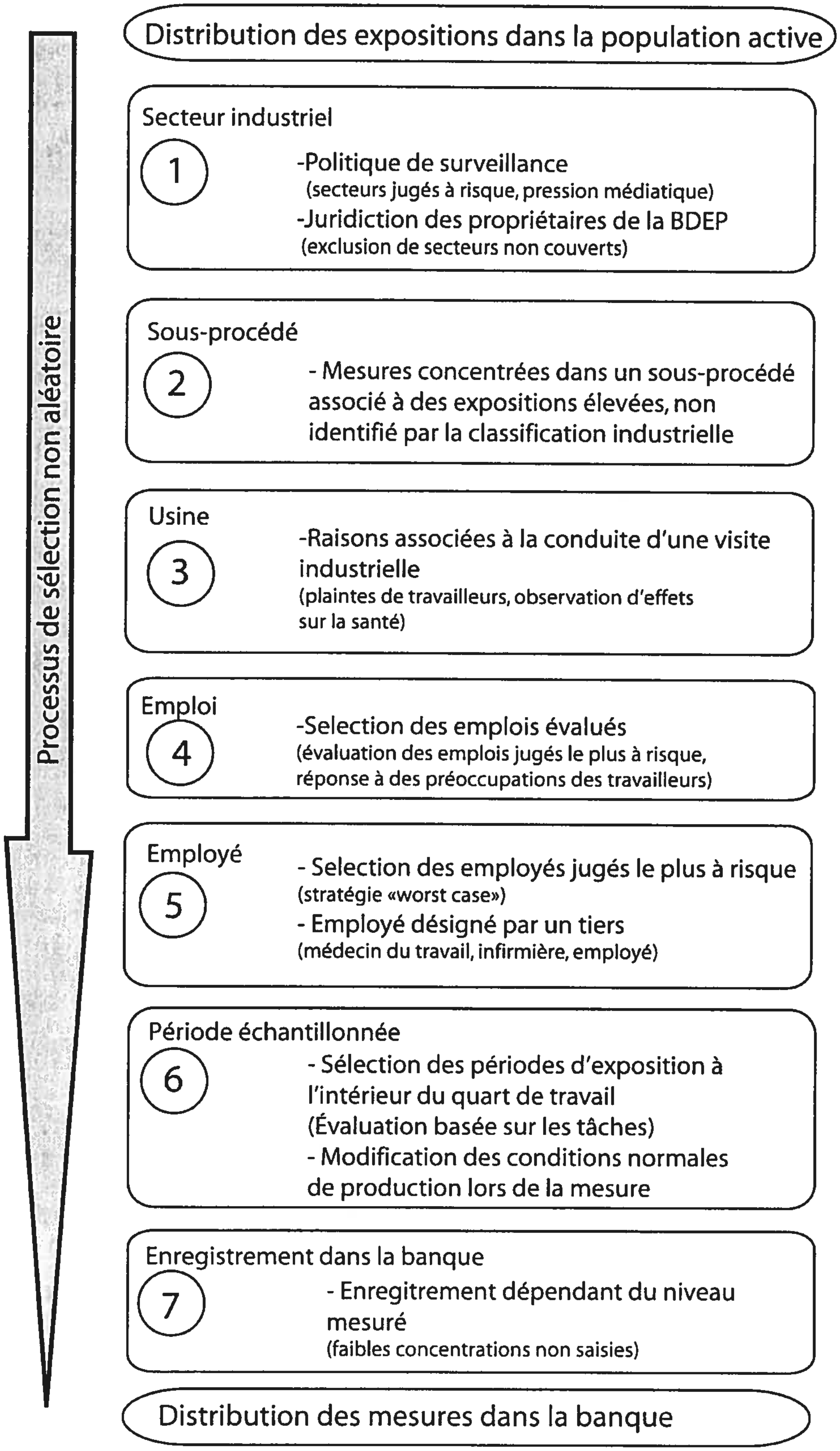

Figure 1 : Schéma conceptuel des biais dans une banque de données d'exposition professionnelle.

L'étude des biais peut être abordée selon deux angles : l'influence de variables internes sur les niveaux, et la comparaison avec des sources de données externes. Ainsi, les BDEP contiennent souvent plusieurs variables descriptives du milieu de travail visité (par exemple taille de l'établissement, statut syndical). S'il s'avère que ces variables ont une influence sur les niveaux d'exposition, un biais sera causé si la distribution des valeurs de ces variables est différente dans la banque de données et dans la population à propos de laquelle on désire tirer des conclusions (par exemple sur- ou sous-représentation des grandes entreprises). De plus, certaines variables internes sont indicatrices d'approches différentes dans la sélection des conditions d'exposition évaluées (par exemple évaluation causée par une plainte / évaluation reliée à un plan de prévention). Encore une fois la mise en évidence d'une influence de telles variables sur les niveaux pourra être indicatrice d'un biais et permettre de le quantifier. Finalement, la comparaison avec des sources externes permet de mesurer les écarts systématiques entre différentes stratégies globales d'évaluation.

Les sept études qui ont été retrouvées dans la littérature portant, au moins partiellement, sur des biais identifiables par l'étude de paramètres internes aux BDEP concernaient la banque IMIS. Les aspects de ces études reliés aux biais sont présentés au tableau N°2. Deux études ont porté spécifiquement sur l'influence des raisons des visites d'entreprise (incluant les visites planifiées dans les usines de secteurs d'activité jugés à risque, ou encore causées par des plaintes d'employé ou la recommandation d'un inspecteur en sécurité du travail), de la portée de ces visites (visite complète de l'établissement par opposition à visite ciblée sur un département spécifique), et du statut syndical de l'établissement visité (Gomez, 1997;Melville et Lippmann, 2001). Gomez, à partir des données de IMIS dans les secteurs présentés au tableau N°2, a formé 9 groupes d'exposition constitués de mesures prises sur des employés effectuant approximativement les mêmes tâches dans les mêmes secteurs industriels dans de multiples établissements. Les groupes formés contenaient entre 87 et 500 mesures de type valeur d'exposition moyenne pondérée (VEMP). L'auteur a utilisé quatre approches de modélisation statistique pour étudier l'influence des variables mentionnées au tableau N°2 pour chaque groupe d'exposition. Pour les groupes du plomb, les modèles utilisés ont expliqué entre 22 et 34% de la variance alors que ces chiffres ne dépassaient pas 9% dans les autres groupes. Melville et Lippman ont utilisé une stratégie similaire à Gomez pour étudier l'influence des même variables sur les mesures issues de IMIS pour trois autres groupes d'exposition (voir tableau N°2) (Melville et Lippmann, 2001). Aucun des modèles n'a expliqué plus de 20% de la variance. Les résultats obtenus (cf. tableau N°2) sont très dépendants des groupes d'exposition et sont dans certains cas en contradiction avec les observations de Gomez.

Tableau N°2 : Littérature concernant l'influence de variables internes aux banques de données d'exposition professionnelles

Référence	Secteur industriel	Agresseur	Fenêtre temporelle	Objet de la validation	Paramètre étudié	Méthode	Résultats
Oudiz et coll., 1983	Fonderies	Silice	1976-1981[a]	Nombre d'employés	% de mesures non-conformes	Calcul des pourcentages pour 3 catégories : <100, [100-500] et >500 employés	Augmentation de 10% entre les catégories 1 et 2, et 2 et 3.
Froines et coll., 1986	Étude multi-sectorielle	Silice	1979-1986	Statut syndical, type d'inspection[b]	Sévérité[c] médiane par code industriel	Calcul des sévérités moyennes et médianes stratifiées par niveaux de variable pour les codes industriels ayant un nombre minimal d'inspection.	Influence variable suivant les secteurs industriels. Tendance approximative à l'augmentation de la sévérité moyenne et médiane pour les inspections causées par des plaintes de compagnies syndiquées
Stewart et Rice, 1990	Étude multisectorielle	Silice	1979-1986	Type d'inspection[b]	Moyenne arithmétique des mesures	Calcul des moyennes par type d'inspection (plainte / planifiée) pour dix secteurs industriels ayant au minimum 5 mesures pour chaque type d'inspection.	Rapport médian entre les moyennes arithmétiques plainte/planifié : 2,4 (3 sont inférieurs à 1 sur les 10 résultats disponibles)
Froines et coll., 1990	6 secteurs industriels prio-ritaires établis par OSHA	Plomb	1979-1985	Nombre d'employés, statut syndical, type d'inspection[b]	Sévérité médiane par inspection	Régression logistique : probabilité d'une sévérité médiane supérieure à 1 en fonction du statut syndical (dichotomique), du nombre d'employés (entier), et du type d'inspection (plainte/planifié)	Rapport de cote estimé de 3 [1,4-6] pour les inspections de type plainte. Pas d'autre variable significative (les interactions de premier ordre ont été testées).

Référence	Secteur industriel	Agresseur	Fenêtre temporelle	Objet de la validation	Paramètre étudié	Méthode	Résultats
Gomez, 1997	Plomb dans l'industrie des batteries Perchloroéthylène chez les nettoyeurs à sec Oxyde de fer chez les soudeurs dans trois industries		1979-1989	Nombre d'employés, statut syndical, type d'inspection, année de mesure, portée de l'étude[d]	Pour chaque groupe d'exposition : -modélisation linéaire simple des moyennes par établissement -modélisation linéaire simple des mesures individuelles -modélisation linéaire simple des mesures individuelles avec prise en compte de la corrélation potentielle entre les mesures prises dans la même inspection (méthodologie équations d'estimation généralisées) -modélisation logistique de la probabilité pour une mesure d'être supérieure au 75[e] percentile des mesures de son groupe d'exposition		Tendance claire observée seulement pour les variables année (réduction dans le temps) et taille (réduction de l'exposition lorsque la taille augmente)
Teschke et coll., 1999	Étude multi-sectorielle	Poussières de bois	1979-1997	Nombre d'employés, type d'inspection, année de mesure	Mesures individuelles de type VEMP	Régression multiple pour les variables d'intérêt et équations d'estimation généralisées pour tenir compte de la corrélation des mesures à l'intérieur d'une inspection. Création de 82 catégories d'emplois standardisés.	Variables nombre d'employés et type d'inspection non significatives. Le modèle, qui incluait les variables année, état, code industriel et emploi, a expliqué 37% de la variabilité. La corrélation intra-inspection a été estimée à 0,31
Melville et Lippmann, 2001	Désamiantage Toluène dans les ateliers de carrosserie Formaldéhyde chez les embaumeurs		1979-1997	Nombre d'employés, statut syndical, type d'inspection, année de mesure, portée de l'étude[d]	Pour chaque groupe d'exposition : -modélisation linéaire simple des moyennes par établissement -modélisation linéaire simple des moyennes par établissement pondérées par les variances correspondantes -modélisation linéaire simple des mesures individuelles		Tendance générale à des expositions plus élevées pour les inspections de portée globale. Pour le toluène et le formaldéhyde les données issues de plaintes étaient supérieures à celle issues d'inspections planifiées.

a : dans cette étude les auteurs ont employé non seulement IMIS (qui débute en 1979) mais aussi les rapports papier des inspecteurs de l'Occupational Safety and Health Administration (OSHA)

b : une variable dans la banque IMIS permet de distinguer plusieurs raisons ayant donné lieu à la visite. On distingue en général les inspections causées par des plaintes, les inspections planifiées, et les inspections de suivi recommandées par des intervenants en sécurité du travail.

c : la sévérité est le ratio de la valeur mesurée sur la valeur limite d'exposition en vigueur au moment de la mesure

d : cette variable dans la banque IMIS précise si la visite portait sur tout l'établissement ou sur une zone à risque en particulier.

Seules deux études portant sur la comparaison de BDEP avec des données externes ont été retrouvées dans la littérature. Olsen et coll. ont comparé les résultats de trois ensembles de mesures d'exposition aux solvants dans l'industrie du meuble au Danemark (Olsen et coll., 1991). Le premier groupe (groupe A, 453 mesures) était constitué des mesures enregistrées à partir de 1982 dans la base danoise ATABAS. Le second groupe (groupe B, 42 mesures) provenait de mesures de courte durée (<20 min.) prises dans la zone respiratoire de travailleurs au moment d'une tâche impliquant la manipulation de solvants, dans un échantillon aléatoire d'usines danoises. L'étude couvrait la période 1985-1986. Le dernier groupe de mesures (groupe C, 240 mesures) a été constitué durant l'année 1988 lors de l'étude d'un échantillon aléatoire de 200 compagnies du secteur industriel du meuble au Danemark. Dans chaque compagnie, 4 mesures couvrant 1 heure et réparties de façon aléatoire dans la journée ont été prises dans la zone respiratoire de travailleurs choisis au hasard parmi ceux travaillant dans des locaux où des solvants étaient présents. Après répartition des mesures dans 3 catégories, la majorité des mesures du groupe C se retrouvaient dans les catégories « faible » ou « moyen » alors que les mesures des deux autres groupes étaient principalement réparties entre les catégories « moyen » et « élevé ». Les auteurs ont également présenté les courbes cumulatives empiriques des mesures d'exposition au toluène provenant des trois groupes. Alors que les courbes des groupes A et B étaient approximativement confondues, la courbe du groupe C était décalée vers des valeurs plus faibles d'un facteur 5 à 10. Le groupe C correspondant à un design aléatoire, ces résultats démontrent selon les auteurs les biais introduits par la stratégie d'échantillonnage dans les groupes A et B. Le groupe B correspond au choix déterministe de périodes d'exposition dans une journée. Pour le groupe A (banque ATABAS), le biais observé peut avoir la même cause que le groupe B, ou encore être dû au choix déterministe de travailleurs, de secteurs de l'entreprise, ou même d'entreprises pour lesquelles l'exposition est supposée importante. La comparaison présentée par Olsen n'a pas pris en compte les différences potentielles entre corps de métiers, départements ou encore une possible diminution des niveaux d'exposition au cours des années.

Vinzents et coll. ont comparé les mesures de xylène dans 5 BDEP nationales européennes : ATABAS (Danemark), MEGA (Allemagne), COLCHIC (France), EXPO (Norvège) et NEDB (Royaume-Uni) (Vinzents et coll., 1995). L'étude a été limitée aux secteurs industriels suivants : travail du bois, peinture par pulvérisation, et peinture par pulvérisation à l'intérieur du travail du bois. Les mesures couvraient approximativement la période 1985-1993. L'analyse s'est limitée à la comparaison des médianes et écart-types géométriques des mesures des BDEPS pour chaque secteur d'activité. Vinzents et coll. ont remarqué des médianes plus faibles pour les bases allemandes et françaises (différence maximale d'un facteur 4), qu'ils attribuent en partie au fait que ces dernières sont constituées de données prises dans le cadre d'assurances plutôt que de conformité à des normes environnementales. Ils ont également observé des écarts-types géométriques élevés (entre 5 et 7), qui seraient expliqués par la définition large des secteurs d'activité.

Globalement, il apparaît que peu d'études se sont attachées à étudier de façon systématique les biais présents dans les BDEP, et que la majorité de celles qui y ont fait référence, la plupart à propos de IMIS, ont utilisé des paramètres descriptifs vulnérables à la confusion par des variables non prises en compte dans l'analyse. Les études ayant utilisé des méthodes plus raffinées, quant à elles, n'ont pas permis en général de conclure sur la présence systématique de biais dans les BDEP analysées, en particulier ceux reliés à la sélection des entreprises et aux stratégies de mesure.

1.2.2 Utilisation de la littérature pour l'établissement de portraits de l'exposition

Plusieurs articles ont récemment déploré l'absence d'un cadre méthodologique rigoureux pour l'utilisation de la littérature dans l'évaluation de l'exposition. Comme Caldwell et coll. ou encore Marquart et coll. le soulignent, de nombreuses études manquent de détails concernant les déterminants de l'exposition (p.ex. procédé, ventilation, tâches) et les caractéristiques de la population ou de la situation évaluée (taille de l'échantillon, échantillons répétés ou pris sur plusieurs postes / travailleurs, mesures court-terme / sur un quart de travail complet). De plus, les paramètres de synthèse des mesures sont rarement rapportés de façon adéquate et diffèrent selon les articles (moyenne arithmétique, moyenne géométrique, fourchette...)(Caldwell et coll., 2001;Marquart et coll., 2001). Ce manque d'information entourant les données d'exposition compromet l'évaluation de la représentativité d'une étude vis-à-vis la situation d'intérêt et augmente ainsi le risque d'introduction de biais. Tielemans et coll., ainsi que Money et Margary ont récemment proposé des lignes directrices pour rationaliser l'utilisation de la littérature, principalement par l'établissement de critères objectifs de qualité des articles et par une pondération de l'influence de chaque article sur l'évaluation finale en fonction de sa qualité (Money et Margary, 2002;Tielemans et coll., 2002). Les auteurs recommandent notamment la documentation systématique d'un certain nombre d'informations, très similaires à celles présentées au tableau N°1. Tielemans et coll. proposent l'évaluation de la qualité d'une étude sur quatres niveaux: un premier niveau concernant la documentation des déterminants de l'exposition, un second niveau se rapportant à la qualité des descripteurs statistiques utilisés, un troisième niveau à propos de la validité interne, et un dernier niveau sur la validité externe. Les auteurs soulignent que le premier niveau d'évaluation est critique puisqu'il détermine la capacité du lecteur à évaluer les trois autres niveaux, et recommandent l'exclusion des études qui ne respectent pas des critères minimaux de qualité pour chacun des niveaux d'information présentés. Money et coll. considèrent plutôt qu'aucune information ne doit être exclue, mais que l'on devrait pondérer les données en fonction de la qualité des études.

Les cadres méthodologiques proposés par Tielemans et coll. et Money et Margary représentent des outils de choix pour l'évaluation de la qualité d'une étude, l'exclusion d'études non pertinentes, ou encore la présentation d'une étude d'hygiène industrielle. Cependant, ils ne fournissent pas de méthodologie permettant l'intégration des résultats issus de multiples études rapportant des paramètres différents. Caldwell et coll. ont résumé la littérature portant sur l'exposition aux hydrocarbures en calculant pour chaque substance une moyenne de moyennes pondérées par la taille des échantillons. Cependant leur étude a été limitée par la disponibilité de ces paramètres dans les articles retrouvés (Caldwell et coll., 2000). Ainsi les études présentant des revues de littérature sur les niveaux d'exposition se limitent-elles le plus souvent à la présentation de tableaux de synthèse (voir par exemple (Burstyn et coll., 2000) ou encore les monographies du CIRC) et à une évaluation globale par expertise, qui devient d'autant plus complexe lorsque de nombreux articles sont disponibles. Il existe donc un besoin de développement de méthodes de type méta-analyse pour l'exploitation des données d'exposition issues de la littérature.

1.3 Le formaldéhyde

1.3.1 Propriétés physicochimiques et utilisations industrielles

Aux conditions normales de température et de pression, le formaldéhyde est un gaz. Incolore, il possède une odeur piquante et suffocante. Le tableau N°3 présente quelques caractéristiques physicochimiques ainsi que des propriétés pertinentes en hygiène industrielle.

Tableau N°3 : Principales caractéristiques physicochimiques et de sécurité du travail du formaldéhyde (Répertoire toxicologique de la CSST, 2005a;Répertoire toxicologique de la CSST, 2005b)

Formule chimique brute	HCHO
Numéro CAS*	50-00-0
Point d'ébullition	- 19°C (1 atm)
Température d'auto-ignition	424°C
Limites d'explosivité dans l'air	7% - 73%
Limite de détection olfactive	0,05 ppm - 1,00 ppm
Concentration posant un danger immédiat pour la vie	20 ppm (24,6 mg/m^3)
Masse molaire	30,03 g/mol
Points d'éclairs des solutions aqueuses à 37% de formaldéhyde Sans méthanol 15% méthanol	 83°C (creuset fermé) 50°C (creuset fermé)
Facteurs de conversion des concentrations dans l'air (20°C, 1 atm)	1 ppm = 1,23 mg/m^3 1 mg/m^3 = 0,81 ppm

* Chemical Abstracts Service Registry Number (American Chemical Society)

Le formaldéhyde est soluble dans l'eau et les solvants polaires comme les alcools et les éthers. C'est un composé très réactif, réagissant vivement avec les composés oxydants. Stable thermiquement, il ne se décompose de façon significative qu'au-dessus de 300°C, formant du dioxyde de carbone et de l'eau (Walker, 1975).

Le formaldéhyde est majoritairement produit par oxydation catalytique du méthanol. Il est principalement stocké et vendu sous forme de solutions aqueuses de concentrations variant généralement entre 30% et 56% en masse. Ces solutions contiennent des proportions variables de méthanol (jusqu'à 36%), ajouté pour inhiber la polymérisation spontanée du formaldéhyde. Les solutions les plus concentrées ne contiennent en général pas d'inhibiteurs mais sont conservées et transportées à température élevée (~60°C) (Répertoire toxicologique de la CSST, 2005a;Répertoire toxicologique de la CSST, 2005b;Walker, 1975).

Le formaldéhyde est utilisé le plus largement dans la production de résines à base d'urée, de phénol, et de mélamine. Ces résines servent d'adhésif dans la production de panneaux de bois aggloméré, de contreplaqué, de meubles et d'autres produits du bois. On les retrouve également dans les matières à mouler thermodurcissables et comme matières premières pour le revêtement de surface et les fertilisants à relarguage contrôlé. Elles sont aussi utilisées dans les industries du textile et du cuir. Les résines aminées et phénoliques sont également utilisées comme liant pour les moules de fonderie en sable, la laine de verre, le papier abrasif, et les garnitures de frein. Le formaldéhyde est aussi largement utilisé comme intermédiaire pour la synthèse d'autres produits chimiques, dans la production de polyuréthane, de polyesters, de lubrifiants, et de plastifiants. La production des matières plastiques de type polyacétals constitue une autre de ses utilisations industrielles. Finalement, le formaldéhyde est aussi utilisé pour ses propriétés bactéricides dans de nombreuses formulations de produits désinfectants, de liquides d'embaumement et de solutions de conservation de tissus biologiques (CIRC, 1995;Pichard, 2005).

1.3.2 Propriétés toxicologiques et exposition professionnelle

Le formaldéhyde est une substance puissamment irritante pour les voies respiratoires supérieures, les yeux, et la peau. Des seuils subjectifs pour l'irritation des voies respiratoires supérieures et des muqueuses oculaires sont rapportés respectivement à partir de 0,15 et 0,40 mg/m^3 de formaldéhyde dans l'air (Lauwerys, 1999). Les symptômes de toux, de gêne respiratoire et d'éternuement sont observés à des niveaux légèrement plus élevés (>0,6 mg/m^3) alors que l'exposition à des concentrations entre 2,5 et 3,5 mg/m^3 cause l'irritation aigue du nez, de la gorge et des yeux (Morandi et Maberti, 2001).

Le formaldéhyde est également reconnu comme une substance allergisante pour la peau, pouvant provoquer des dermites de contact allergiques. Il est également considéré comme un allergisant respiratoire pouvant causer de l'asthme mais le risque associé semble faible (Leroyer et Dewitte, 1999).

Le formaldéhyde a récemment été classé comme cancérogène pour l'humain (groupe 1) par le CIRC, sur la base de preuves considérées suffisantes qu'il cause le cancer du rhinopharynx chez l'humain. Le groupe de travail du CIRC a également conclu qu'il y avait des preuves limitées pour le cancer sino-nasal chez l'humain et des présomptions fortes mais insuffisantes pour établir une relation causale entre l'exposition professionnelle au formaldéhyde et la leucémie (CIRC, Sous presse). Le formaldéhyde a induit le cancer des cavités nasales dans des études chez les rats et a été démontré génotoxique dans des modèles *in vitro*, animaux et humains. Dans son document critère sur le formaldéhyde, le comité formé par le Nordic Expert Group for Criteria Documentation on Health Risks from Chemicals et le Dutch Expert Committee on Occupational Standards (DECOS) soutient que les données mécanistiques sur la toxicité et la cancérogénicité du formaldéhyde suggèrent que des concentrations non cytotoxiques de formaldéhyde ne peuvent causer le cancer. Les experts concluent que l'effet critique du formaldéhyde est l'irritation sensorielle, pour laquelle ils considèrent la concentration 0,3 mg/m^3 comme un seuil au-delà duquel l'irritation sensorielle apparaît chez l'humain dans une proportion faible mais significative de la population (Wibowo, 2003). Plusieurs auteurs ont également proposé des modèles d'analyse du risque cancérogène du formaldéhyde. Dès 1987, l'U.S. Environmental Protection Agency (U.S. EPA) a estimé par extrapolation à partir d'études chez l'animal le risque cancérogène posé par le formaldéhyde pour une exposition vie-entière. Le modèle a été mis en jour en 1991 en utilisant des données chez le singe et en prenant en compte des données relatives aux liaisons ADN-formaldéhyde / formaldéhyde-protéines (« DNA-protein crosslink »). Le risque unitaire était estimé à $2{,}7.10^{-4}$ $(mg/m^3)^{-1}$ (U.S. Environmental Protection Agency, 1991). Pour une exposition à 1,2 mg/m^3 (soit 1 ppm, une valeur utilisée comme limite d'exposition en milieu de travail), cela correspond à un risque de cancer de 2,7 pour 10 000 pour une exposition vie-entière. En 1999 le Chemical Industry Institute of Toxicology (CIIT) a proposé des estimations basées sur des études animales et utilisant un modèle de cancérogenèse qui inclut l'effet mutagène du formaldéhyde et la prolifération des cellules causée par sa cytotoxicité (CIIT, 1999). Pour une exposition durant la vie active (40 ans, 5 jours par semaine, 8 heures par jour) à une concentration de 1,2 mg/m^3, le risque était estimé respectivement à 0,09 pour 10 000, 1,5 pour 10 000 et 2,1 pour 10 000 chez des populations de non fumeurs, mixtes, et de fumeurs. Pour une exposition à une concentration de 0,6 mg/m^3 (0,5 ppm), le risque était estimé respectivement à 0,00025 pour 10 000, 0,005 pour 10 000 et 0,007 pour 10 000. Conolly et coll. ont récemment présenté une mise à jour du modèle du CIIT qui intègre une modélisation complète du système respiratoire chez l'humain (Conolly et coll., 2004). Les scénarios les plus pessimistes présentés dans ces travaux, pour lesquels plusieurs modèles de régénération cellulaire ont été testés, indiquent un risque de 13,7 pour 10 000 pour une population de fumeurs exposés à 1,2 mg/m^3 durant 40 ans (5 jours par semaines, 8 heures par jour) pour un travail comportant une activité physique importante. Le risque passe à 0,003 pour 10 000 pour une exposition à 0,6 mg/m^3. Ces différents modèles suggèrent un risque cancérogène faible pour des concentrations de formaldéhydes inférieures à 1 mg/m^3 et négligeable pour des expositions sous le seuil de 0,3 mg/m^3 proposé par le DECOS.

L'exposition professionnelle au formaldéhyde par inhalation provient principalement de trois types de sources : la décomposition thermique ou chimique des polymères à base de formaldéhyde (par exemple les résines urée-formaldéhyde), l'émission de formaldéhyde par ses solutions aqueuses (par exemple à partir des liquides d'embaumement), ou la formation de formaldéhyde comme produit secondaire de la combustion d'une variété de composés organiques (par exemple les gaz d'échappement ou la fumée de cigarette) (CIRC, 1995).

Plusieurs VLE en milieu de travail ont été proposées pour le formaldéhyde. Le tableau N°4 présente quelques valeurs limites légales ou recommandées au Canada, aux Etats-Unis et en France. Les experts du DECOS proposent une valeur limite d'exposition sur huit heures de 0,15 mg/m^3 accompagnée d'une limite sur 15 minutes de 0,5 mg/m^3 (Dutch Expert Committee on Occupational Standards, 2003).

Tableau N°4 : Valeurs limites d'exposition professionnelle au formaldéhyde en vigueur au Canada

Juridiction	**Valeur limite**
Canada	0,37 mg/m^3 (0,3 ppm), plafond
Alberta	0,92 mg/m^3 (0,75 ppm), VEMP 8h 2,46 mg/m^3 (2 ppm), plafond
Colombie-Britannique	0,37 mg/m^3 (0,3 ppm), VEMP 8h 1,23 mg/m^3 (1 ppm), plafond
Ontario	1,23 mg/m^3 (1 ppm), VEMP 8h 2,46 mg/m^3 (2 ppm), VECD 15 min
Québec	2,46 mg/m^3 (2 ppm), plafond, mentions C2, EM et RP
France	0,62 mg/m^3 (0,5 ppm), VEMP 8h 1,23 mg/m^3 (1 ppm), VECD 15 min
États-Unis	0,92 mg/m^3 (0,75 ppm), VEMP 8h (intervention: 0,5ppm) 2,46 mg/m^3 (2 ppm), VECD 15 min
ACGIH	0,37 mg/m^3 (0,3 ppm), plafond
NIOSH	0,020 mg/m^3 (0,016 ppm), VEMP 8h 0,12 mg/m^3 (0,1 ppm), plafond

VEMP = Valeur d'exposition moyenne pondérée; VECD = Valeur d'exposition de courte durée; C2 : Cancérogène soupçonné; EM : Réduction de l'exposition au minimum; RP : Recirculation interdite

L'examen de la section de la monographie du CIRC de 1995 (la plus récente publiée) sur l'exposition professionnelle au formaldéhyde montre un manque de données d'exposition au formaldéhyde pour les deux dernières décennies, en particulier après 1990. De plus peu d'études ont présenté des portraits historiques multisectoriels de l'exposition à cette substance. Dans le cadre de la surveillance de l'exposition professionnelle au formaldéhyde, sa classification comme cancérogène pour l'humain par le CIRC a créé un besoin en données récentes sur les niveaux d'exposition pour permettre d'identifier les secteurs d'activité à risque. Elle constitue de surcroît un incitatif à la conduite d'études épidémiologiques supplémentaires pour préciser le risque incitatif à la conduite d'études épidémiologiques supplémentaires pour préciser le risque relié au cancer. Il existe donc également un besoin en données quantitatives historiques sur les niveaux d'exposition.

1.4 Objectifs de la thèse

1.4.1 Objectif principal

Objectif général

L'objectif principal de ce travail, de nature méthodologique, est d'analyser plusieurs catégories de données préexistantes en vue de leur utilisation pour l'évaluation de l'exposition professionnelle.

Objectifs spécifiques

1. Caractériser les biais présents dans les sources de données préexistantes
2. Proposer une méthode d'exploitation des données de la littérature

1.4.2 Objectif secondaire

Objectif général

Le second objectif de cette étude, d'intérêt plus immédiat pour la santé publique, est d'établir un portrait historique de l'exposition professionnelle au formaldéhyde et d'identifier les déterminants de l'exposition à cette substance.

Objectifs spécifiques

1. Établir un portrait de l'exposition au formaldéhyde dans le secteur industriel des panneaux de bois aggloméré et identifier des déterminants de l'exposition dans ce secteur d'activité.

2. Établir un portrait multisectoriel historique de l'exposition au formaldéhyde et identifier des déterminants génériques de l'exposition à cette substance à partir de banques de données d'exposition professionnelle IMIS et COLCHIC.

1.5 Organisation de la thèse

La thèse est organisée autour de la présentation de quatre manuscrits d'articles scientifiques qui en constituent la contribution principale.

Le présent chapitre a présenté le contexte dans lequel ce travail a été entrepris et la revue de la littérature portant sur l'utilisation des sources de données préexistantes pour l'évaluation de l'exposition professionnelle. Il a également énoncé les objectifs de cette étude.

Le chapitre II, sous la forme d'un manuscrit d'article scientifique publié dans la revue « Annals of Occupational Hygiene », présente l'analyse par modélisation statistique des données d'exposition au formaldéhyde mesurées dans les usines québécoises de panneaux de bois aggloméré. Ces données comprennent des mesures effectuées par une équipe de recherche de l'IRSST en 2001-2002 et des mesures prises de 1984 à 2002 par des équipes d'hygiène du travail des CLSC dans les mêmes usines dans le cadre de leurs activités de prévention.

Le chapitre III, sous la forme d'un manuscrit d'article scientifique soumis pour publication dans la revue « Annals of Occupational Hygiene », présente l'analyse des données d'exposition au formaldéhyde issues d'une revue de littérature dans le secteur des panneaux de bois aggloméré. Il comprend également la description d'une méthode originale d'exploitation des données de la littérature et une comparaison des résultats avec les données québécoises.

Le chapitre IV, sous la forme d'un manuscrit d'article scientifique publié dans la revue « Annals of Occupational Hygiene », présente l'analyse par modélisation statistique des données d'exposition au formaldéhyde enregistrées dans la banque de données d'exposition professionnelle multisectorielle française COLCHIC entre 1986 et 2003.

Le chapitre V, sous la forme d'un manuscrit d'article scientifique soumis pour publication dans la revue « Journal of Ocupational and Environmental Hygiene », présente l'analyse par modélisation statistique des données d'exposition au formaldéhyde enregistrées dans la banque de données d'exposition professionnelle multisectorielle états-unienne IMIS entre 1979 et 2001. Il comprend également une comparaison des résultats avec les données de COLCHIC.

Le chapitre VI présente la discussion générale des résultats obtenus durant cette étude ainsi que les conclusions tirées de ce travail.

CHAPITRE II

ÉTUDE DES DÉTERMINANTS DE L'EXPOSITION AU FORMALDÉHYDE DANS L'INDUSTRIE DES PANNEAUX DE BOIS AGGLOMÉRÉ AU QUÉBEC

Article publié dans "Annals of occupational hygiene", Vol 49 (7), pp.587-602

Investigation of determinants of past and current exposures to formaldehyde in the reconstituted wood panel industry in Quebec

Jérôme Lavoué[(1)], Charles Beaudry[(1)], Nicole Goyer[(2)], Guy Perrault[(2)], and Michel Gérin[(1)*].

(1) Groupe de recherche interdisciplinaire en santé (GRIS)
Département de santé environnemtale et santé au travail
Faculté de médecine
Université de Montréal
P.O Box 6128, Main Station
Montreal (QC) Canada H3C 3J7

(2) Quebec Research Institute for Occupational Health and Safety (IRSST)
505, De Maisonneuve blvd. West
Montréal (QC) Canada H3A 3C2

Keywords: formaldehyde, determinants of exposure, mixed-effects models, particle board, oriented-strand board, medium density fibre board.

* Author to whom correspondence should be addressed.

2.1 Abstract

Introduction

Past and present formaldehyde measurements made in facilities manufacturing reconstituted wood panels in Quebec have been collected in order to assess formaldehyde exposure and its determinants in this industry.

Methods

All 12 plants manufacturing Oriented-strand board (OSB), Medium density fibreboard (MDF), and Particle board (PB) in Quebec were visited by a research team which took area and personal measurements. Past measurements taken by governmental occupational health teams in these plants were also collected. Log-transformed formaldehyde concentrations were analysed with extended linear mixed-effects models.

Results

During 2001-2002, 275 measurements were taken by the research team, while 590 measurements dating back to 1984 were collected from governmental files. The area measurements had a global geometric mean (GM) of 0.28 ppm (geometric standard deviation (GSD): 3.1). The GM of the personal measurements was 0.17 ppm (GSD: 2.3). The fixed-effects of the models for personal and area measurements explained 61% and 57% of the variance respectively. Job (working area for area concentrations), process (PB, MDF, OSB), season of sampling, origin of the data (research, governmental), and year of sampling were significant determinants of exposure.

Proximity to the press, winter conditions, PB and MDF processes, and governmental data resulted in the highest exposures. Significant within-sampling campaign correlation was found for both personal and area models. The final models include different residual variances by process for personal measurements and by working area for area measurements.

Conclusions

Several determinants of exposure to formaldehyde in the reconstituted wood-panel industry were successfully identified. Higher levels found in governmental data as compared to research data may be explained by a "worst-case" strategy bias. The observed intra-sampling campaign correlation supports existing results suggesting that measurements taken in a small time frame tend to be correlated. Exposures in this sector are low compared to most 8h-TWA occupational exposure limits (e.g 1 ppm) but close to the most demanding ones (e.g. 0.3 ppm).

2.2 Introduction

2.2.1 Context

The manufacture of reconstituted wood panels has long been associated with exposure to formaldehyde, which is an irritant gas considered by the International Agency for Research on Cancer (IARC) as carcinogenic to humans (group 1) {IARC, in press #918}. Most of the processes in this industry involve the use of formaldehyde-based resins as the binding agent. The recent change in the IARC classification of formaldehyde from group 2A (probably carcinogenic) to group 1, based on sufficient evidence that it causes nasopharyngeal cancer in humans, constitutes an incentive for improved exposure surveillance in workplaces where formaldehyde is present. Furthermore, information on levels and determinants of exposure to this substance are needed to help improve exposure assessment in epidemiological studies on the carcinogenicity of formaldehyde at other sites (such as the nasal cavity or the blood-forming system). Within the framework of a more global project aimed at estimating the economic and health impacts of lowering the occupational exposure limit (OEL) for formaldehyde in Quebec, an exposure assessment was conducted in the reconstituted wood panels industry (Goyer et al., 2004). Scenarios being studied for full shift exposure include an 8h-TWA OEL of 1, 0.75, and 0.3 ppm. With nearly 2000 workers in 2001-2002, this industry constitutes the largest industrial sector with formaldehyde exposure in Quebec.

2.2.2 Description of the industry

The reconstituted wood panel industry includes several processes and can be classified as either producing plywood products (in which panels are formed by the assembly of thin layers of wood) or composition boards (in which wood particles are bonded together to form the panels) (U.S. Census Bureau, 2002;Zimowski, 1986). The processes included in this study are limited to particle board (PB), medium density fibre board (MDF), and oriented-strand board (OSB), which all belong to the composition board category. These three processes involve the mixing of wood chips with a binding agent and compressing the mixture under high temperature (USEPA, 1998). For PB, the main raw materials comprise wood chips, saw dust, and planer shavings. In the MDF process the chips are formed into fibres prior to mixing with the resin. Liquid urea-formaldehyde (UF) and melamine-urea-formaldehyde (MUF) resins are generally used in the manufacture of MDF and PB. OSB panels are made from wood wafers produced on site from logs and mixed with liquid or powder phenol-formaldehyde (PF) resins.

2.2.3 Exposure to formaldehyde

The main sources of occupational exposure to formaldehyde in this sector include emissions from resins before pressing, mostly during the formation of the mat and along the system conveying it to the press, emissions from the press, and emissions from the newly formed panels in the finishing, storage, and shipping areas. Formaldehyde emissions from resins are caused in part by their free formaldehyde content, but most emissions at the press and from the hot panels arise from the hydrolysis of the cured resin and condensation reactions between wood compounds (Tohmura et al., 2001).

Tohmura et al. have observed a decrease in formaldehyde emissions when the melamine content in MUF resins is increased, caused by an increased resistance of the cured resins to hydrolysis (Tohmura et al., 2001). Wolcott et al. have investigated variables affecting formaldehyde emissions from the press during laboratory production of UF-bonded PB (Wolcott et al., 1996). They found that emissions increased when the following parameters increased: press time, press temperature, mat humidity, mat content in resin, and formaldehyde / urea molar ratio of the resin. Formaldehyde emissions were inversely associated with the thickness of the panels.

The OSHA Health Response Team sampled formaldehyde in four facilities in 1986, three manufacturing PB and one producing MDF (Zimowski, 1986). They report geometric means (GMs) of personal measurements for different jobs ranging from 0.10 to 0.32 ppm and from 0.18 to 1.8 ppm in PB and MDF respectively. All measurement durations were > 250 min. Kauppinen and Niemelä, in a review of formaldehyde exposure levels in facilities manufacturing PB in Finland prior to 1985, report geometric means of area measurements between 0.4 and 2.3 ppm depending on the localization in the plant. All measurement durations were below 120 min (Kauppinen and Niemelä, 1985). Niemelä et al. also report an arithmetic mean (AM) of 1.15 ppm for 220 measurements made between 1977 and 1979 (Niemelä and Vainio, 1981). More recently, Niemelä et al., presenting historical trends in this industry based on measurements taken in 8 facilities between 1980 and 1994, report successive measurement medians of 0.91 (n=21), 0.26 (n=31) and 0.46 ppm (n=9) for the periods 80-85, 86-90, and 91-94 (Niemelä et al., 1997). In their review of occupational exposure to formaldehyde, the authors of the 1995 IARC monograph, citing the results of a

Swedish study, report an AM of 0.2 ppm (obtained from 19 values) measured between 1980 and 1989 in facilities manufacturing MDF (International Agency for Research on Cancer, 1995). A value of 0.3 ppm for facilities manufacturing PB is also reported. Edling *et al.* (1988) studied the effects of formaldehyde on the physiology of nasal mucosa in workers of 3 plants manufacturing PB. They report that measurements taken by in-house hygienists between 1975 and 1983 were in the range 0.08-0.9 ppm with peaks reaching 4.1 ppm (Edling et al., 1988). In a study of the effects of formaldehyde on the mucous membranes and lungs of workers in a PB facility, Horvath et al. report that area and personal measurements taken in the plant during the study (supposedly in 1987 or 1988) had a median of 0.62 ppm (Horvath et al., 1988). Herbert et al. studied the effects of formaldehyde on the respiratory system of workers in the OSB industry in Alberta. The authors took 21-hour long measurements at different fixed stations in a facility using a PF resin. All 10 reported values were below 0.2 ppm (Herbert et al., 1995). Imbus et al. report that the results of 15 measurements taken in an OSB facility using PF resin were all below 0.05 ppm (Imbus and Tochilin, 1988).

The available literature provides little information about determinants of occupational exposure to formaldehyde. A few influential factors seem to modify formaldehyde emissions but their quantitative influence on occupational exposure is not documented. Moreover, there is a lack of recent data on exposure levels, most of the few reported exposure levels being prior to 1990. The present study aimed at documenting current formaldehyde exposure levels in Quebec in this industry and identifying their determinants.

2.3 Methods

2.3.1 Industrial hygiene surveys

The 12 plants manufacturing PB, MDF or OSB in Quebec were visited during the period from June 2001 to March 2002. The visits were conducted by a team of 2-4 industrial hygienists and technicians and lasted 1 to 2 days.

A standardized form was created to facilitate and systematize the information gathering process. One member of the research team completed the form either from observation of the workplace or from interview with employees. In order to describe industry-wide exposure trends, lists of standardized jobs and work zones in the plants were created (Table 1).

Table 1: Standardized jobs and zones in the reconstituted wood panel industry in Quebec

Standardized zones	**Description**
Raw materials receiving – Chip preparation	Includes the raw material storage area, and wood chips preparation area (different for each process) and the resin blending area.
Resin production	Area where the resin is stored and is produced from formaldehyde and other ingredients.
Main production	The mat formation, the press and the panel cooling wheel areas
Finishing	The panel aging, sawing, sanding, laminating, and painting areas
Storage – shipping	Packaging, storage, and shipping areas
Operator booth – internal ventilation	Any operator booth in the plant which is supplied with air coming from the inside of the facility
Operator booth – external ventilation	Any operator booth in the plant which is supplied with air coming from the exterior of the facility
Other departments	All non-production areas (administration, labs, maintenance shops)
Standardized jobs	
Raw materials receiving and supply	Employees working from the receiving area to the wood-resin mixture area
Chip preparation operator	Any employee working in the chip preparation area (OSB only)
Resin operator	Mixes the ingredients required for the production of the resin and takes samples at regular intervals for quality control purposes
Press operator	Operates the press from within an operator booth
Assistant press operator	Spends his shift partly in the press operator booth and partly close to the mat forming, the press, and the cooling wheel, to take readings of process parameters for quality control purposes
Finisher	Includes workers operating the sanding, trimming or painting machines, workers visually assessing the quality of the panels, and floaters assigned exclusively to the finishing operations
Flat mill operator	Includes all workers assigned to the operation of laminating the panels with printed paper overlays
Shipper	Includes forklift operators in the shipping area and automatic packing machine operators
Lab technician	Includes workers responsible for taking and/or analyzing samples of panels and other materials for quality control purposes
Cleaner	Worker operating the mechanical sweeper throughout the facility
Foreman	Foreman or production supervisor
Mech. Elec.	Any mechanic or electrician working inside the facility
Press-Miscellaneous tasks	Includes all workers performing tasks in close proximity to the press during their shift, mostly outside operator booths (e.g. surveillance of machinery, cleaning around the press with compressed air).
Floater	Includes workers assigned to various locations in the facility (mainly finishing and main production areas), mostly outside operator booths
Administration	Includes employees in the administration, management, or engineering departments

Time-weighted type monitoring was performed in each plant at fixed stations (area sampling) and in the respiratory zone of employees (personal sampling) by using adsorbent tubes coupled with sampling pumps. Employees and locations monitored were not chosen randomly but were selected with the aim to cover the greatest number of exposure circumstances in a facility given the available sampling resources.

Although all jobs/areas in the facilities were sampled, more sampling resources were allocated to those thought to be associated with the highest exposures. Sampling was not task oriented but rather aimed at representing full-shift exposure.

2.3.2 Governmental data

In Quebec, all companies in certain regulated industries are visited by governmental occupational health teams that identify health hazards, evaluate health risks, define required medical surveillance and devise corrective measures with the employers and employees. The visits may be conducted several times in the same plant if periodical reassessment is deemed necessary. All but one of the plants identified as manufacturing PB, MDF or OSB panels in Quebec were visited by these teams, as early as 1984. During each visit by the research team, all governmental exposure data were collected from the corresponding local health center along with the available ancillary information.

Quebec legislation requires sampling strategies to be designed according to the recommendations of the Quebec Research Institute for Occupational Health and Safety (IRSST) (IRSST, 2000). While mentioning several types of strategies (e.g. 'worst-case', random sampling), the IRSST guide does not provide precise guidelines to be used by governmental health teams. Moreover, the purpose of sampling may differ from situation to situation, i.e., hazard identification, follow up, task monitoring, compliance monitoring, or exposure profiling. Therefore, since the sampling strategies associated with the collected exposure data were not explicitly stated in the paper records, they are mostly unknown and may vary considerably across the exposure data.

2.3.3 Analytical methods

During the visits by the research team, formaldehyde concentrations were evaluated with sampling pumps coupled with solid sorbent tubes (type XAD-2) impregnated with 2-(hydroxymethyl) piperidine. Analysis was conducted by gas chromatography with Nitrogen Phosphorus detection (GC-NPD) (IRSST, 1995). Governmental teams used three different methods sequentially from 1984 to 2002. Before 1985, sampling was mostly performed with an impinger filled with a collecting solution containing dinitrophenylhydrazine and perchloric acid and the analysis conducted with liquid chromatography (IRSST, 1985). From 1985 to 1995, the sampling device used was a solid sorbent tube (type ORBO 22) impregnated with n-benzylethanolamine while the analysis was conducted by GC-NPD (IRSST, 1988). Since 1995, the sampling and analytical methods have been those used by the research team.

2.3.4 Data formatting

Area and personal measurements were analysed separately, which was motivated by the fact that area measurements rarely represent personal exposures adequately and may be influenced by different determinants. Inside the *Area* and *Personal* datasets, data originating from the research team and governmental records were merged for analysis. This allowed for increased statistical power during the analysis. A variable identifying the source of data was included in the statistical modelling to explore potential systematic differences.

A limit of detection (LOD) was calculated according to the sampling volume using the lowest reported formaldehyde mass of each analytical method. Values reported under a LOD were treated according to the recommendations of Hornung and Reed, separately for the area and personal datasets (Hornung and Reed, 1990). This led to assigning the value of the LOD divided by 2 to the non detected area measurements and the value of the LOD divided by the square root of 2 to the non detected personal measurements.

2.3.5 Statistical analysis

Based on a graphical assessment of the frequency distributions of area and personal concentrations, the response variable selected for analysis was the natural logarithm of formaldehyde concentrations. The data were analysed with extended linear mixed-effects models. Linear mixed-effects models have recently been used successfully to analyze occupational exposure data (Burstyn et al., 2000;Egeghy et al., 2002;Lagorio et al., 1997;Leena et al., 1999;Peretz et al., 2002;Raaschou-Nielsen et al., 2002;Rappaport et al., 2003;Rappaport et al., 1999;Symanski et al., 2001;Teschke et al., 1994;Van Tongeren and Gardiner, 2001;Weaver et al., 2001). Extended linear mixed-effects models, in addition to the possibility of modelling correlation patterns, allow exploring changes in the variability of the response as a function of other variables (Pinheiro and Bates, 2000). The model framework used in this study is described by the following equation:

$$Ln(C)_{ijk} = \sum(Fixed.effects) + (Random.effectA)_i + (Random.effectB)_{ij} + (Error)_{ijk} \quad (1)$$

$i = 1,\ldots,M,\ j = 1,\ldots,M_i,\ k = 1,\ldots,M_{ij}$

where there are M groups for variable A, M_i groups for variable B in the i^{th} group of variable A, and M_{ij} observations in the j^{th} group of variable B in the i^{th} group of variable A. The total number of observations is $\sum_{i=1}^{M}\sum_{j=1}^{M_i} M_{ij}$. $Ln(C)_{ijk}$ is the logarithm of the k^{th} observation in the j^{th} group of variable B in the i^{th} group of variable A. The model assumptions are: $(Random.effectA)$ and $(Random.effectB)$ are distributed normally with mean 0; $(Random.effectA)$, $(Random.effectB)$, and $(Error)$ are statistically independent; and $(Error)$ follows a multinormal distribution with mean 0 and different possible variance-covariance structures.

All fixed effects tested for inclusion in the model are presented in table 2. The year of sampling was tested as either a continuous variable or a nominal variable representing different time periods. The facility variable was tested as a random effect for the first level of grouping, with the sampling campaign as a random effect for the second level of grouping. The facility variable was tested as a random effect to determine the extent of similarity of exposures among different plants manufacturing the same product with the same process. The sampling campaign was tested as a random effect to explore potential correlation of measurements made in a small time window after controlling for the fixed effects. The correlation between measurements taken during the same campaign can be determined from the estimated inter- and intra-group variances by the calculation of the intra-class correlation coefficient (ICC, equation 2).

$$ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \quad (2)$$

where σ_B^2 is the inter-group variance and σ_W^2 is the intra-group or residual variance.

The variance structures tested for the error term were: a different residual variance (σ_W^2) for each level of one of the available nominal variables (equation 3) and a residual standard deviation varying linearly (equation 4) or exponentially (equations 5 and 6) with one of the continuous variables. Because of the unbalanced nature of our datasets and the limited number of measurements, interactions were not tested in the models.

Different models tested for the residual standard deviation (σ_w):

$\sigma_w = \beta_i$ (3), where β_i is to be estimated, i=1,…,n; n is the number of categories of the nominal variable

$\sigma_w = C * (\beta X)$ (4)

$\sigma_w = C * \exp(\beta X)$ (5)

$\sigma_w = C * \exp(\beta \ln(X))$ (6), where C and β are to be estimated and X is a continuous variable.

The S-plus software provides two ways of modelling random effects, which yields to a slightly different interpretation of the parameters when the residual error is not modelled as homoscedastic: the *lme* function provides an estimate of the intergroup variability (σ_B^2 in equation 2), therefore yielding a variable intra-group correlation (ICC) if the residual (or intra-group) variability (σ_W^2 in equation 2) has a heteroscedastic structure. The *gls* function provides an estimate of the intra-group correlation, therefore yielding a variable inter-group variability if the residual (or intra-group) variability has a heteroscedastic structure. The *lme* function was preferred in our study because it allows modelling of several levels of nested random effects.

REML optimization was used to choose the random effects and residual variance structures and estimate the final model parameters. ML optimization was used to compare models with different fixed effects structures (Pinheiro and Bates, 2000). Model building was performed by means of the following procedure: The best fixed effects model was first constructed with a forward stepwise routine using the Bayesian information criterion (BIC) as a discrimination criterion. Then the best random effect structure was added by comparing the BIC of the 4 possible models (no random effect, random effect A, random effect B, random effects A and B). Finally the variance structure for the residual error was assessed in a similar way. The next step consisted in retesting the fixed effects for removal or addition of variables. The random effects and variance structure were adjusted again if the fixed effect model had changed.

In order to illustrate the quantitative influence on exposures of the fixed effects coded as nominal variables, relative indices of exposure (RIE, equation 7) were calculated. Hence, for each variable, the category corresponding to the highest number of observations (the reference category) was assigned the value 100%. The RIE of each of the other categories of the variable was then calculated by finding the exponent of the difference between the estimated coefficient for that category and the one for the reference category.

$$RIE_{levelA}(\%) = 100 * \exp\left(Coeff_{levelA} - Coeff_{level\,Ref}\right) \quad (7)$$

where RIE_{levelA} is the relative index of exposure for level A of the variable in question, $Coeff_{level\ A}$ is the estimated coefficient corresponding to the category A and $Coeff_{levelRef}$ is the estimated coefficient corresponding to the reference category. $Coeff_{levelRef}$ is 0 when the reference category is included in the intercept. Thus, relative to the reference category, exposure levels associated with other categories are estimated as percentages.

Internal validation was primarily conducted by graphical assessment of residuals and estimates of random effects regarding the assumptions underlying the estimation. There was no external validation of the final models.

All analyses were conducted with the statistical software S-Plus 6.1 professional edition for Windows Release 1 (Insightful Corp., Seattle, WA).

2.4 Results

2.4.1 Information on the plants

Among the 12 plants visited, 6 manufactured OSB panels and had a median workforce of 120. Three plants manufactured PB panels, with a median workforce of 140. Three plants produced MDF panels, with a median workforce of 71. All plants except one OSB facility had exhaust ventilation systems extracting air above the press. The remaining plant had only a roof opening above the press. In all the plants most of the production jobs were on a 12-hour shift structure, with 3 different teams working alternately. No operation requiring the use of respiratory protection was monitored during the visits performed by the research team.

2.4.2 Data collection

The visits performed by the research team yielded 275 measurements while 590 were available from the governmental files. Fifty nine measurements were removed from this dataset prior to further analysis: 10 corresponding to a job performed in only one facility and for only a few months, 7 corresponding to task sampling or infrequent events, 6 samples considered 'dubious' by the technician, 1 sample taken directly above the press beside the exhaust system, 22 because they were labelled 'other' regarding the job or work zone classification, and 13 because the year of sampling was missing. In addition, 47 measurements were not included in the modelling datasets because they corresponded to jobs existing in only one or two of the three processes in the study. Inclusion of these data in the statistical modelling would have required nesting the variable job inside the variable process, which could not be achieved due to the unbalanced nature of our data. These 47 measurements were nevertheless included in the descriptive analysis. The variables documented for both sources of data are presented in Table 2.

Table 2: Variables tested in the statistical models

variable	Variable type (number of categories)	Description
Analytical method	Nominal (3)	Analytical method used for formaldehyde quantification (the three methods are consecutive over time with some overlap)
Measurement duration	Continuous (hours)	Time covered by the measurement
Number of tubes	Continuous	Number of sampling tubes used during the measurement
Origin of data	Nominal (2)	Origin of data, believed to be indicative of differing sampling strategies (the two categories are: research team, governmental data)
Standardized job	Nominal(15)	For each personal measurement, the workers were assigned to a standardized job category based on their job description (e.g. floater, see Table 1)
Standardized zone	Nominal (8)	Each area measurement was assigned to a standardized zone corresponding to a broadly defined location in the plant (e.g. finishing area, see Table 1)
Sampling date	Continuous	Date on which measurement was taken
Plant	Nominal(12)	Code of the plant sampled
Process	Nominal(3)	Code of process in operation at the plant (OSB/PB/MDF)
Season of sampling	Nominal (4)	Season of sampling as defined by the following cut-off dates: winter (12/22 to 3/20), spring (3/21 to 6/21) summer (6/22 to 9/22) autumn (9/23 to 12/21)
Sampling campaign	Nominal(varying number of classes)	A sampling campaign was defined as a group of measurements taken <4 days apart in the same plant.

2.4.3 Area measurements

Summary statistics along with the empirical distributions of area measurements stratified by their origin (research or governmental) are presented in Figure 1. The percentages of measurements below the LOD were 12 and 8 for the research measurements and the governmental data, respectively. The research and governmental measurements were taken, respectively, during 11 and 63 sampling campaigns. Concentrations reported by the research team covered a median duration of 4.9 h (interquartile interval: [2.4-5.9]). The corresponding values were 2.1 h [2.0-5.6] for the governmental data. Table 3 presents the GMs, geometric standard deviations (GSDs), and corresponding number of measurements stratified by work zone, process, and origin of the data.

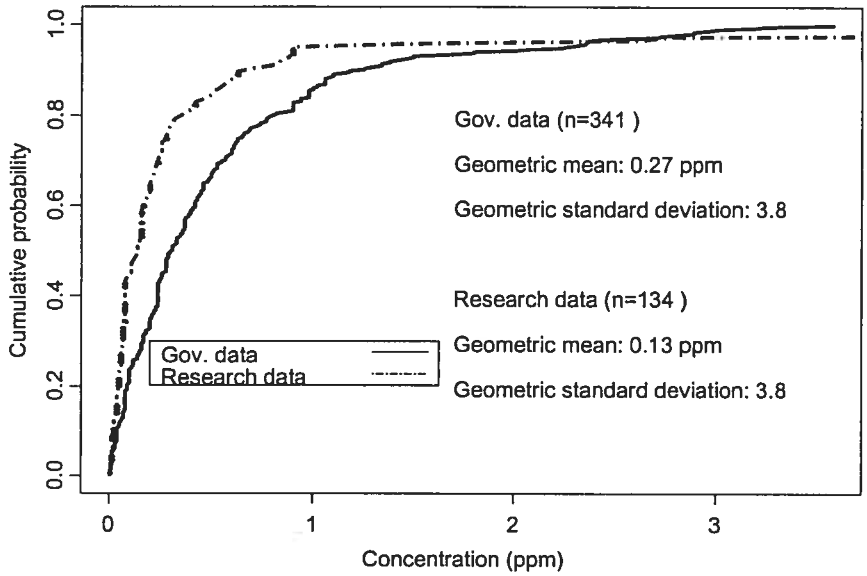

Figure 1: Cumulative distributions of area measurements stratified by origin of the data

Table 3: Geometric means (ppm) and geometric standard deviations of all formaldehyde exposure concentrations stratified by job/zone, process, and origin of the data

	OSB process				MDF process				PB process			
	Research		Governmental		Research		Governmental		Research		Governmental	
	GM	GSD (n)	GM	GSD (n)	GM	GSD (n)	GM	GSD (n)	GM	GSD (n)	GM	GSD (n)
Personal data												
Administration	-	-	0.04	1.1(7)	-	-	0.03	1.3(6)	-	-	-	-
Assistant press operator	0.04	1.3(4)	0.07	1.8(7)	0.09	2.0(5)	0.27	1.6(15)	0.20	1.3(5)	0.28	1.4(6)
Foreman	-	-	0.05	(1)	-	-	-	-	0.08	(1)	-	-
Shipper	0.06	1.5(7)	0.10	(2)	0.19	2.0(4)	-	-	0.16	1.7(7)	0.20	2.3(20)
Finisher	0.05	1.6(17)	0.07	1.2(10)	0.16	1.8(6)	0.24	1.4(10)	0.19	1.6(12)	0.26	1.9(20)
Cleaner	0.05	1.6(11)	0.54	(1)	0.08	(1)	-	-	0.12	2.3(4)	0.26	3.0(8)
Floater	0.04	(2)	0.22	1.9(5)	-	-	-	-	-	-	-	-
Press-Miscellaneous	-	-	-	-	-	-	-	-	0.23	(2)	0.45	1.8(9)
Press operator	0.03	(1)	0.08	(2)	-	-	0.07	(1)	-	-	0.26	1.7(7)
Lab technician	0.05	1.4(4)	0.06	1.4(4)	0.14	(1)	0.31	(2)	0.16	(1)	0.42	(2)
Maintenance worker	0.06	1.7(8)	0.08	2.7(4)	0.12	1.6(4)	0.11	2.0(13)	0.10	2.1(5)	0.22	2.2(9)
Resin operator	-	-	-	-	0.22	(1)	0.34	3.2(5)	0.16	(1)	0.41	2.1(8)
Chip preparer	0.04	1.3(4)	0.11	2.9(5)	-	-	-	-	-	-	-	-
Flat mill operator	-	-	-	-	-	-	-	-	0.12	1.6(8)	0.17	2.0(15)
Area data												
Operator booth (internal vent)	-	-	-	-	-	-	-	-	0.68	(1)	0.31	1.5(12)
Operator booth (external vent)	0.02	1.3(3)	0.01	3.0(4)	0.07	3.2(3)	-	-	0.09	2.1(7)	-	-
Other departments	0.04	(2)	0.03	2.0(9)	0.04	(1)	0.06	4.1(21)	0.14	(2)	0.09	4.1(5)
Storage-shipping	0.04	2.3(5)	0.06	1.4(4)	0.17	1.9(3)	0.37	2.1(17)	0.09	3.9(5)	0.17	2.7(21)
Finishing	0.09	1.8(9)	0.09	3.1(10)	0.22	2.5(6)	0.33	2.1(19)	0.20	2.0(13)	0.29	1.7(33)
Main production	0.05	2.5(29)	0.15	2.0(25)	0.26	2.1(15)	0.75	2.5(95)	0.75	3.4(22)	0.56	2.1(29)
Resin production - Storage	0.05	(2)	0.09	1.1(3)	-	-	0.35	2.1(18)	0.08	(1)	2	(1)
Raw materials, Chip preparation	0.08	(2)	0.10	1.9(8)	0.16	(2)	-	-	0.42	(1)	0.03	2.3(7)

GM: geometric mean
GSD: Geometric standard deviation
n: number of available measurements

The final 'whole dataset' model includes the following fixed effects, which explain 57% of the total variance: *Process*, *Work zone*, *Season of sampling*, *Origin of data*, and *Year*. The coefficients of all fixed effects of the model are presented in Table 4, along with their standard error. The estimated REIs associated with each work zone are presented in Figure 2 along with an ~95% confidence interval (95% CI). The presented CIs correspond to the comparison of the levels of each variable to the reference category. Their exact interpretation is that if they exclude the 100% value, the exposure level associated with the category is significantly higher (or lower) than the exposure level associated with the reference category. The REIs corresponding to the other categorical variables are presented in Figure 3.

The best fit for the sampling year was obtained with a linear spline structure containing one knot, the line-break year being 1991. The other tested structures were a nominal variable with three levels (1984-1989, 1990-1994, and 1995-2002), a linear, and a quadratic trend (one 1st order term and one 2nd order term). The line-break year was determined by graphical assessment and refined by using the BIC. The estimated coefficient showed a yearly decrease of 19.7% until 1991 followed by a yearly increase of 9.4%.

A one-level random effect structure with the variable *Sampling campaign* as the random effect yielded the best fit compared with no random effect, *Facility* alone or *Sampling campaign* within *Facility*. The best residual variability structure in terms of BIC was one with different residual standard deviation for each work zone. With the *Finishing* *zone* as the reference (i.e value 100%), the relative residual standard deviations estimated by the model for the other work zones are as follows: *Main production* (153%), *Resin production-storage* (136%), *Storage-shipping* (131%), *Raw materials* (206%), *Other department* (226%), *Operator booths* (both interior and exterior) (119%). The estimated within-sampling campaign residual standard deviation is 0.57 for the reference category *Finishing zone* while the between-sampling campaign standard deviation is 0.36, giving ICC values varying between 0.07 and 0.29, with an average of 0.17.

Graphical assessment of the distribution of the normalized residuals yielded satisfactory fit to the normal distribution. Likewise, the estimates of the random effects indicated graphical conformity to the normal distribution. No systematic trend was found during examination of the variations of residuals and estimated random effects stratified by the different levels of the other available variables, indicating satisfactory conformity to the independence hypothesis.

Table 4: Coefficient estimates of the final models

Area measurement model			Personal measurement model		
Fixed effect	estimate	SE	Fixed effect	estimate	SE
Intercept[a]	-0.89	0.29	Intercept[b]	-2.48	0.54
Process OSB	-1.82	0.19	Process OSB	-0.97	0.15
Process PB	-0.47	0.17	Process PB	0.01	0.16
Operator booth (external vent)	-1.62	0.19	Job group 1[c]	-0.86	0.16
Operator booth (internal vent)	-0.25	0.26	Job group 2[c]	-0.16	0.07
Finishing	-0.52	0.09	Job group 4[c]	0.62	0.14
Other departments	-1.96	0.22	Research data	-1.06	0.23
Raw materials - Chip preparation	-1.25	0.28	Automn	-0.27	0.16
Resin production - storage	-0.59	0.17	Spring	-0.65	0.17
Storage-shipping	-0.86	0.12	Summer	-0.58	0.25
Research data	-1.09	0.26	Measurement duration (h)	-0.05	0.07
Autumn	-0.39	0.17	(year-1984)	-0.08	0.03
Spring	-0.74	0.20	max(year-1984, 11)[d]	0.19	0.06
Summer	-0.70	0.19	% total variance explained by the fixed effects		
(year-1984)	-0.22	0.07	Personal :	61%	
max(year-1984,7)[d]	0.31	0.08	Area :	57%	

(a) The intercept represents the estimated AM of the log-transformed concentrations (ppm) for the following combination: MDF, main production zone, governmental data, winter, and 1984. Estimates of the AM of the the log-transformed concentrations for other combinations of the levels of the variables are obtained by adding the appropriate coefficients, e.g. substracting 1.82 from the intercept to change the process to OSB.
(b) The intercept represents the estimated arithmetic mean of the log-transformed concentrations for the following combination: MDF, Job group 3, governmental data, winter, duration of measurement 0 hour, and year 1984.
(c) group 1 includes *Administration* and *Foreman*, group 2 includes *Lab technician*, *Maintenance worker*, and *Cleaner*, group 3 includes *Press operator*, *Assistant press operator*, *Finisher*, *Shipper*, and group 4 includes *Floater* and *Press-miscellaneous*.
(d) Coefficient corresponding to the spline: multiply the coefficient by either (year-1984) or the given number (7 or 11) whichever is greater. Practically this corresponds to a constant added to the intercept until the spline knot at which point the coefficient increases linearly with years.

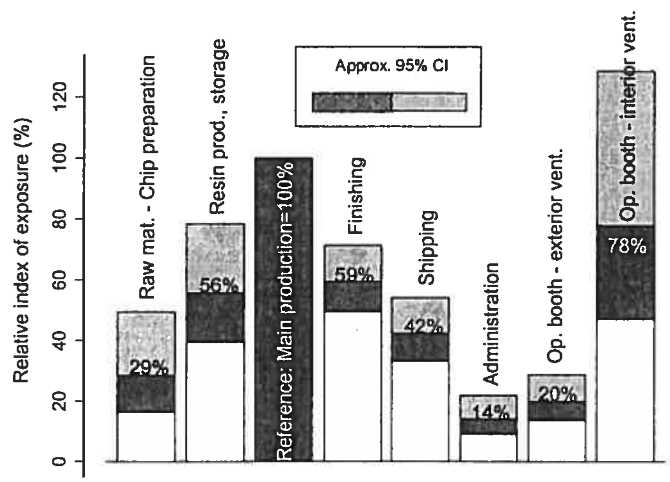

Figure 2: Relative exposure indices of the different zones in the final area model

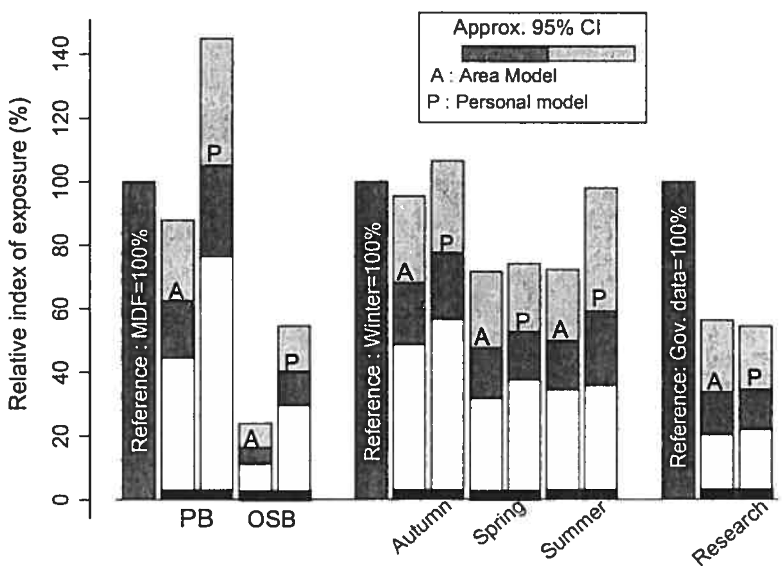

Figure 3: Relative exposure indices of predictive variables common to the area and personal models

2.4.4 Personal measurements

Summary statistics of the personal measurements are presented in Figure 4. The percentages of measurements below the LOD were 23 and 9 for the research measurements and the governmental data respectively. The research and governmental personal measurement were taken during 12 and 67 sampling campaigns, respectively. Concentrations reported by the research team covered a median duration of 4.8 h (interquartile interval: [3.3-6.0]). The corresponding value was 5.9 [4.4-8.0] for the governmental data. GMs, GSDs, and corresponding number of measurements stratified by job, process, and origin of the data can be found in Table 3.

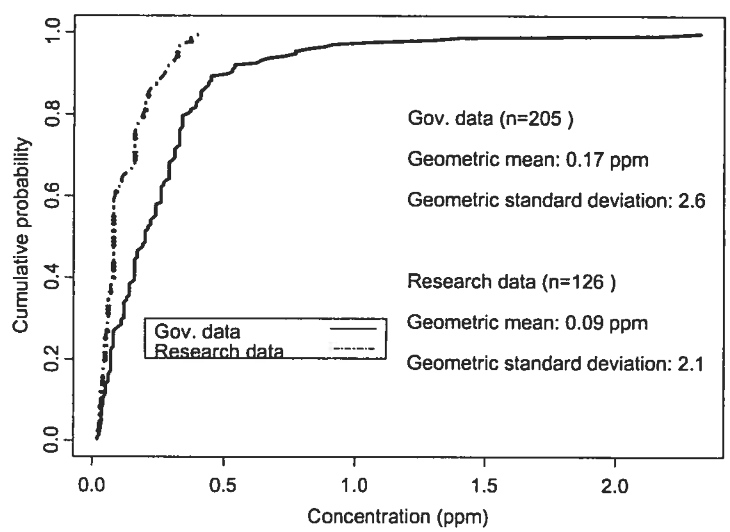

Figure 4: Cumulative distributions of the personal measurements stratified by origin of the data

The initial 'whole dataset' model included the following fixed effects, which explained 62% of the total variance: *Process*, *Job*, *Season of sampling*, *Origin of data*, *Year*, and *Measurement duration*. Further examination of the coefficients for the different jobs led to the testing of the grouping of jobs into 4 exposure groups: group 1 includes *Administration* and *Foreman*, group 2 includes *Laboratory technician*, *Maintenance worker*, and *Cleaner*, group 3 includes P*ress operator*, *Assistant press operator*, *Finisher*, *Shipper*, and group 4 includes *Floater* and *Press-miscellaneous tasks*. The resulting model yielded a better fit than the original one in terms of BIC. The percentage of total variance explained by the fixed effects was marginally reduced (to 61%). The RIEs associated with each job group are presented in Figure 5 while the RIEs corresponding to the other nominal variables in the model can be found in Figure 3. The coefficients for the fixed effects of the reduced model can be found in Table 4.

The best fit for the sampling year was obtained with a linear spline structure containing one knot, the line-break year being 1995. The estimated trend corresponds to a yearly decrease of 7% until 1995 followed by a yearly increase of 12.2%. The estimated influence of the measurement duration on exposure levels corresponds to a decrease of 5.2 % when the duration is increased by 60 min. (95% CI: 2.1-8.2).

A one-level random effect structure with the variable *Sampling campaign* as the random effect yielded the best fit compared with *Facility* alone or *Sampling campaign* within *Facility*. The best residual variance structure in terms of BIC was one with different residual standard deviations for each process. With the process PB as the reference (i.e value 100%), the relative residual standard deviations estimated by the model for the other processes are as follows: *MDF* (68%), *OSB* (63%). The estimated within-sampling campaign residual standard deviation (of the log-transformed concentrations) is 0.53 for the reference category while the between-sampling campaign standard deviation is 0.46, giving ICC values varying between 0.42 and 0.66, with an average of 0.56.

As for the area measurements, graphical assessment of the distribution of the normalized residuals and of the estimates of the random effects yielded satisfactory fit to the normal distribution and satisfactory conformity to the independence hypothesis.

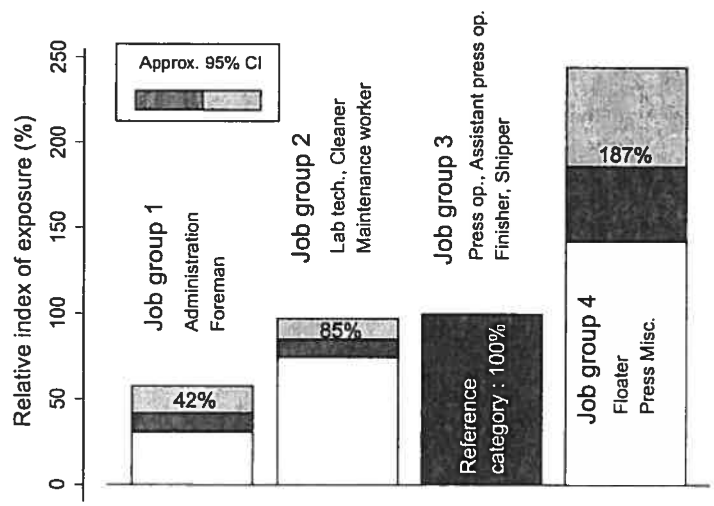

Figure 5: Relative exposure indices of the different exposure groups in the final personal model

Table 5 presents, side by side for *Research* and *Government* data, yearly geometric means of area and personal levels stratified by work zone/exposure group and estimated for the year 2002 by the statistical models. For the personal measurements, the estimates were calculated for a duration equal to the median duration of the measurements (5.6 h)

Table 5: Yearly geometric means (in ppm) for research and government data stratified by standardized zones and jobs, estimated by the statistical models for 2002.

	OSB		PB		MDF	
	Research	Government	Research	Government	Research	Government
Area measurements						
Operator booth (external vent)	0.01	0.04	0.05	0.16	0.09	0.25
Operator booth (internal vent)	0.05	0.16	0.21	0.62	0.34	1.00
Finishing	0.04	0.12	0.16	0.48	0.26	0.76
Other departments	0.01	0.03	0.04	0.11	0.06	0.18
Raw materials - Chip preparation	0.01	0.04	0.05	0.16	0.09	0.25
Resin prod - storage	0.04	0.12	0.15	0.45	0.24	0.71
Storage-shipping	0.03	0.09	0.11	0.34	0.18	0.54
Main production	0.07	0.21	0.27	0.80	0.43	1.28
Personal measurements						
Job group1[a]	0.02	0.05	0.05	0.14	0.05	0.14
Job group2[b]	0.04	0.11	0.10	0.28	0.10	0.28
Job group3[c]	0.04	0.13	0.12	0.33	0.11	0.33
Job group4[d]	0.08	0.23	0.22	0.62	0.21	0.62

(a) includes administration and foreman
(b) includes lab technician, maintenance worker, and cleaner
(c) includes press operator, assistant press operator, finisher, shipper
(d) includes floater and press-miscellaneous

2.5 Discussion

2.5.1 Determinants of exposure to formaldehyde in the reconstituted wood panel industry

The percentages of total variability explained by the fixed effects of both personal and area models (respectively 61 and 57%) compare favourably with similar studies in the field of occupational exposure assessment (ranging between 20 and 70%) (Bakke et al., 2002;Burstyn and Teschke, 1999;Van Tongeren and Gardiner, 2001). Thus, strong determinants of exposure to formaldehyde in this industry have been identified.

The estimated coefficients for the variable *Process* confirm, in both area and personal models, the lower potential for formaldehyde emissions when PF resins are used (all OSB facilities in our study used PF resins). However, while personal exposures tend to be similar for MDF and PB, MDF ambient levels seem higher than PB levels. This difference might be partly explained by the fact that measures to prevent worker exposure, such as reducing time spent in proximity to the press, might have been implemented to a greater extent in MDF. However, confounding by variables not accounted for in the modelling process cannot be ruled out as an explanation of this observation.

For both area and personal models, the origin of data was also a strong predictor of formaldehyde levels, research data being lower than governmental data by a factor of 3

(Figure 3). Several hypotheses may explain this observation also apparent in the empirical cumulative distributions shown in Figures 1 and 4. Firstly, the research visits having been planned long in advance with the management of the plants and after several meetings with representatives of this industry, it is possible that 'favourable' production conditions were implemented on the days of the visits. This would cause a negative bias compared to real exposure conditions. While Olsen et al., among others, have mentioned this source of bias as a possibility, it has, to our knowledge, not been further discussed in the published literature (Olsen et al., 1991). A second explanation is confounding by the temporal variable in the models. All research measurements were taken in 2001-2002 while some governmental data date back to 1984. This points to a potential collinearity issue between the two variables. This possibility was further explored in two different ways. First, the area and personal models were fitted to the data restricted to years common to both research and governmental datasets (2001 for the area measurements and 2001 and 2002 for the personal measurements). For the area analysis, with 109 observations (of which 26 were governmental), the estimated relative exposure index for the research data (the governmental data constitutes the reference category) was 47% compared to 34% for the initial model. For the personal analysis, with 138 observations (of which 24 were governmental), the estimated relative exposure index for the research data was 45% compared to 35% for the initial model. These results support the presence, albeit limited, of some confounding by the temporal trends. Furthermore, the full models were fitted without the 'origin of data' variable to datasets restricted to governmental data in order to assess the robustness of the observed temporal trend. For the area analysis, with 341 observations, the coefficients for *[year-1984]* and *[max(year-1984,7)]* (see Table 4) were respectively -0.17 (-0.22 for the initial model) and 0.25 (0.31 for the initial model). For the personal analysis, with 171 observations, the coefficients for *[year-1984]* and *[max(year-1984,7)]* were respectively -0.10 (-0.07 for the initial model) and 0.23 (0.29 for the initial model). These results tend to show that a temporal trend actually exists independently of the origin of the data, and that the estimated trend is close to that observed in the global models. Thus, although a degree of overestimation of the difference between research and governmental data may exist because of confounding, its extent should be small. We believe that, because of the limited resources that can be devoted to sampling during industrial visits, the hygienists tend to monitor worst-case scenarios in order to optimize the interpretation of their results. Since, by devoting more sampling resources to monitor supposedly 'high' exposure jobs or area, the strategy used by the researchers might be regarded as biased if jobs or work-zones are not controlled for, it must be concluded that within the jobs and area classification used in this study, governmental teams tended to sample tasks or sub-areas specifically associated with high formaldehyde concentrations.

The 'worst-case' bias in data from governmental sources, often associated with compliance monitoring, has already been mentioned in the literature (Stewart and Rice, 1990;Vinzents et al., 1995). Olsen et al. reported an actual comparison of measurements taken with a random sampling strategy with measurements existing in a governmental database. The authors found that the governmental data were higher than the 'random' data by a factor from 5 to 10, which is compatible with our interpretation. On the other hand, a variable identifying the sampling strategy as either 'research' or 'compliance' did not improve the fit of a multiple regression model applied to softwood dust levels in British Columbia lumber mills (Hall et al., 2002).

It might be argued that research and governmental data should not be analysed together because they may not be influenced by the same determinants. In order to explore that possibility, both area and personal models were fitted to a restricted dataset to allow the testing of 1^{st} order interactions between *Origin of data* and the other fixed effects, and of a possible difference of residual variance structure for the research and governmental data. With respectively 392 and 236 data available for analysis for the area and personal measurements, none of the tested additional models improved the fit compared to the original ones. In addition to the increased statistical power and the fact that the research and governmental measurements were taken in the same workplace, we believe that the merged analysis of both types of data was justified in our case.

A significant time trend, best modelled by a one knot linear spline, existed both for the area and personal measurements. The use of linear splines to model irregular temporal trends in occupational exposures has already been reported (Friesen et al., 2003;Raaschou-Nielsen et al., 2002). Although the spline knot year was different for the personal and area measurements (1995 and 1991), both estimated trends show high exposures at the beginning of the study period, decreasing until 1991-1995 and then increasing, although moderately, until the end of the study period (2002). Both the nominal and quadratic coding of the year of sampling yielded the same temporal pattern. This pattern differs from the reported generic occupational exposure decrease over time reported by Symanski et *al.* (Symanski et al., 2001;Symanski et al., 1998). A possible explanation might be the documented important increase in production in that industry in Northern America in the middle of the 90s (Spelter, 1997).

The other influent variable common to both the area and personal models is the season of sampling. Thus, winter, and, to a lesser extent, autumn, are associated with higher exposures than summer and spring. From direct observation and interviews with the employees of the visited plants, it appears that during the cold season (in Quebec temperatures are commonly below -15°C during this period) the ratio of recirculated air to fresh outside air is increased because of associated heating costs. Thus, the facilities tend to be in a negative pressure relative to the outside, which in turn causes the exhaust systems to loose efficiency. Furthermore, most doors and windows are open during the hot season, improving the air replacement rates. Van Tongeren et Gardiner have reported lower exposure levels during summer for some job titles in the carbon black manufacturing industry (Van Tongeren and Gardiner, 2001).

For the area model, the work zone had a significant influence on exposure levels. The main production area is associated with the highest exposures, with levels decreasing as one gets further from the press to the finishing, and then the shipping areas. As expected, departments separated from the main production area are associated with low formaldehyde levels. The results shown in Figure 2 also demonstrate the importance of supplying operator booths with outside air. Thus, the exposure levels inside operator booths ventilated with air from the plant are comparable to those found outside the booths (most operator booths with such ventilation were in the finishing zone).

Regarding the personal measurements, the modelling of the variable 'job' allowed the identification of four similar exposure groups, presented hereafter from the least to the most exposed: group 1 includes employees spending most of their time outside the production zone (e.g. foreman), group 2 includes employees spending part of their shift in the production zone (e.g. mechanics), group 3 includes workers spending their whole shift in the production zone but with a variable proportion of time spent in operator booths (e.g. press operator), and group 4 includes workers spending their whole shift in the production zone unprotected by operator booths (e.g. floater).

Higher personal exposure levels were also associated with shorter measurement duration. This trend was not modified when fitting the model separately to research and governmental data. It may be partly explained by the fact that longer measurements may include unexposed periods such as breaks (Raaschou-Nielsen et al., 2002). Since there was no significant interaction between duration and the source of data for the personal measurements, together with the absence of any influence in the case of area measurements, we conclude that the upward bias found in governmental data is not due to merely shorter measurement durations associated with a task-based strategy.

The significance of the variables identifying work-zones and jobs in both models underlines the usefulness of the standardized lists created during this study. They appeared as potentially strong predictors of exposure to formaldehyde, based on technical literature and field observation: this was validated and refined by the statistical modelling. Consequently, their use should be explored in other studies of formaldehyde exposure in the same industry.

The determinants identified in our study will be useful from the industrial hygiene standpoint for devising sampling strategies in this industrial sector, for example to help identify a priori potential overexposure situations. Within the framework of epidemiology, the identification of time trends is important for retrospective exposure assessment, while determinants such as jobs or process can help in the elaboration of specialized questionnaires in population-based case-control studies.

2.5.2 Structures of variance-covariance of exposure data

Random effects models have recently been used to explore patterns of correlation among occupational exposure measurements. However, the studies have mostly focused on modelling intra- and inter-worker variability (Burstyn et al., 2000;Kromhout et al., 1993;Rappaport et al., 2003;Rappaport et al., 1999;Van Tongeren and Gardiner, 2001). In our study, information on the identity of workers was not available in the majority of governmental files. Moreover, the sampling campaigns performed by the research team, lasting 2 days at the most, did not allow resampling the same workers because of the rotating teams work structure. Failing to model intra-and inter-worker variability constitutes a limitation of the present study in that this hampers rigorous conclusions about the long term risks posed to workers in this industrial sector. In particular we can not estimate probabilities of overexposure of a random worker as described by Rappaport et al. (Lyles et al., 1997;Rappaport et al., 1995). However the estimated yearly GMs that we present in table 5 allow drawing a general picture of exposure levels in this industrial sector, together with our results on the determinants of exposure.

Although the variable *Facility* was tested as a random effect, the results of both area and personal models show no significant intra-facility correlation, indicating similarity of exposure levels across facilities when all other variables are taken into account. This result is plausible since during the industrial hygiene visits by the research team, it was observed that most facilities were quite similar in terms of architecture (the press being the main emission source inside the plant, contaminating other areas), machinery used, and exposure control measures.

The variable *sampling campaign* improved the fit significantly when tested as a random effect. This corresponds to the fact that after controlling for all fixed effects in the models, there are systematic differences between formaldehyde levels measured during different campaigns, and is equivalent to the existence of correlation among measurements taken during the same campaign. The average intra-sampling campaign correlation coefficients estimated in our study were respectively 0.17 and 0.56, respectively, for the area and personal measurements. Teschke et al. report a coefficient of correlation of 0.31 between wood dust levels measured during the same inspection in data taken from OSHA's occupational exposure database (Teschke et al., 1999). While other authors have reported evidence of correlation in shift-long measurements taken on consecutive days (Buringh and Lanting, 1991;Deadman et al., 1996;Symanski and Rappaport, 1994), some observed no evidence of such correlation (Francis et al., 1989;George et al., 1995). Our results confirm the presence of correlation between shift-long exposures measured in short time periods in this industrial sector, implying a potential for underestimation of the day-to-day exposure variability when an assessment is based only on one sampling campaign conducted over a few consecutive days.

Significant structures of heteroscedasticity of the error term were found in both area and personal models. Hence, personal exposure appeared more variable for the process PB than for MDF and OSB. Moreover, ambient formaldehyde levels from the work zones *Main production*, *Department other than production*, and *Raw material reception* were much more variable than those measured in other locations in the facilities. The variability of exposure levels determines the number of samples necessary to assess an exposure situation with adequate precision. Furthermore, with regard to statistical modelling, estimates of other parameters in the model depend on the variance-covariance structures in the case of unbalanced data, which is present in most datasets in this field of research. Therefore it appears important to take into account and explore such structures of variability when modelling occupational exposures (Pinheiro and Bates, 2000).

2.5.3 Formaldehyde exposure levels in the reconstituted wood panel industry in Quebec

The observed (Table 3) and estimated (Table 5) levels of exposure to formaldehyde in this industry are consistent with the few results reported in the most recent literature (posterior to 1985). In particular, the exposure levels reported by Zimowski show very close agreement with those measured in Quebec. Exposures in the OSB manufacturing industry are mostly below 0.1 ppm with some work zones / job associated with somewhat higher levels according to the governmental data. Exposure levels in the MDF and PB manufacturing industries are similar, with observed and estimated GMs of personal and area levels between 0.1 and 0.4 ppm depending on the source of data. The highest estimated ambient GM is 0.43 ppm in the main production area of MDF manufacturing facilities (this estimate is increased to 1.28 ppm when governmental data are used). The highest personal estimated GM is 0.22 ppm for workers in close proximity to the press in the PB process (this estimate is increased to 0.62 ppm if the governmental data are considered). Levels reported prior to 1985 are consistently higher than those in our database. This maybe explained by the generalized implementation, around 1985, of low formaldehyde emission resins (OSHA, 1987).

Area measurements are consistently higher than personal measurements. This is consistent with the fact that the most exposed workers are those who spend the most time in the production area unprotected by ventilated booths. While none of the personal measurements in this study was greater than 2 ppm (the current Québec ceiling OEL), 4 and 6% of the research and governmental ambient concentrations, respectively, were greater than this value.

Globally, our results point to generally low personal full-shift exposure to formaldehyde in this industrial sector. However, the potential for short term high exposures associated with specific and occasional tasks cannot be ruled out on the basis of our study. Several

OELs exist for formaldehyde, varying both in type and level. In the US, OSHA enforces an 8-h TWA limit of 0.75 ppm with a short term exposure limit (STEL) of 2 ppm. The ACGIH recommends a ceiling limit of 0.3 ppm (ACGIH, 2003).

2.5.4 Validity of the statistical models

The internal validity of our models appears satisfactory considering the results of the graphical assessments of residuals and estimates of random effects described earlier. In addition, bootstrapping the model estimates 1000 times (results not shown) resulted in the observation that among 29 coefficient estimates for the 2 models the absolute relative difference between the model and the bootstrap estimate was >5% only in 6 cases, with a maximum of 17%. The sign of the difference between the two estimates was negative almost the same number of times as it was positive. The standard errors of the bootstrap estimates were on average 30% smaller than the corresponding model estimates, indicating moderate overestimation of the widths of the confidence intervals presented in Figures 2, 3, and 5. These results support reliance on the asymptotic assumptions linked to the estimation by ML or REML of confidence intervals for the model parameters (Harrel, 2001).

Although there was no formal external validation of the models, several elements suggest that our results may be applicable outside the restricted scope of our dataset. Thus, as seen in Figure 3, the estimates of the effects of variables common to the area and personal measurements are similar. Moreover, our estimates of the influence of several potential determinants of exposure are consistent with published observations.

Finally, the exposure levels estimated by the models are similar to those reported in the most recent literature. These observations, while not constituting an external validation *per se*, provide some insight into the potential for generalization of our results. However, the presence in our dataset of a bias not accounted for can not entirely be excluded since the data were not generated through a randomized sampling process.

2.6 Conclusion

Through statistical modelling of area and personal exposure measurements performed in the reconstituted wood panel industry in Quebec, several determinants of exposure to formaldehyde in this industrial sector were identified. The MDF and PB processes were associated with high exposure levels compared to OSB. Higher exposures also occurred during winter conditions compared to other seasons. While plant was not a strong predictor of exposure levels, work zones and jobs were strongly associated with area and personal measured concentrations, respectively. The highest levels were measured in areas close to the press and on workers spending most of their time in the press area. Moderate historical variations in exposure levels were also identified, best modelled by a one-knot linear spline. Governmental measurements were consistently found higher than those measured by the research team, pointing to the probable existence of a 'worst-case' bias in governmental data. The use of extended linear mixed-effects models allowed the identification of a moderate correlation between measurements taken during the same sampling campaign. Significant heteroscedasticity structures of the error term were also identified during modelling, stressing the need to take them into account in similar studies to reduce bias and error in the estimation of other model parameters. The measured and estimated time-weighted average levels of exposure to formaldehyde in this sector can be considered low compared to the 8h-TWA OELs of most jurisdictions (e.g 0.75 ppm) but close to the most demanding ones (e.g. 0.3 ppm). The possibility of higher short-term exposures can not be ruled out. The successful identification of several determinants of exposure to formaldehyde in the reconstituted wood panel industry will allow for better sampling strategies in this industrial sector. Furthermore, these determinants may be used in future epidemiological studies to improve prospective and retrospective exposure assessments.

2.7 Acknowledgements

The authors would like to thank Denis Bégin for his help regarding the interpretation and identification of the published literature relevant to the reconstituted wood panel industrial sector. We would also like to thank Jan-Erik Deadman for his useful and much appreciated comments on the present manuscript. This research project was funded by the Quebec Research Institute for Occupational Health and Safety (IRSST, grant number 099-011). J.L was supported by the IRSST.

2.8 References

ACGIH. (2003) TLVs and BEIs Threshold Limit Values for Chemical Substances and Physical Agents / Biological Exposure Indices. Cincinnati, OH: American Conference of Governmental Industrial Hygienists.

Bakke B, Stewart P, Eduard W. (2002) Determinants of Dust Exposure in Tunnel Construction Work. Appl. Occup. Environ. Hyg.; 17 783-796.

Buringh E, Lanting R. (1991) Exposure variability in the workplace: Its implications for the assessment of compliance. Am. Ind. Hyg. Assoc. J.; 52 6-13.

Burstyn I, Kromhout H, Kauppinen TP, Heikkila P, Boffetta P. (2000) Statistical modelling of the determinants of historical exposure to bitumen and polycyclic aromatic hydrocarbons among paving workers. Ann. Occup. Hyg.; 44 43-56.

Burstyn I, Teschke K. (1999) Studying the Determinants of Exposure: A Review of Methods. Am. Ind. Hyg. Assoc. J.; 60 57-72.

Deadman JE, Armstrong BG, Thériault GP. (1996) Exposure to 60-Hz magnetic and electric fields at a Canadian electric utility. Scand. J. Work Environ. Health; 22 415-424.

Edling C, Hellquist H, Oedkvist L. (1988) Occupational exposure to formaldehyde and histopathological changes in the nasal mucosa. British Journal of Industrial Medicine; 45 761-765.

Egeghy P, Nylander-French L, Gwin KK, Hertz-Picciotto I, Rappaport SM. (2002) Self-collected Breath Sampling for Monitoring Low-level Benzene Exposures among Automobile Mechanics. Ann. Occup. Hyg.; 46 489-500.

Francis M, Selvin S, Spear R, Rappaport SM. (1989) The Effect of Autocorrelation on the Estimation of Workers' Daily Exposures. Am. Ind. Hyg. Assoc. J.; 50 37-43.

Friesen MC, Demers P, Spinelli J, Nhu DL.(2003) From expert-based to quantitatve exposure assessment: Updating a job exposure matrix at a Söderberberg aluminum smelter. In: CARWH symposium; 2003 October 26th; Montreal: Canadian association for research on work and health; p. 45.

George DK, Flynn MR, harris RL. (1995) Autocorrelation of Interday Exposures at an Automobile Assembly Plant. Am. Ind. Hyg. Assoc. J.; 56 1187-1194.

Goyer N, Perrault G, Beaudry C, Bégin D, Bouchard M, Carrier G, et al. (2004) Impact d'un abaissement de la valeur d'exposition admissible au formaldéhyde. Montréal: Institut de recherche Robert-Sauvé en santé et en sécurité du travail (R-386).

Hall A, Teschke K, Davies H, Demers P, Marion S. (2002) Exposure Levels and Determinants of Softwood Dust Exposures in BC Lumber Mills, 1981-1997. Am. Ind. Hyg. Assoc. J.; 63 709-714.

Harrel FEJ. (2001) Regression modeling strategies - With applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY: Springer.

Herbert FA, Hessel PA, Melenka LS, Yoshida K, Nakaza M. (1995) Pulmonary effects of simultaneous exposures to MDI, formaldehyde and wood dust on workers in an oriented strand board plant. J. Occup. Env. Med.; 37 461-465.

Hornung R, Reed LD. (1990) Estimation of Average Concentration in the Presence of Nondetectable Values. Appl. Occup. Environ. Hyg.; 5 46-51.

Horvath EP, Anderson H, Pierce WE. (1988) Effects of formaldehyde on the mucous membranes and lungs. A study of an industral population. J. Am. Med. Assoc.; 259 701-707.

Imbus HR, Tochilin SJ. (1988) Acute Effect upon Pulmonary Function of Low Level Exposure to Phenol-formaldehyde-Resin-Coated Wood. Am. Ind. Hyg. Assoc. J.; 49 434-437.

International Agency for Research on Cancer. (1995) IARC Monographs on the evaluation of carcinogenic risks to humans Vol.62: Wood dust and formaldehyde. Lyon: World Health Organization.

International Agency for Research on Cancer. (in press) IARC Monograph on the evaluation of carcinogenic risks to humans Vol.88: Formaldehyde, 2-Butoxyethanol and 1-tert-Butoxy-2-propanol. Lyon: World Health Organization.

IRSST. (1985) Guide d'échantillonnage des contaminants de l'air en milieu de travail - Aldehyde formique. Montréal, QC: Institut de recherche en santé et en sécurité du travail

IRSST. (1988) Guide d'échantillonnage des contaminants de l'air en milieu de travail - Methode 216-1 -Aldehyde formique. Montréal, QC: Institut de recherche en santé et en sécurité du travail

IRSST. (1995) Analyse du formaldehyde dans l'air - Méthode 295-1. Montréal, QC: Institut de recherche Robert-Sauvé en santé et en sécurité du travail

IRSST. (2000) Sampling Guide for Air Contaminants in the Workplace - 7th edition revised and updated. Montreal, QC: Research Istitute for Occupational Health and Safety (IRSST).

Kauppinen TP, Niemelä RI. (1985) Occupational exposure to chemical agents in the particleboard industry. Scand. J. Work Environ. Health; 11 357-363.

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

Lagorio S, Iavarone I, Iacovella N, Proieto AR, Fuselli S, Bladassari LT, et al. (1997) Variability of benzene exposure amonf filling station attendants. Occup. Hyg.; 4 15-30.

Leena A, Nylander-French L, Kupper LL, Rappaport SM. (1999) An investigation of factors contributiong to styrene and styren-7,8-oxide exposures in the reinforced-plastics industry. Ann. Occup. Hyg.; 43 99-109.

Lyles RH, Kupper LL, Rappaport SM. (1997) A lognormal distribution-based exposure assessment method for unbalanced data. Ann. Occup. Hyg.; 41 63-76.

Niemelä RI, Priha E, Heikkila P. (1997) Trends of formaldehyde exposure in industries. Occup. Hyg.; 4 31-46.

Niemelä RI, Vainio H. (1981) Formaldehyde exposure in work and the general environment. Scand. J. Work Environ. Health; 7 95-100.

Olsen E, Laursen B, Vinzents PS. (1991) Bias and Random Errors in Historical Data of Exposure to Organic Solvents. Am. Ind. Hyg. Assoc. J.; 52 204-211.

OSHA. (1987) Regulatory impact and regulatory flexibility analysis of the formaldehyde standard. Washington, DC: United States Departement of Labor, Occupational Safety and Health Administration (Docket No. 225B. Exhibit No. 206).

Peretz C, Goren A, Smid T, Kromhout D. (2002) Application of Mixed-effects Models for Exposure Assessment. Ann. Occup. Hyg.; 46 69-77.

Pinheiro JC, Bates DM. (2000) Mixed-Effects Models in S and S-plus. New York: Springer-Verlag.

Raaschou-Nielsen O, Hansen J, Thomsen BL, Johansen I, Lipworth L, McLaughlin JK, et al. (2002) Exposure of Danish Workers to Trichloroethylene, 1947-1989. Appl. Occup. Environ. Hyg.; 17 693-703.

Rappaport SM, Goldberg M, Herrick RF. (2003) Excessive Exposure to Silica in the US Construction Industry. Ann. Occup. Hyg.; 47 111-122.

Rappaport SM, Lyles RH, Kupper LL. (1995) An exposure-assessment strategy accounting for within- and between-worker sources of variability. Ann. Occup. Hyg.; 39 469-495.

Rappaport SM, Weaver MA, Taylor D, Kupper LL. (1999) Application of Mixed Models to Assess Exposures Monitored by Construction Workers During Hot Processes. Ann. Occup. Hyg.; 43 457-469.

Spelter HN. (1997) Capacity, Production and Manufacture of Wood-Based Panels in North America. Madison, WI: United States Department of Agriculture, Forest Service, Forest Products Laboratory.

Stewart PA, Rice C. (1990) A source of Exposure Data for Occupational Epidemiology Studies. Appl. Occup. Environ. Hyg.; 5 359-363.

Symanski E, Chan W, Chang C-C. (2001) Mixed-Effects Models for the Evaluation of Long-term Trends in Exposure Levels with an Example from the Nickel Industry. Ann. Occup. Hyg.; 45 71-81.

Symanski E, Kupper LL, Rappaport SM. (1998) Comprehensive evaluation of long-term trends in occupational exposure: Part 1. Description of the database. Occup. Environ. Med.; 55 300-309.

Symanski E, Rappaport SM. (1994) An investigation of the dependence of exposure variability on the interval between measurements. Ann. Occup. Hyg.; 38 361-372.

Teschke K, Hertzman C, Morrison B. (1994) Level and Distribution of Employee Exposures to Total and Respirable Wood Dust in Two Canadian Sawmills. Am. Ind. Hyg. Assoc. J.; 55 245-250.

Teschke K, Marion SA, Vaughan TL, Morgan MS, Camp J. (1999) Exposure to Wood Dust in U.S. Industries and Occupation, 1979 to 1997. Am. J. Ind. Med.; 35 581-589.

Tohmura SI, Inoue A, Sahari SH. (2001) Influence of the melamine content in melamine-urea-formaldehyde resins on formaldehyde emissions and cured resin structure. Journal of wood sciences; 47 451-457.

U.S. Census Bureau. (2002) North American Industry Classification System (NAICS). Washington, DC: United States Census Bureau.

USEPA. (1998) Compilation of Air Pollutant Emission Factors AP-42 Volume 1, Fifth edition. Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards

Van Tongeren M, Gardiner KG. (2001) Determinants of inhalable dust exposure in the European carbon black manufacturing industry. Appl. Occup. Environ. Hyg.; 16 237-245.

Vinzents PS, Carton B, Fjeldstad P, Rajan B, Stamm R. (1995) Comparison of Exposure Mesurements Stored In European Databases on Occupational Air Pollutants and Definitions of Core Information. Appl. Occup. Environ. Hyg.; 10 351-354.

Weaver MA, Kupper LL, Taylor D, Kromhout H, Susi P, Rappaport SM. (2001) Simultaneous Assessment of Occupational Exposures from Multiple Worker Groups. Ann. Occup. Hyg.; 45 525-542.

Wolcott JJ, Motter WK, Daisy NK, Tenhaeff SC, Detlefsen WD. (1996) Investigation of variables affecting hot-press formaldehyde and methanol emissions during laboratory production of urea-formaldehyde-bonded particleboard. Forest Products Journal; 46 62-68.

Zimowski EF. (1986) Final report of the OSHA Health Response Team on the Wood Product Industry. Washington, DC: United States Department of Labor, Occupational Safety and Health Administration (Docket No. 225B Exhibit No. 206)

CHAPITRE III

UTILISATION DE LA SIMULATION MONTE CARLO POUR RECONSTRUIRE DES NIVEAUX D'EXPOSITION AU FORMALDÉHYDE À PARTIR DE PARAMÈTRES DE SYNTHÈSE RAPPORTÉS DANS LA LITTÉRATURE

Article soumis à "Annals of occupational hygiene"

Monte Carlo simulation to reconstruct formaldehyde exposure levels from summary parameters reported in the literature

Lavoué, J., Bégin, D., Beaudry, C., Gérin, M.*

Groupe de recherche interdisciplinaire en santé (GRIS)
Département de santé environnementale et santé au travail
Faculté de médecine
Université de Montréal
P.O. Box 6128, Main Station
Montreal (QC) Canada H3C 3J7

Keywords: occupational exposure assessment, mixed-effects models, lognormal model, particle board, oriented-strand board, medium density fibre board

*** Author to whom correspondence should be addressed.**

3.1 Abstract

Objectives

This study presents a procedure allowing the numerical synthesis of exposure data reported in different ways in the literature, including summary parameters and single measurements. The procedure was applied to literature regarding formaldehyde exposure in the reconstituted wood panels industry, including Oriented-Strand Board (OSB), Medium Density Fibre board (MDF) and Particle Board (PB).

Methods

For each publication providing summary parameters we estimated geometric means (GM) and geometric standard deviations (GSD) by assuming lognormality of exposure levels. Monte Carlo simulation was performed to re-create datasets from the sample sizes and estimated GMs and GSDs, allowing their subsequent formatting together with the single measurements. The precision and bias of the methods used to estimate GMs and GSDs were evaluated.

Results

Altogether, the 13 articles included in our study yielded a final database of 874 data, of which 732 were simulated. For both area and personal data, exposures corresponding to MDF and PB were similar while OSB levels were lower. The most recent available personal levels (1985-1994) were highest in PB for jobs performed in the vicinity of the press (GM=0.63 mg/m^3). Corresponding area levels were highest for PB in the main production zone (GM=0.43 mg/m^3). Mixed-effects models fitted to area PB data explained 38% of the total variability. A 6 fold decrease in exposures from 1965 to 1995

was estimated. Replication of the simulation process yielded relative standard deviations of the calculated GMs and GSDs between 10 and 20%. The relative biases of the methods used to estimate GMs and GSDs varied across methods and decreased with higher sample sizes (from ~15% for n=5 to less than 5% for n=30, in absolute value). The precision also varied across methods and improved with higher sample sizes (from ~30% for n=5 to ~10% for n=30).

Discussion

This methodology constitutes a new meta-analysis tool that should improve the interpretation of industrial hygiene literature data, but needs to be further validated.

3.2 Introduction

In many exposure assessment situations, either because of a lack of resources for prospective sampling or because of the need to characterize past exposures, literature is the main source of information. In addition to information describing exposure generating processes and tasks, quantitative exposure measurements are reported in a number of studies. Several authors have underlined the limits of the use of published data in the scope of risk analysis (Caldwell et al., 2001;Marquart et al., 2001;Money and Margary, 2002;Tielemans et al., 2002). Hence, some studies lack details about determinants of exposure, about characteristics of the study population, or about statistical parameters. It then becomes difficult to integrate all the available numerical data, especially when numerous studies are available and different statistical parameters are reported. Thus, most literature analyses are reported in the form of tables presenting the results of each individual study and expert opinion is used to make a global assessment. On the other hand, quantitative assessments of exposure, particularly for occupational epidemiology, become increasingly preferred over qualitative or semi-quantitative assessments (Ahrens and Stewart, 2003).

Recently, Caldwell et al., in a review on solvent exposure, calculated averages of the reported arithmetic means weighted by the associated sample sizes. The authors had to exclude numerous studies that did not report results as arithmetic means (Caldwell et al., 2000). Tielemans et al. (Tielemans et al., 2002), and Money and Margary (Money and Margary, 2002) have proposed theoretical frameworks for the use of exposure data available in the published literature, mainly by presenting quality criteria regarding the internal and external validity of the studies.

The main objective of our study was to develop a method to summarize exposure data reported in different ways in the literature by estimating common statistical exposure parameters from different types of exposure metrics and by simulating exposure data from the estimated parameters to allow their combined analysis with single measurements. This paper presents in details the proposed procedure and its application to formaldehyde exposure data in the reconstituted wood panels industry. This study should not be regarded as a literature review of formaldehyde exposure in this sector, which would include information of a much wider scope than the summary of exposure levels presented here.

The reconstituted wood panels industry is part of the larger Veneer, Plywood and Engineered Wood Product Manufacturing group of the North American Industry Classification System (NAICS). It includes several processes that can be classified as either plywood products or composition boards. The processes included in this study are limited to particle board (PB), medium density fibre board (MDF), and oriented strand board (OSB), which all belong to the composition board category. Occupational exposure to formaldehyde in this industry comes mainly from the degradation of the resins during and after the pressing operation.

3.3 Methods

An exhaustive literature review of formaldehyde exposure levels published before and up to 2001 was conducted in the reconstituted wood panels industry. The study was limited to PB, MDF, and OSB. All publications reporting formaldehyde levels measured in workplaces in this industrial sector were retained for further analysis. The exposure levels were formatted into a relational database (the 'initial' database, see figure 1) and allocated to specific jobs/work zones based on the work by Lavoué et al. (Lavoué et al., 2005). The personal measurements were also classified in exposure groups identified in the same study: group 1 includes administration and foremen, group 2 includes laboratory technicians, maintenance workers, and cleaners, group 3 includes press operators, assistant press operators, finishers, shippers, and group 4 includes floaters and press-miscellaneous tasks. The area measurements were classified in 7 zones: Raw materials receiving – Chip preparation, Resin production – storage, Main production – press, Finishing, Storage – shipping, Operator booth and Other departments (non production areas).

3.3.1 Step 1 of the procedure: Calculation of common statistical parameters

The initial database contained two types of exposure data: single exposure concentrations, each representing one measurement (SM record, see figure 1), and sets of summary parameters, each set summarizing a number of measurements (SS record, see figure 1). Each record of this database was associated with values for different variables such as work zone, job group, source article. The SS records also contain some of the following parameters: sample size (N), arithmetic mean (AM), arithmetic standard deviation (ASD), geometric mean (GM), geometric standard deviation (GSD), range (R, with minimum a and maximum b), median (M), or an empirical percentile of the distribution of the measurements ($\hat{P}_x$, with $\frac{x}{100}$ the cumulative probability associated with the value of the percentile). The first step of the methodology consisted, for each of the SS records without GM and GSD, in estimating them from available parameters. This was performed using the following equations:

Estimation of GM from GSD and AM:

$$GM = \frac{AM}{\exp\left(\frac{(\ln(GSD))^2}{2}\right)} \quad (1)$$

Estimation of GSD from GM and AM:

$$GSD = \exp\sqrt{2 \times \ln\left(\frac{AM}{GM}\right)} \quad (2)$$

Estimation of GSD from AM and ASD

$$GSD = \exp\sqrt{\ln\left(1 + \frac{ASD^2}{AM^2}\right)} \quad (3)$$

Estimation of GM from AM and ASD

$$GM = \frac{AM}{\sqrt{1 + \frac{ASD^2}{AM^2}}} \quad (4)$$

Estimation of GM from [a,b]

$$GM = \exp\left(\frac{\ln(a)+\ln(b)}{2}\right) \qquad (5)$$

Estimation of GSD from GM and $\hat{P}_x$ (see appendix for details)

$$GSD = \exp\left(\frac{\ln(\hat{P}_x)-\ln(GM)}{Z_x}\right) \qquad (6)$$

with Z_x the x^{th} percentile of the standard normal distribution.

Estimation of GSD from [a,b] (see appendix for details)

$$GSD = \exp\left(\frac{\ln(b)-\ln(a)}{W_{median}}\right) \qquad (7)$$

with W_{median} the theoretical median standardized range (see appendix)

Equations 1 to 4 are based on the theoretical correspondence between the different parameters characterizing the lognormal distribution (Rappaport, 2000). Equation 5 is based on the fact that the expected values of the maximum and minimum of a sample from a normal distribution are symmetrical around the mean of the distribution (Zwillinger and Kokoska, 2000). Equation 6 is based on the fact that the x^{th} percentile of a sample from a normal distribution is a good approximation of the x^{th} percentile of the parent distribution (see appendix). The assumptions linked to equation 7 are detailed in the appendix.

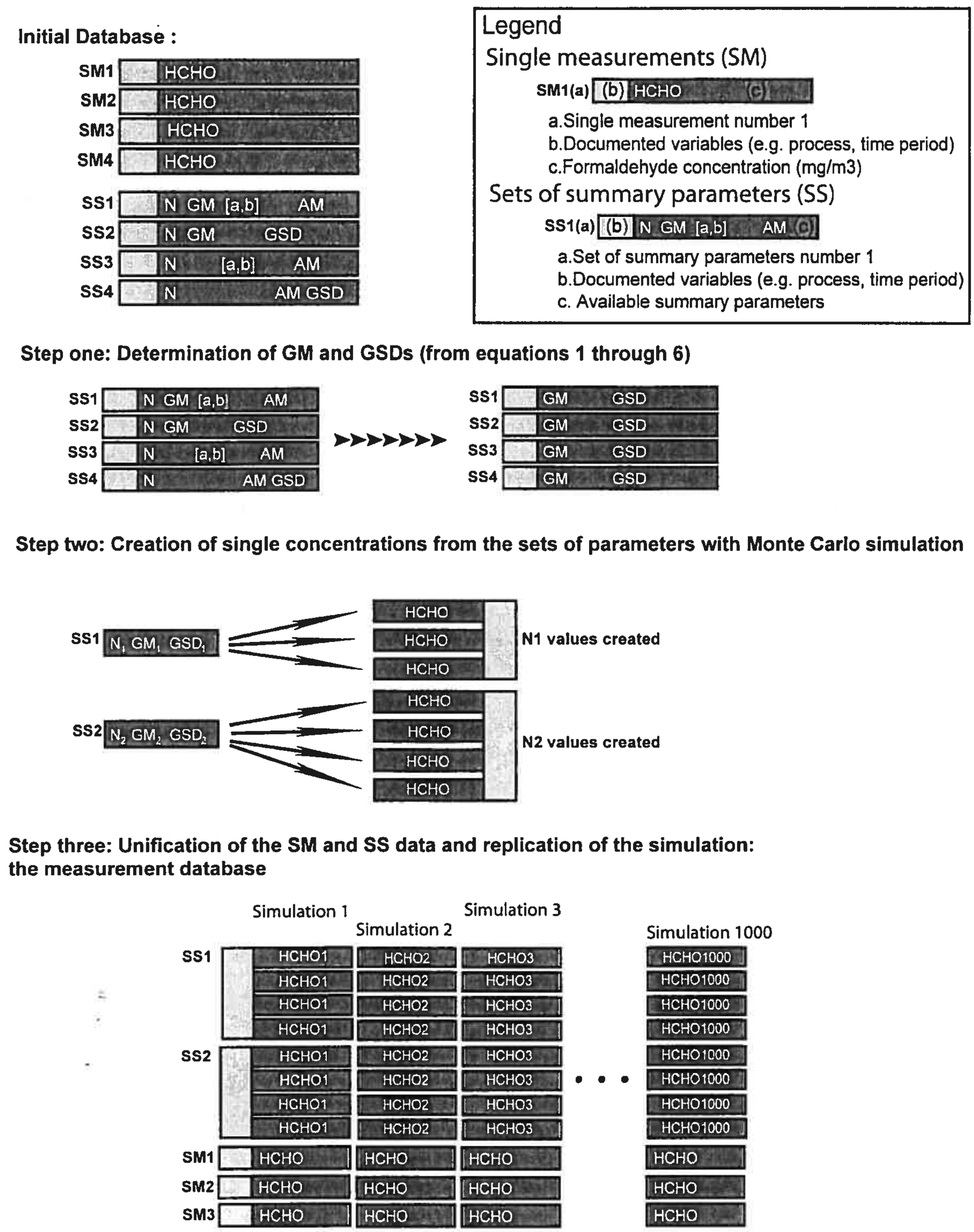

Figure 1: Conceptual schema for the creation of the measurement database from literature data

If GM was not available but the sample median was, the latter was used as an estimate of GM. If neither GM nor the median was reported, GM was estimated from AM and GSD (equation 1) or from AM and ASD (equation 4). GM was estimated from the range (equation 5) only when it was the only available parameter.

GSD, if not available, was determined from AM and ASD (equation 3), from AM and GM (equation 2), or from GM and $\hat{P}_x$ (equation 6). When none of the previous methods could be used, GSD was estimated from the theoretical median of the standardized range (equation 7, see appendix). When the estimation of GSD was not possible, the median of the GSDs estimated for the other sets of measurements was used.

3.3.2 Steps 2 and 3 of the procedure: Simulation of exposure data and creation of the measurement database

In step two of the methodology, each SS record was replaced with a number of concentrations equal to the reported sample size, drawn at random from a lognormal distribution with the GM and GSD estimated for that particular SS. In step three of the methodology, a new measurement database was created by combining the original single measurements and the simulated measurements. The procedure was repeated 1000 times creating 1000 replicas of the measurement database. Personal and area measurements were then separated.

3.3.3 Analysis of the measurement database

For personal and area concentrations, GMs and GSDs stratified by process and by period of time were calculated for each of the 1000 replicas of the analysis database. For each stratum, the median of the 1000 resulting GMs was used as the exposure metric and the relative standard deviations (RSD) of the 1000 GMs was computed as a measure of variability caused by the simulation.

For the personal measurements 90% of the data were in the PB process. Moreover 70% of the PB data corresponded to the 'unknown' category of the job classification and the data for which the job was known were almost entirely (87 out of 96 data) in the most recent time period category (1984-1995). Therefore we restricted the analysis of personal data to calculation of stratified GMs for the most populated categories of process, job group, and time period. For the area measurements, PB also represented ~90% of the data, but inside the PB process data were approximately evenly distributed across categories of work-zone and time period. Therefore statistical modelling was used to analyse the area PB data.

A total of 200 replicas of the area PB data were randomly selected among the 1000 initially created and linear mixed-effects models were fitted to the log-transformed concentrations. The models were constructed in a manual forward stepwise procedure similarly to those described in Lavoué et al. (Lavoué et al., 2005), using the Bayesian information criterion (BIC) as a discrimination tool. Median values of the BIC were used to build a model common to the 200 datasets. The fixed effects tested in the models included time period (1965-1974, 1975-1984, 1985-1994), work zone and sampling duration (<1 hour, 1-6 hours, > 6 hours, unknown). The source of the data (i.e. article) was tested as a random effect to evaluate the extent of the difference between sources after taking into account the fixed effects. Since the measurements simulated from SS records were generated with differing variances, a different residual variance for each record was modelled. In addition, single measurements were allowed to vary differently from source to source. The parameters of the final model were computed as the mean of the 200 restricted maximum likelihood (REML) estimates resulting from fitting the common model to the 200 replicas of the area PB data. RSDs were also computed to assess the variability of the results. Internal validation of the models was conducted graphically for a random subsample of 10 models.

During the creation of the measurement database, personal and area single measurements reported as below a limit of detection were replaced with the limit of detection divided, respectively, by $\sqrt{2}$ and 2, based on their global GSD (Hornung and Reed, 1990). Ranges including a limit of detection (e.g. [<0.2-3]) were used for estimation of GM and GSD only when R was the only available parameter because of the potential for important bias. In this case, the lower limit of R was replaced by the limit of detection divided by $\sqrt{2}$ for personal measurement or 2 for area measurements in equations 5 and 7.

3.3.4 Partial validation of the equations used to estimate GMs and GSDs

In order to provide some insight on the validity of the estimation tools presented above (equations 1-7), a limited simulation study was conducted. A total of 500 samples of sizes 5, 10, and 30 were generated from a lognormal distribution with GM=1 and GSD=2.5. For each generated sample, the sample GM and GSD were calculated, as well as all the parameters used in equations 1-7. From these parameters, four estimations of the GMs and GSDs were, in turn, determined using the estimators in equations 1-7. For each estimator, the relative bias was calculated as the average relative difference between the estimate and the sample GM or GSD, and the relative precision as the relative standard deviation of the bias. The relative bias and precision were stratified by sample size.

3.4 Results

3.4.1 Simulation procedure

A total of 23 publications were abstracted into the relational database. Ten publications were excluded from the analysis: 2 only reported results from semi-quantitative colourimetric tubes; one described a plant located in a tropical area; the sample size could not be estimated in two publications; one was excluded because the facilities surveyed did not use a formaldehyde-based resin to bond the wood-particles; in another the results summarized a mix of area and personal measurements, which could not be distinguished. Finally three publications were excluded because they used the same dataset presented in a paper of wider scope; this left 13 publications for further analysis.

Table 1 shows, for each of the 13 publications included in the analysis, the number of single concentrations (SM records) and of summarized sets of measurements (SS records) provided. Among the journal articles retained for analysis, 3 presented exposure data taken by different organizations in several facilities in a country (Kauppinen and Niemelä, 1985;Niemelä et al., 1997;Triebig et al., 1989). Three were epidemiological studies, mainly on the respiratory effects of formaldehyde exposure (Herbert et al., 1994;Horvath et al., 1988;Imbus and Tochilin, 1988). One article reported the summary of formaldehyde levels in the French occupational exposure database COLCHIC (Carton, 1995) while another reported results of different sampling methods for formaldehyde in a facility manufacturing wood panels (Wentrup et al., 1986).

Table I: Number of single measurements and summarised sets of measurements associated with each study entered in the database.

Reference	Publication type	Number of unique measurements	Number of sets (number of measurements summarized)
(Sussell, 1995)	HHE [(a)]	41	-
(Lee, 1988)	HHE	4	-
(Mortimer, 1982)	HHE	-	1(10)
(Horvath et al., 1988)	Article	3	1(109)
(Kauppinen and Niemelä, 1985)	Article	14	17(287)
(Imbus and Tochilin, 1988)	Article	15	-
(IARC, 1995)	IARC[(b)] Monograph	-	2(40)
(Carton, 1995)	Article	-	2(81)
(Triebig et al., 1989)	Article	-	3(581)
(Niemelä et al., 1997)	Article	-	3(67)
(Herbert et al., 1994)	Article	10	-
(Wentrup et al., 1986)	Article	24	-
(Centaur Associates, 1986)	Report	5	20(82)

(a): NIOSH Health Hazard Evaluation report
(b): International Agency for Research on Cancer

Neither GM nor GSD were reported in any of the analysed papers. AM was available in 44 of the 49 measurement sets while ASD appeared in 1. M was available in 26 sets. R was reported in 42 sets and one publication, representing two measurement sets, reported the sample's 95th percentile. This resulted in 53% of the estimated GMs determined from M, 41% from AM and R (GSD estimated from equation 7 then GM estimated from equation 1), 4% from AM and GSD set to the median of the other GSDs (equation 1), and 2% from AM and ASD (equation 4). The GSDs were estimated from the different parameters in the following proportions: R for 59% (equation 7), AM and M for 26% (equation 2), median of the other GSDs for 8%, AM and $\hat{P}_x$ for 4% (equation 6), AM and ASD for 2% (equation 3).

Ten of the 49 measurement sets represented less than 10 measurements each while 20 sets represented more than 10 measurements each. For 19 of the sets, the sample size was not provided explicitly in the publication but could be estimated from other information provided (e.g. sample size of other similar measurement sets in the source article).

Regarding the final datasets, including simulated and actual measurements, the analytical and sampling methods, and a crude description of the sampling strategy were reported respectively in 10 and 7 of the 13 analysed publications. Most of the reported sampling strategies were of the type 'placed in areas representative of normal exposure conditions' or 'presumed maximal exposure jobs'. For 33% of the measurements, information about the associated work-zone or job was not provided. The sample duration was not reported for 20% of the measurements. Twenty percent of measurement durations were below 1 hour, 52% were between 1 and 6 hours, and 8% were greater than 6 hours.

3.4.2 Analysis of the measurement database

Personal measurements

The final database contains 376 personal measurements, of which 320 were simulated from 11 SS records. The measurement sets comprised a median of 22 measurements (range 5-109). The personal database represents data from 8 publications. None of the 56 single measurements were reported under a limit of detection. The personal measurements had a median GM of 0.53 mg/m^3 (RSD=3%) and a median GSD of 2.80 (RSD=3%). Table 2 shows the median GMs, 90th percentiles, and their respective RSDs for personal measurements after stratification by job group, time period and process, for strata with more than 5 values.

Table 2: Median GMs and 90th percentiles in mg/m^3 of personal measurements after stratification by job group, time period, and process.

Process PB			
Time period	Job group (sample size)	Median GM (RSD[(A)])	Median 90th percentile (RSD)
1975-1984	G3[(B)] (9)	0.38 (18%)	1.28 (25%)
	Unknown (77)	0.96 (9%)	3.50 (17%)
1985-1994	G1/G2[(D)](46)	0.17 (13%)	0.47 (21%)
	G3 (41)	0.63 (7.6%)	1.23 (10%)
	Unknown (154)	0.58 (5%)	1.28 (7%)
All	All (332)	0.56 (4%)	1.42 (7%)
Process MDF			
1985-1994	G2 (14)	0.23[(C)]	0.47[(C)]
	G3 (22)	0.37[(C)]	0.84[(C)]
All	All (42)	0.41 (2%)	1.46 (26%)
Process OSB			
All	All (2)	0.05[(C)]	0.04[(C)]
Stratification by time period for all processes			
1965-1974	All (5)	1.85 [(C)]	3.03 [(C)]
1975-1984	All (86)	0.87 (8%)	2.72 (18%)
1985-1994	All (285)	0.45 (3%)	0.99 (5%)

(A) Relative standard deviation of the 1000 estimates; (B) group 1 (G1) includes administration and foreman, group 2 (G2) includes laboratory technician, maintenance worker, and cleaner, group 3 (G3) includes press operator, assistant press operator, finisher, and shipper; (C) Non variable parameters since they were estimated from single measurements (SM records, see figure1); (D) all data coming from one source, labelling the task as 'chipboard pasting', which we interpreted as corresponding to a mix of G1 and G2.

Area measurements

The final database contains 498 area measurements, of which 412 were simulated from 38 SS records which comprised a median of 21 measurements (range 4-61). The area database represents data from 10 publications. Twelve of the 86 single measurements were reported under a limit of detection. The area measurements had a median GM of 0.79 mg/m^3 (RSD=3%) and a median GSD of 4.36 (RSD=3%). The median GMs after stratification by process were: MDF 0.22 mg/m^3 (N=32, RSD=10%, data from 1975 to 1994), PB 0.99 mg/m^3 (N=443, RSD=4%) and OSB 0.05 mg/m^3 (N=23, 23 original single measurements, data from 1985 to 1994). The median GMs after stratification by time period were: 1965-1974: 2.52 mg/m^3 (N=136, 5 RSD=5%), 1975-1984: 0.68 mg/m^3 (N=295, RSD=4 %) and 1985-1994: 0.14 mg/m^3 (N=67, RSD=10%).

There were 443 data available for modelling of PB area data. The fixed effects of the final models explained an average of 38% of the total variability (the inter-quartile interval of the 200 obtained R^2 is [0.36-0.39]). A summary of the mean parameters of the model along with their RSDs for the 200 replications are presented in Table 3. Table 4 presents the GMs for the different time periods and work zones predicted from the models by using the average coefficients presented in table 3. A significant within-source correlation was found in the area data, with an average intra-class correlation coefficient of 0.23 (RSD=26%). The variable modelling a different variability for the different sets of measurements was refined in a 'post hoc' manner by aggregating categories with estimates less than 10% different from each other. This aggregation was performed because the important initial number of categories yielded a non definite positive approximate variance-covariance matrix in the REML optimization. The refined variable (from 42 to 9 categories) improved the model fit significantly in terms of BIC. This corresponds to within-source GSDs varying from 1.36 to 5.79 (RSDs not shown). Graphical assessment of 10 randomly chosen models yielded satisfactory conformity to the model assumptions.

Table 3: Summary of the final area model

Fixed effect	Estimate[(b)]	RSD(%)[(c)]
Intercept[(a)]	0.05	50
Finishing	-0.63	13
Other – Unknown	-1.24	10
Raw materials – Chip preparation	-0.90	20
Storing-shipping	-0.86	29
1965-1974	0.99	8
1985-1994	-0.91	25
% total variance explained by the fixed effects 38%		
Within – source correlation coefficient 0.23		

(a) The intercept represents the estimated AM of the log-transformed concentrations (mg/m^3) for the following combination of variable: main production zone, and period 1975-1984. Estimates of the AM of log-transformed concentrations for other combinations of levels of the variables are obtained by adding the appropriate coefficients, for example adding 0.99 to the intercept to change the period to 1965-1974; (b) Arithmetic mean of the estimates from the 200 fitted models; (c) Relative standard deviation of the estimates from the 200 fitted models

Table 4: Predicted GMs (mg/m^3) for PB area measurements stratified by time period and work zone

Work zone	1965-1974	1975-1984	1985-1994
Main production zone	2.83	1.05	0.42
Raw materials – Chip preparation	1.15	0.43	0.17
Finishing	1.51	0.56	0.23
Storing-shipping	1.19	0.44	0.18
Other - unknown	0.82	0.30	0.12

3.4.3 Partial validation of the equations used to estimate GMs and GSDs

Table 5 shows the relative bias and precision of the estimates of GM and GSDs obtained with equations 1-7 compared to the sample GMs and GSDs, for sample sizes 5, 10, and 30.

Table 5: Bias and precision associated with the estimation of GM and GSD from other summary parameters

	Relative bias (%)			Relative precision (%)		
	N=5	N=10	N=30	N=5	N=10	N=30
Sample GM(A)	14	4	2	55	31	17
GM1(B)	5	5	1	38	22	12
GM2(C)	-18	-9	-4	18	10	9
GM3(D)	8	8	5	19	10	7
GM4(E)	4	3	3	30	30	30
Sample GSD (F)	-1	-1	0	35	22	12
GSD1 (G)	-12	-6	-2	14	8	6
GSD2 (H)	-19	-14	-8	26	15	12
GSD3 (I)	9	4	2	13	12	11
GSD4 (J)	-24	-13	-5	21	14	11

(A) Relative bias and precision of the sample GM compared to the true GM; (B) GM estimated from M; (C) GM estimated from equation 1 and GSD in equation 1 estimated from equation 7; (D) GM estimated from equation 4; (E) GM estimated from equation 5; (F) Relative bias and precision of the sample GSD compared to the true GSD; (G) GSD estimated from equation 2; (H) GSD estimated from equation 3; (I) GSD estimated from equation 7; (J) GSD estimated from equation 6

3.5 Discussion

3.5.1 Information available in the literature

The analysis of the literature database revealed several limitations, in the context of exposure assessment, of the information reported in the publications, similar to those described by Money and Margary, and Caldwell et al.(Caldwell et al., 2001;Money and Margary, 2002). Hence although only looking systematically at crude information (statistical summary parameters, job / work-zone sampled, sample size, crude sampling strategies and analytical methods), our results show high percentages of data for which this information was lacking or not adequate.

Most available area exposure data came from the main-production and finishing work-zones (74% of the data for which this information was available). Likewise, only three job strata had more than 10 measurements available (Press operator, finisher, and chipboard pasting, representing 71% of the documented data). These results underline a potential for misinterpretation of exposure levels for which no information on the job / workplace is provided.

3.5.2 Statistical modelling

Our simulation procedure yielded area and personal datasets different from exposure datasets commonly described in the literature only in that they contain multiple concentration data 'columns' due to the replication of the simulation process. While these datasets could have been interpreted using several numerical analysis methods, we initially had planned to use linear mixed-effect models, which are increasingly considered as the state of the art analysis tool in occupational hygiene (Burdorf and Van Tongeren, 2003). However, because personal data were severely unbalanced and jobs were unknown for a majority of data we rather performed a stratified analysis of these measurements in categories with enough data. Mixed effects models were used to analyse the area PB data, in which data was spread approximately evenly between categories of the different variables.

The fixed-effects of the models constructed for the area data explained an important fraction of the variability of the simulated data (38%), which indicates the usefulness of such an analysis even with crude information on potential determinants of exposure.

The models identified a clear time trend in the data, with an estimated 5 to 6 fold reduction in exposure levels between the periods 1965-1974 and 1985-1994. The main production area was also shown as the highest exposure zone, with the zones corresponding to post- or pre-pressing operations corresponding to lower exposure levels. These results are very similar to those reported by Lavoué et al., who analysed formaldehyde exposure data measured in 12 plants manufacturing wood panels in Quebec (Lavoué et al., 2005). The 'unknown-other' category in our study corresponds to low exposure levels compared to the other zones. This is due to the fact that a significant proportion of these data were classified as 'other' in the initial database, and, as such, probably corresponds to very low exposure locations such as administration or the exterior of the facility.

Moderate correlation was found between concentrations coming from the same publication for area measurements when simple random-effect models were fitted to the data. This suggests that there were differences among the various sources due to undocumented factors. Thus, significant bias could be present in any assessment based on a small fraction of the available publications if sufficient information is not provided.

3.5.3 Exposure levels estimated / predicted in our study

Personal measurements for the most recent available period (1985-1994) showed formaldehyde levels around 0.60 mg/m^3 for the most exposed jobs (i.e job group 3 including press operators) in the PB process and slightly lower levels in the MDF process (0.40 mg/m^3). Other job groups corresponded to lower exposures, by a factor of

1.5 to 2. The presence of a decrease of exposure over time was less clear in the case of personal measurements than that observed in the area models. There were only 5 values for 1965-1974, and table 2 shows that although exposures in the 'unknow' job category for PB decreased from 1975-1984 to 1985-1994, they increased in the case of job group 3. Non stratified GMs nevertheless show a trend similar to that observed in the area models. The GM corresponding to the 'unknown category' is higher than both group 2 and 3, which suggests that most of the unknown data come from high exposure jobs, most probably a mix of group 3 and 4 (Group 4 corresponded to the highest exposures in the study of Lavoué et al.). This confirms the risk of exposure misclassification when interpreting data with little ancillary information.

Area measurements were markedly lower in OSB (GM=0.05 mg/m^3) than in the two other processes. This is not unexpected and is due to the fact that phenol-formaldehyde resins used in OSB are more resistant to hydrolysis than urea-formaldehyde and melamine-urea-formaldehyde resins, which are used in MDF and PB. MDF concentrations (GM=0.22 mg/m^3) also seemed lower than PB levels, but did not differ from PB levels after correction by time period and zone (most of the MDF levels correspond to 1984-1995 and the 'Other-unknown' zone). PB concentrations, predicted for the most recent period, show formaldehyde GMs at 0.42 mg/m^3 in the main production zone, and between 0.12 mg/m^3 and 0.23 mg/m^3 in the other zones.

The personal and area exposure levels reported above are very similar to predictions made from the models presented in the study of Lavoué et al. in the same industrial sector in Quebec for the year 1990 and for what the authors labelled 'governmental data' as opposed to 'research' data (calculations not shown). These predictions were generally less than 20% different from the GMs presented in table 2 and 5. In their study, Lavoué et al. found that levels taken by governmental hygienists were consistently higher than those measured by a research team in the same facilities after correction for other determinants of exposure. They concluded that 'governemental' data probably corresponded to 'worst case' sampling strategies. Our observations would therefore imply that such bias also exist in the data we extracted from the literature. Unfortunately, only 3 of the 7 publications reporting a crude description of the sampling strategy stated explicitly that it was a 'worst-case' strategy; they represented 20 % of all the measurements. Moreover these data were not higher than the rest of the data after correction for process, time period and job/zone. Thus it remains unclear how the different and mostly unknown sampling strategies used in the literature influenced our results.

The data in our study corresponded to variable sampling times (20% <1hour, 52% 1-6hours, 8%>6 hours, 20% unknown). Several authors have observed an influence of the sampling duration on exposure levels, generally a decrease in exposures when the duration increases (Kolstad et al., 2005;Lavoué et al., 2005;Raaschou-Nielsen et al., 2002). Including a nominal variable corresponding to the sampling time didn't improve the fit of the area models. In the case of personal data, median GMs stratified only by the sample time category showed increasing exposures with increasing sample time : <1 hour (0.36 mg/m^3), 1-6 hours (0.52 mg/m^3), >6 hours (0.77 mg/m^3), unknown (0.76 mg/m^3). However this trend was inversed when using MDF data from 1984-1995 stratified by job group and category of sample time. It is therefore difficult to conclude on the existence of an influence of the sample time in our data. This may be due to the classification scheme we used, which was warranted by the lack of precision in the information provided in the articles.

3.5.4 Validity of the simulation methodology

Monte Carlo simulations used to merge aggregated and single measurements yielded moderately variable results. Indeed the RSDs of the summary parameters (GMs and GSDs) calculated from the 1000 replications are between 10% and 20%. These results are confirmed by the similar variability observed in the parameters of the mixed effects models fitted to the simulated data and point to the possibility of using less replications in future studies.

The first main limitation of our methodology regards the calculations used to estimate distribution parameters from very limited information. The simulation study we performed in order to evaluate the accuracy of the different estimators in equations 1-7 showed moderate bias and precision for most methods, with errors decreasing when the sample size increased. Indeed, the maximum bias for the estimation of GMs decreased (in absolute value) from 18% for a sample size of 5 to 5% for a sample size of 30. Table 5 also shows that the different methods used are associated with different biases and precisions (e.g. GM2 in table 5 is negatively biased whereas the other estimators are positively biased). Since the set of available parameters is often different between sources of data, the error introduced by the estimation methods will therefore also vary across sources. We chose an average value of GSD (2.5) to perform the simulation (Buringh and Lanting, 1991). Since the GSD we estimated during this study were slightly lower (median values of respectively 1.7 and 2.2 for personal and area measurements), the actual error in our estimates should be lower than that observed in Table 5. Altogether, we believe these results are promising since the accuracy of the estimation methods seems acceptable compared to the uncertainty generally associated with occupational exposure assessment. However, more extensive simulation studies should be conducted to fully evaluate the methods presented here. Such studies could permit the determination of a prioritizing scheme for the choice of specific equations and allow the specification of sample sizes below which some methods should not be used.

The second main limitation of our methodology regards the assumption that every set of data summarized in a study follows a lognormal distribution. This assumption is central in our methodology because it is used in the methods of estimation of GM and GSD and during the simulation process. It is now well established that airborne concentrations of contaminants in the workplace tend to follow, at least approximately, a lognormal distribution, and most methods of interpretation of exposure levels rely heavily on this assumption (Mulhausen and Diamano, 1998;Rappaport, 2000). We believe there was little risk of important departure from the assumption of lognormality in most of our data since each set of measurements came from the same occupational setting and was further characterized by process, time period and job/zone. However, we did not quantify the robustness of our simulation method to such departures, and recommend that other studies be conducted to evaluate it.

While we believe that the procedure we propose permitted to recreate exposure data representative of the data summarized in the articles we analysed, their representativeness of occupational exposure to formaldehyde in the general population cannot be assessed directly. It depends on the validity of the publications themselves. In our study, all data available in the literature was retrieved, and only very crude criteria were used to discard irrelevant data before analysis. In particular, the authors did not use the criteria proposed by Tielemans et al., according to which a large part of the data summarised in this study would have been excluded (Tielemans et al., 2002). We chose to include as much data as possible in order to maximize the dataset available to test the feasibility of our methodology and the variability caused by the aggregation of different studies. However, it is plausible that the inclusion of multiple studies in the analysis permitted to compensate to some extent for study-specific biases. Moreover the analysis of the simulated area datasets with statistical models yielded plausible quantitative results regarding the influence of the different variables and the models explained an important percentage of the variability of the simulated data. Finally, the observations drawn from the analysis of the simulated data were similar to the results obtained in an analysis made on an external source of exposure data in the same industrial sector (Lavoué et al., 2005).

3.6 Conclusion

The new method we used in this study allowed the inclusion in the analysis of data that would have been discarded if conventional methods (e.g. average of the means weighted by the sample size or variance) had been used. Moreover, single measurements could be analysed along with summarized measurements. In addition, the equations and assumptions used to simulate the exposure data are explicit and permit the computational aggregation of all available data. This ensures reproducibility of the results by researchers other than the initial assessor(s) and permits quantitative assessment of the uncertainty, as opposed to the "black box" of expert-only assessments. Finally the database-like format resulting from the simulation procedure enables to produce the same kind of analyses one would conduct on a standard exposure dataset, in particular the use of statistical models to explore potential exposure determinants. The authors would like to emphasize that such 'all numerical' analysis should not be taken as a replacement for expert analysis of the literature but merely as a tool for industrial hygiene meta-analyses, available to help the exposure assessor to integrate in a consistent and transparent way the results available from several exposure studies. Further analysis of this methodology, by the use of quality criteria to down-weigh or exclude some data, or by using simulations to assess the accuracy of the estimation equations, will allow for a better appraisal of the potential of this methodology for exposure assessment. Similar studies in other industrial settings and for other contaminants are also warranted to assess the generalizability of our results.

3.7 Acknowledgements

The authors wish to thank Dr. Patricia Stewart of the U.S. National Cancer Institute for her helpful comments during the initial drafting of this manuscript. J.L was supported by the Institut de Recherche Robert-Sauvé en Santé et en Sécurité du travail.

3.8 References

Ahrens W, Stewart PA. (2003) Retrospective exposure assessment. In: Nieuwenhuijsen M, Editor. Exposure assessment in occupational and environmental epidemiology. New York, NY: Oxford University Press. p. 103-118.

Blom G. (1958) Statistical Estimates and Transformed Beta Variables. New York, NY: John Wiley.

Burdorf A, Van Tongeren M. (2003) Variability in Workplace Exposures and the Design of Efficient Measurement and Control Strategies. Ann. Occup. Hyg.; 47 95-99

Buringh E, Lanting R. (1991) Exposure variability in the workplace: Its implications for the assessment of compliance. Am. Ind. Hyg. Assoc. J.; 52 6-13

Caldwell DJ, Armstrong TW, Barone MJ, Suder JA, Evans MJ. (2001) Lessons Learned While Compiling a Quantitative Exposure Database from the Published Litterature. Appl. Occup. Environ. Hyg.; 16 174-177

Caldwell DJ, Armstrong TW, Barone NJ, Suder JA, Evans MJ. (2000) Hydrocarbon Solvent Exposure Data: Compilation and Analysis of the Litterature. Am. Ind. Hyg. Assoc. J.; 61 881-894

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

Centaur Associates. (1986) Case Studies of Formaldehyde Exposure Control in Six Industries - Prepared for the Occupational Safety and Health Administration under contract with the Office of Regulatory Analysis. Washington, DC: United States Department of Labor, Occupational Safety and Health Administration (OSHA Docket No. H-225, Exhibit No. 85-116).

Hartley HO. (1942) The range in random samples. Biometrika; 32 334-348

Herbert FA, Hessel PA, Melenka LS, Yoshida K, Nakaza M. (1994) Respiratory consequences of exposure to wood dust and formaldehyde of workers manufacturing oriented strand board. Arch. Environ. Health; 49 465-470

Hornung R, Reed LD. (1990) Estimation of Average Concentration in the Presence of Nondetectable Values. Appl. Occup. Environ. Hyg.; 5 46-51

Horvath EP, Anderson H, Pierce WE. (1988) Effects of formaldehyde on the mucous membranes and lungs. A study of an industrial population. J. Am. Med. Assoc.; 259 701-707

IARC. (1995) IARC Monographs on the evaluation of carcinogenic risks to humans Vol.62: Wood dust and formaldehyde. Lyon: International Agency for Research on Cancer, World Health Organization.

Imbus HR, Tochilin SJ. (1988) Acute Effect upon Pulmonary Function of Low Level Exposure to Phenol-formaldehyde-Resin-Coated Wood. Am. Ind. Hyg. Assoc. J.; 49 434-437

Insightfull. (2001) S-plus 6 for Windows. Seatle, WA: Insightfull corporation.

Kauppinen TP, Niemelä R. (1985) Occupational Exposure To Chemical Agents In The Particleboard Industry. Scand. J. Work Environ. Health; 11 357-363

Kolstad HA, Sonderskov J, Burstyn I. (2005) Company-Level, Semi Quantitative Assessment of Occupational Styrene Exposure when Individual Data are not Available. Ann. Occup. Hyg.; 49 155-165

Lavoué J, Beaudry C, Goyer N, Perrault G, Gérin M. (2005) Investigation of past and current exposures to formaldehyde in the reconstituted wood panels industry in Quebec. Ann. Occup. Hyg.; 49 587-600

Lee SA. (1988) Health Hazard Evaluation Report No. HETA-87-309-1906, Louisiana-Pacific, Corporation, Missoula, Montana. Cincinnati, OH: United States Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health.

Marquart H, Van Drooge H, Groenewold M, Van Hemmen J. (2001) Assessing Reasonable Worst-Case Full Shift Exposure. Appl. Occup. Environ. Hyg.; 16 210-217

Money CD, Margary SA. (2002) Improved Use of Workplace Exposure Data in the Regulatory Risk Assessment of Chemicals within Europe. Ann. Occup. Hyg.; 46 279-285

Mortimer VD. (1982) Preliminary Survey Report No. 108-17a: Particleboard Plant, Timber Products Company, Medford, Oregon. Cincinnati, OH: United States

Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health.

Mulhausen JR, Diamano J. (1998) A Strategy for Assessing and Managing Occupational Exposures. Fairfax, VA: AIHA Press.

Niemelä RI, Priha E, Heikkila P. (1997) Trends of formaldehyde exposure in industries. Occup. Hyg.; 4 31-46

Raaschou-Nielsen O, Hansen J, Thomsen BL, Johansen I, Lipworth L, McLaughlin JK, et al. (2002) Exposure of Danish Workers to Trichloroethylene, 1947-1989. Appl. Occup. Environ. Hyg.; 17 693-703

Rappaport SM. (2000) Interpreting Levels of Exposures to Chemical Agents. In: Harris RL, Editor. Patty's Industrial Hygiene. New York, NY: John Wiley & Sons, Inc. p. 679-745.

Sussell A. (1995) Health Hazard Evaluation Report No. HETA-91-0239-2509, Medite of New Mexico, Las Vegas, New Mexico. Cincinnati, OH: United States Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health.

Tielemans E, Marquart H, de Cock J, Groenewold M, Van Hemmen J. (2002) A proposal for Evaluation of Exposure Data. Ann. Occup. Hyg.; 46 287-297

Triebig G, Schaller KH, Beyer B, Müller J, Valentin H. (1989) Formaldehyde Exposure at Various Workplaces. Sci. Total Environ.; 79 191-195

Wentrup GJ, Brenk FR, Wenzel M, Striefler B.(1986) Field Measurements of Formaldehyde for Workplace Monitoring, Using Various Active and Passive Methods for Personal and Area Sampling. In: Berlin A, Brown RH, Saunders KJ, editors. Diffusive Sampling - An Alternative Approach to Workplace Air Monitoring; 1986 22-26 September 1986; Luxembourg: Royal Society of Chemistry, London; p. 328-332.

Zwillinger D, Kokoska S. (2000) Standard Probability and Statistics Tables and Formulae. Boca Raton, FL: Chapman & Hall / CRC.

3.8 Appendix

Let Y be a random variable following a normal distribution with mean μ and standard deviation σ. Let y(1),…,y(n) be the order statistics of any sample of size n taken from this distribution, R the range of the sample, and $\hat{P}_x$ the empirical x^{th} percentile of the sample.

Justification of equation 6

Usually, $\hat{P}_x$ is obtained by finding the ith order statistic (usually the ith smaller value) for which $\frac{i}{n} \geq \frac{x}{100}$. Formally, each order statistic follows a distribution of which the expected value can be estimated. In particular, Blom proposed the following formula for samples drawn from a normal distribution (Blom, 1958):

$p = \left(\frac{i-0.375}{n+0.25}\right) \times 100$ (7) with n the sample size, i the rank of the value of $\hat{P}_x$ in the sample, and p such that $\hat{P}_x$ is an estimate of the pth percentile of the population. Thus, in theory, Z_x in equation 5 should be replaced by Z_p, p being calculated from equation 7. However, the authors found that there was less than 1% difference between p and x when the sample size is greater than 14. Therefore to simplify calculations, equation 6 was used in our study since all sample sizes for which this equation was to be used were greater than 30.

Justification of equation 7

The standardized range of a random sample from a normal distribution, defined by $W = \frac{R}{\sigma} = \frac{y_{(n)} - y_{(1)}}{\sigma}$, follows a specific sampling distribution, described in equation form by Hartley (Hartley, 1942). The cumulative density function of this distribution can be estimated by numerical integration. This concept is much used in R-charts in process quality control, where the ranges of sequential process samples are plotted against control limits, with excessive values indicating departure from the initial distribution and loss of control. The control limits are computed as chosen percentiles of the theoretical sampling distribution of the ranges. The equations proposed by Hartley show that the sampling distribution of the range of the standard normal depends only on the sample size; therefore the sampling distribution of the standardized range of a sample of any normal distribution depends only on the sample size. If we know the quantity y(n)-y(1) for one sample of size n from a normal distribution, and we assume that this single value is close from its theoretical median, we can then estimate σ as : $\sigma = \frac{y_{(n)} - y_{(1)}}{W_{median}}$. Applied to our situation with a lognormal distribution, we estimate GSD=exp(σ) from a range [ln(b),ln(a)] and a given sample size.

The determination of W_{median}, the theoretical median of the standardized range was performed with the function *qnrange* of the S-Plus 6.1 statistical software (Insightfull, 2001). This function solves the equations proposed by Hartley by numerical integration. As an alternative to this function, a table giving the cumulative probability of the sampling distribution for values of W ranging from 0 to 7.25 and for sample sizes between 2 and 20 can be found in Zwillinger and Kokoska p69-76 (Zwillinger and Kokoska, 2000).

CHAPITRE IV

MODÉLISATION STATISTIQUE DES NIVEAUX D'EXPOSITION PROFESSIONNELLE AU FORMALDÉHYDE DANS L'INDUSTRIE FRANÇAISE, 1986-2003

Article publié dans "Annals of occupational hygiene", Vol 50 (3), pp.305-321

Statistical modelling of formaldehyde occupational exposure levels in French industries 1986-2003

Jérôme Lavoué[(1)], Raymond Vincent [(2)], and Michel Gérin[(1)*].

(1) Groupe de recherche interdisciplinaire en santé (GRIS)
Département de santé environnemtale et santé au travail
Faculté de médecine
Université de Montréal
P.O Box 6128, Main Station
Montreal (QC) Canada H3C 3J7

(2) Institut national de recherche et de sécurité
Département de métrologie des polluants
Vandoeuvres-les-Nancy
France

Keywords: formaldehyde, determinants of exposure, mixed-effects models, occupational exposure database, COLCHIC.

***Author to whom correspondence should be addressed**

4.1 Abstract

Occupational exposure databanks (OEDB) have been cited as sources of exposure data for exposure surveillance and exposure assessment in epidemiology. In 2003, an extract was made from COLCHIC, the French national OEDB, of all concentrations of formaldehyde. The data were analysed with extended linear mixed-effects models in order to identify influent variables and elaborate a multisector picture of formaldehyde exposures. Respectively, 1401 and 1448 personal and area concentrations were available for the analysis. The fixed effects of the personal and area models explained, respectively, 57 and 53% of the total variance. Personnal concentrations were related to the sampling duration (short-term higher than TWA levels), decreased with the year of sampling (-9% per year) and were higher when local exhaust ventilation was present. Personal levels taken during planned visits and for occupational illness notification purpose were consistently lower than those taken during ventilation modification programmes or because the hygienist suspected the presence of significant risk or exposure. Area concentrations were related to the sampling duration (short-term higher than TWA levels), decreased with time (-7% per year) and when the measurement sampling flow increased. Significant within-facility (correlation coefficient 0.4 to 0.5) and within-sampling campaign correlation (correlation coefficient 0.8) was found for both area and personal datasets. The industry/task classification appeared to have the greatest influence on exposure variability while the sample duration and the sampling flow were significant in some cases. Estimates made from the models for year 2002 showed elevated formaldehyde exposure in the fields of anatomopathological and

biological analyses, operation of glueing machinery in the wood industry, operation and monitoring of mixers in the pharmaceutical industry, and garages and warehouses in urban transit authorities.

4.2 Introduction

Formaldehyde is an irritant gas recently classified as carcinogenic to humans (International Agency for Research on Cancer, in press). Exposure in a wide array of workplaces is mainly due to its presence in amino and phenolic resins used in several products such as varnishes, glues, and plastics. Formaldehyde is also found in sanitizing products, as an intermediate product in chemical synthesis, in histological fixative products, and embalming fluids. The recent change in the IARC classification of formaldehyde from group 2A (probably carcinogenic to humans) to group 1 (Carcinogenic to humans), based on *sufficient evidence* that it causes nasopharyngeal cancer in humans, constitutes an incentive for improved exposure surveillance in workplaces where formaldehyde is present.

The COLCHIC database is the French national occupational exposure data bank (Carton and Goberville, 1989). Set up in 1987, it contains the results of measurements taken since 1986 by eight French regional health insurance fund interregional chemical laboratories and the laboratories of the Institut National de Recherche et de Sécurité (INRS). As of 2001, COLCHIC contained more than 400,000 measurement results taken in 14,000 facilities, corresponding to 600 different substances (Vincent and Jeandel, 2001). In 1995, Carton presented a succinct summary of the 2700 formaldehyde measurements then available in COLCHIC (Carton, 1995).

The potential of occupational exposure databases (OEDB) as a source of exposure data has already been mentioned in the literature for purposes including mainly exposure surveillance (Goldman et al., 1992;LaMontagne et al., 2002), exposure assessment for epidemiology (Stewart and Rice, 1990), or regulatory impact assessment (Botkin and Conway, 1995). Several limits of nation-wide OEDBs have also been discussed in the literature (Gomez, 1997;Stewart and Rice, 1990;Ulfvarson, 1983). The main objective of our study was to explore the extent to which the ancillary information provided in COLCHIC allows predicting the exposure levels in this OEDB, in order to gain insight on the usefulness of this OEDB for exposure assessment and produce a multi-sector picture of formaldehyde exposures.

4.3 Methods

The COLCHIC database has been described in details previously (Carton, 1995;Carton and Goberville, 1989;Vincent and Jeandel, 2001). COLCHIC covers only the workplaces under the authority of the French national security general insurance scheme, thus excluding State services (e.g army, education), agriculture, mines, energy production, and national mass transit. For the purpose of this study, all formaldehyde concentrations recorded in COLCHIC since 1986 up to September 2003 were extracted from the database. The extract was refined by eliminating data without the following characteristics: quantitative personal or area measurement, sampling device used is a sampling tube. The analytical method used in the COLCHIC formaldehyde data was presented previously (INRS, 2003). Results reported only as 'detected' or 'superior to' were excluded (n=126). A total of 65 records for which the concentration was >10 mg/m^3 were investigated by contacting the institution having made the measurements, which caused the exclusion or correction of 7 records.

The resulting dataset was split into 'area' and 'personal' datasets. Within each dataset, values reported as under a detection limit (concentration < x) were replaced by the detection limit divided by 2 (x/2) (Hornung and Reed, 1990). Values reported as 'non detected' were replaced by the median of the reported detection limit divided by two [median(x)/2].

Each measurement in COLCHIC can be related to variables indicative of economic activity and task. The ECA variable is a codification of economic activities used to classify companies for occupational illnesses compensation purposes in France. It is based on the NAF classification with the addition of one character to allow for more precision. The NAF classification is the French four-character classification of economic activities (NAF, Rev. 1) (Ministère de l'économie des finances et de l'industrie, 2003) and is related to European NACE classification (European community (EC), 2002). The TASK variable is a five-character code (with ~700 categories), specific to COLCHIC, which identifies the task or workstation corresponding to a sample (Carton and Goberville, 1989). In order to improve stratified analyses with these variables a partial aggregation of data across categories was performed, yielding to the modified variables Mod-ECA and Mod-TASK (see Table 1). Hence, when less <40 values were available for a five-character category, one character was taken out to create a new, broader category. The process was repeated until the category contained 40 values or the code was reduced to two characters. This procedure allowed maximizing the number of data per strata while keeping a precise codification for categories with more values. A peculiarity of this classification algorithm is that when results are presented for a broadened category (e.g. code reduced to 2 characters), the results exclude data in this broad category which belong to finer categories containing > 40 data. The variable ECA-TASK was obtained by combining Mod-ECA and Mod-TASK.

4.3.1 Statistical modelling

Separate statistical modelling was performed with the personal and area datasets. The response variable selected for analysis was the natural logarithm of formaldehyde concentrations. The data were analysed with extended linear mixed-effects models, which allow modelling complex variance-covariance structures in the data (Pinheiro and Bates, 2000). The model framework used in this study can be described by the following equation:

$$Ln(C)_{ijkl} = \sum(Fixed.effects) + (Lab.effect)_i + (Facility.effect)_{ij} + (Campaign.effect)_{ijk} + (Error)_{ijkl}$$

(1)

$i = 1,\ldots,M$, $j = 1,\ldots,M_i$, $k = 1,\ldots,M_{ij}$, $l=1,\ldots,M_{ijk}$

where there are M regional laboratories, M_i facilities corresponding to the i^{th} Lab, M_{ij} sampling campaign in the j^{th} facility corresponding to the i^{th} lab, and M_{ijk} observations in the k^{th} sampling campaigns in the j^{th} facility corresponding to the i^{th} lab. The total number of observations is $\sum_{i=1}^{M}\sum_{j=1}^{M_i}\sum_{k=1}^{M_{ij}} M_{ijk}$. $Ln(C)_{ijkl}$ is the logarithm of the lth observation in the kth sampling campaigns in the jth facility corresponding to the ith lab. The model assumptions are that 1) $(Lab.effect)$, $(Facility.effect)$ and $(Campaign.effect)$ are distributed normally with mean 0 with respective standard deviations σ_{lab}, $\sigma_{facility}$, and $\sigma_{campaign}$; 2) $(Lab.effect)$, $(Facility.effect)$, $(Campaign.effect)$ and $(Error)$ are statistically independent; and 3) $(Error)$ follows a multinormal distribution with mean 0.

All fixed and random effects tested for inclusion in the models are presented in Table 1, along with descriptive statistics for the continuous variables.

Several modelling approaches were used in order to address specific complexities in our data: First, Mod-ECA and Mod-TASK were not tested simultaneously for inclusion in the statistical models. Indeed their simultaneous presence would imply that for each industrial category, the model could predict exposure for every kind of task, which is not possible since many tasks are industry-specific. Therefore, for both the personal and area datasets, three models were constructed, one in which Mod-ECA was tested along with the other potential predictors, one with the variable Mod-TASK and one with the variable ECA-TASK.

Moreover, depending on the classification used, the number of data belonging to categories with sufficient data to permit estimation was much variable. To address this issue and allow comparison of the different models, a dataset common to all industrial classification schemes was used. This dataset only included data belonging to categories of ECA-TASK with at least 30 values.

Finally, the measurement durations in COLCHIC ranged from a few minutes to several hours. Peak, short-term and TWA measurements might not be influenced by the same determinants. Therefore, the predictions of the final models were compared with those obtained with the same models fitted to a dataset restricted to sample durations > 1 h.

The sample duration was tested both as a continuous (LENGTH, Table 1) and nominal (TYPE, Table 1) variable. The categories of TYPE, initially defined by the cut-points 15, 30, 60, and 120 min. were refined during the modelling by comparison to models fitted with aggregated categories. For all nominal variables, estimates of the model coefficients for categories containing <10 values were not reported. Data which were classified in the 'unknown' or 'other' category of a nominal variable were excluded. The random effect structure tested in our model was a nested structure "campaign in facility in lab" (equation 1). The correlation between measurements taken in the same group of the 1st, 2nd, and 3rd levels of classification were estimated as follows :

i. Correlation of measurements taken in the same lab but different facilities and sampling campaigns

$$\rho_{lab} = \frac{\sigma^2_{lab}}{\sigma^2_{lab} + \sigma^2_W + \sigma^2_{facility} + \sigma^2_{campaign}} \quad (2)$$

σ^2_W is the residual variance.

ii. Correlation of measurements taken in the same facility but during different sampling campaigns

$$\rho_{facility} = \frac{\sigma^2_{lab} + \sigma^2_{facility}}{\sigma^2_{lab} + \sigma^2_{W} + \sigma^2_{facility} + \sigma^2_{campaign}} \quad (3)$$

iii. Correlation of measurements taken during the same sampling campaign

$$\rho_{campaign} = \frac{\sigma^2_{lab} + \sigma^2_{facility} + \sigma^2_{campaign}}{\sigma^2_{lab} + \sigma^2_{W} + \sigma^2_{facility} + \sigma^2_{campaign}} \quad (4)$$

The *lme* function of the S-plus software, which was used in our study, provides estimates of the different intergroup variabilities (e.g. $\sigma^2_{campaign}$). This implies variable intra-group correlation coefficients (e.g $\rho_{campaign}$) when σ_W is modelled as dependant on other variables. The intra-group correlation coefficients presented in the results section were therefore calculated for an average residual standard deviation.

The potential dependency of σ_W on other variables (i.e heteroscedastic model) was tested in several ways in our study: σ_W modelled as different for each category of a nominal variable (all nominal variables in Table 1 were tested) as illustrated in equation 5; σ_W varying linearly (equation 6) or exponentially (equations 7 and 8) with a continuous variable (the response and all continuous variables in Table 1 were tested).

$\sigma_w = \beta_i$ (5), with β_i to be estimated, i=1 to the number of categories of the nominal variable

$\sigma_w = C*(\beta X)(6)$

$\sigma_w = C * \exp(\beta X)$ (7)

$\sigma_w = C * \exp(\beta \ln(X))$ (8), with C and β to be estimated and X the continuous variable tested.

Only first order interactions were tested for inclusion in the models.

Table 1: Variables tested in the statistical models

Variable	Type	Descritpion
	Fixed effects	
SEASON	Nominal (4 categories)	Season of the sampling as defined by the following break dates: winter (12/22 to 3/20), spring (3/21 to 6/21), summer (6/22 to 9/22), autumn (9/23 to 12/21)
MOD-ECA	Nominal (respectively, 17 and 20 categories in the personal and area restricted datasets)	Constructed from the aggregation (see Methods) of the five-character industrial classification ECA.
MOD-TASK	Nominal (respectively, 15 and 19 categories in the personal and area restricted datasets)	Constructed from the aggregation (see Methods) of the five-character classification of workstations TASK.
ECA-TASK	Nominal (respectively, 27 and 31 categories in the personal and area restricted datasets)	Constructed as a combination of MOD-ECA and MOD-TASK
YEAR	Continuous (integer) 1986 to 2003	Year of sampling
LENGTH	Continuous (min.) Interquartile interval: 20-89 min. for personal data 43-111 min. for area data	Sample duration
TYPE	Dichotomous Personal dataset : <15min / >15min Area dataset : <60min / >60min	Constructed from the continuous variable LENGTH
FLOW	Continuous (L/min.) Interquartile interval : 0.5-1 L/min for personal data 0.6-0.1 L/min for area data	Sample flow rate Calculated from the variables LENGTH and VOLUME
VOLUME	Continuous (L) Interquartile interval : 15-60 L for personal data 30-77 L for area data	Volume of sampling, only FLOW and LENGTH were tested in the models
AGMOTIV	Nominal (9 categories) 1. Suspicion of exposure	Constructed from a variable indicating the

	2. Suspicion of health risk 3. Observation of health effects 4. Occupational illness notification 5. Law enforcement 6. systematic survey 7. Initial survey 8. Modification of the ventilation system	reason of sampling by aggregating categories with similar purposes (reduction from 17 to 9 categories)
AMBTEMP	Nominal (4 categories) 1. <10ºC 2. 10-25ºC 3. >25ºC 4. Unknown)	Information on the temperature on the sampling site
COLPROT	Nominal (6 categories) 1. No local exhaust ventilation 2. Local exhaust ventilation (non enclosing) 3. Local exhaust ventilation (enclosing) 4. Ventilated booth 5. Employee far from emission source 6. Unknown	Information on collective exposure control measures
VENTIL	Nominal (4 categories) 1. No mechanical general ventilation 2. No mechanical general ventilation but local exhaust ventilations systems 3. Mechanical general ventilation 4. Unknow	Information on the pattern of general ventilation
Random effects		
LAB	Nominal (9 categories)	Identification of the laboratory who conducted the survey
FACILITY	Nominal (respectively, 123 and 172 categories in the personal and area modelling datasets)	Code identifying anonymously each facility visited
CAMPAIGN	Nominal (respectively, 223 and 240 categories in the personal and area modelling datasets)	Sampling campaign: regrouped all measurements made in the same plant during a period of 4 or less consecutive days.

REML optimization was used to choose the random effects and residual variance structures, and estimate the final model parameters. ML optimization was used to compare models with different fixed effects structures. Two series of models were constructed, one using the Akaike information criterion (AIC), and the other using the

Bayesian information criterion (BIC). Model building was performed firstly by means of a manual forward stepwise routine using the discriminating criterion for the fixed-effects structure. Then the best random effect structure was added by comparing the criteria of the possible models. Finally the variance structure for the residual variance was assessed the same way. The next step consisted in retesting the fixed effects for removal or addition of variables. The random effects and variance structure were adjusted again if the fixed effect model had changed.

In order to maximize the power of the analysis, the datasets were not parted between a 'model building set' and a 'validation set'. Rather, the whole datasets were used for the construction of the models. Internal validation was conducted by graphical assessment of residuals and estimates of random effects.

In order to illustrate the quantitative influence on exposures of the fixed effects coded as nominal variables, relative indices of exposure (RIE, equation 9) were calculated (Lavoué et al., 2005). For each variable, the category corresponding to the highest number of observations (the reference category) was assigned the value 100%.

$$RIE_{levelA}(\%) = 100 * \exp\left(Coeff_{levelA} - Coeff_{levelRef}\right) \quad (9)$$

with RIE_{levelA} the relative index of exposure for level A of the variable in question, $Coeff_{levelA}$ the estimated coefficient corresponding to the category A and $Coeff_{levelRef}$ the estimated coefficient corresponding to the reference category. $Coeff_{levelRef}$ is 0 when the reference category is included in the intercept. Thus, relative to the reference category, exposure levels associated with other categories are estimated as percentages.

All analyses were conducted with the statistical software S-plus 6.1 professional edition for windows Release 1 (Insightfull corp., Seattle, WA).

4.4 Results

4.4.1 Descriptive analysis

The extract from COLCHIC contained 7392 formaldehyde measurements corresponding to the preliminary criteria mentioned in the methods, with 44 and 56% of personal and area measurements, respectively. These were taken between 1986 and 2003 in 692 facilities covering 259 five-character economic activity codes. Respectively, 52 and 43% of the personal and area data belonged to categories of ECA-TASK with >30 values, leading to modelling datasets with 1401 and 1448 values.

During the compilation of the datasets, it was found that each result in COLCHIC corresponds to one analytical result instead of an 'exposure' result. Hence, several results may actually correspond to one 'full shift exposure' evaluated with several consecutive samples (e.g. one personal full shift exposure evaluated with a 'morning' and an 'afternoon' sample). Since there is no variable in COLCHIC allowing the automated pairing of this type of result, all data were taken as separate measures of exposure. Respectively, 2.7 and 4.3% of the personal and area measurements were reported as not detected or under a limit of detection.

4.4.2 Statistical modelling

Six models were developed to accommodate the separate analyses of Mod-ECA, Mod-TASK and ECA-TASK for both personal and area measurements. Table 2 summarizes the mains features of the final models and presents the differences between models built based on AIC and BIC. The AIC models are described in details below.

Table 2: Main features of the six final mixed-effects models for the TWA measurements

	Personal measurements (n=1401)			Area measurements (n=1448)		
Model	**Economic activity**	**Task**	**Combined**	**Economic activity**	**Task**	**Combined**
$P^{(A)}$	73	68	103	73	69	105
$R^{2\,(B)}$	0.39	0.47	0.57	0.46	0.35	0.53
Fixed effects	**MOD-ECA**[D] **TYPE** **YEAR** **COLPROT** **LENGTH** AGMOTIV Interaction of TYPE with: MOD-ECA AGMOTIV **LENGTH** **COLPROT**	**MOD-TASK** **TYPE** **YEAR** COLPROT **LENGTH** AGMOTIV Interaction of TYPE with: MOD-TASK AGMOTIV **LENGTH** COLPROT	**ECA-TASK** **TYPE** **YEAR** **COLPROT** **LENGTH** AGMOTIV Interaction of TYPE with: ECA-TASK AGMOTIV **LENGTH** **COLPROT**	**MOD-ECA** **TYPE** **YEAR** **LENGTH** **FLOW** **VENTIL** Interaction of TYPE with: **MOD-ECA** FLOW **LENGTH**	**MOD-ECA** **TYPE** **YEAR** **LENGTH** **FLOW** **VENTIL** Interaction of TYPE with: **MOD-TASK** FLOW **LENGTH**	**MOD-ECA** **TYPE** **YEAR** **LENGTH** **FLOW** **VENTIL** Interaction of TYPE with: **ECA-ECA** **FLOW** **LENGTH**
Random effects	**CAMPAIGN** **In: FACILITY**	**CAMPAIGN** **In: FACILITY**	**CAMPAIGN** **In: FACILITY**	**CAMPAIGN** **In: FACILITY** **In: LAB**	**CAMPAIGN** **In: FACILITY** **In: LAB**	**CAMPAIGN** **In: FACILITY** **In: LAB**
Residual standard deviation[C]	**MOD-ECA (5)** [E] **LENGTH (7)** **TYPE (5)**	**MOD-TASK (5)** **LENGTH (7)** **TYPE (5)**	**ECA-TASK (5)** LENGTH (7) TYPE (5)	**MOD-ECA (5)** **FLOW (7)**	**MOD-TASK (5)** **FLOW (7)**	**ECA-TASK (5)** **FLOW (7)**

(A) Number of parameters; (B) Fraction of the total variance explained by the fixed effects; (C) Variables found to affect the residual variance; (D) variables included in the models built with the Bayesian information criterion (BIC) instead of the AIC criteria are in bold font; (E) the number in brackets refers to the equation used to model the influence of the variable on the residual standard deviation

For both area and personal measurements the model based on ECA-TASK explained the most variance (53 and 57%, respectively) and had the lowest AIC value (it had generally not the lowest BIC however because of the substantial number of additional parameters). In the following section only numerical results from the area and personal ECA-TASK models are presented, except when significant differences existed with models based on only Mod-ECA or Mod-TASK. Table 3 and 4 show, for personal and area measurements, respectively, the quantitative influence on exposure of all variables except the ECA-TASK classification.

In the personal model, both variable representing the sample duration (LENGTH and TYPE) were related to the response. The best fit was obtained with TYPE as a dichotomous variable defining short-term (<15 min) and TWA (>15 min) measurements, and interacting with LENGTH. This corresponds to different intercepts for short-term and TWA measurements as well as a different slope for the influence of the sample duration as a continuous variable. The TYPE variable also interacted with ECA-TASK, the reason of sampling (AGMOTIV) and the type of local exhaust ventilation (COLPROT). Estimated at 10 min. for the short-term measurements and 100 min for the TWA measurements (the respective median sample durations for these categories), the median ratio of short-term to TWA concentrations was 2.5 (varying between 1 and 4 across categories of ECA-TASK). The variable ECA-TASK explained the highest proportion of the total variability compared to the other fixed effects (partial r^2 0.32).

Table 3: Effects on exposure of the fixed effects in the personal ECA-TASK model

Variable / Category	Short-term measurements	TWA measurements
Sample duration (LENGTH)		
Rate of reduction of exposures caused by an increase	18% per 5 min. [-6;36] (A)	6% per 60 min. [-2;13]
Year of sampling		
Yearly decrease in exposures	9% per year [5;13]	
Type of LEV (COLPROT)		
No LEV (Reference)	100%	100%
Non-enclosing LEV	110% [80;151]	153% [133;176]
Enclosing LEV	143% [79;248]	120% [90;159]
Reason of sampling (AGMOTIV)		
Suspicion of exposure (Reference)	100	100
Suspicion of health risk	52% [18;147] (B)	122% [50;301]
Modification of the ventilation	103% [55;196]	85% [60;122]]
Systematic survey	n.a (C)	56% [29;110]
Notification of occupational illness	n.a	5% [0.4;64]
Observation of health effects	n.a	44% [19;101]

(A) Approximate 95% confidence interval (B) The estimates for the models based on ECA and TASK were, respectively, at 95 and 73% (C) Less than 10 values in the category

For the area model, a structure similar to the personal models was found to best represent the influence of sample duration on concentrations: a combination of LENGTH and TYPE. In this case, refining the TYPE variable also led to a dichotomous variable but the cut point that best differentiated short-term and TWA area measurements was 60 min. The TYPE variable interacted with ECA-TASK, LENGTH and FLOW. Estimated at 30 min. for the short-term measurements and 100 min. for the TWA measurements, the median ratio of short-term to TWA concentrations was 1.9 (varying between 0.1 and 20 across categories of ECA-TASK). The variable ECA-TASK explained the highest proportion of the total variability compared to the other fixed effects (partial r^2 0.38).

Table 4: Effects on exposure of the fixed effects in the area ECA-TASK model

Variable / Category	**Short-term measurements**	**TWA measurements**
Sample duration (LENGTH)		
Rate of reduction of exposures caused by an increase	6% per 5 min. [3;8] [A]	13% per 60 min. [7;20] [B]
Sample flow (FLOW)		
Rate of reduction of exposures caused by an increase	2% per 0.1L/ min. [-2;5]	6% per 0.1L/ min. [1;10]
Year of sampling		
Yearly decrease in exposures	7% per year [3;12]	
Type of ventilation (VENTILGLOB)		
No mechanical ventilation (Reference)	100%	
Mechanical ventilation	70% [50;80]	

(A) Approximate 95% confidence interval (B) The estimates for the models based on ECA and TASK were, respectively, at 42 and 32% per 60min.

Inclusion of a random effect structure resulted in an improvement of the fit for all models with an average reduction of the AIC and BIC of 13%. The variables identifying the sampling campaign and the facility were included in personal and area models. The estimated within-facility correlation coefficients were 0.5 and 0.4, respectively, for the personal and area models. The within-campaign correlation coefficients were 0.8 for both models. Inclusion of the variable lab as a random effect resulted in an improvement of the fit only for the area models, with an estimated within-lab correlation coefficient of 0.08.

Addition of a heteroscedastic structure for the error term in the models also improved the model fits with an average reduction in the AIC and BIC of 5%. For the personal model the residual variance was different for each category of ECA-TASK and TYPE (equation 5) and varied exponentially with the sample duration (equation 7). The residual standard deviation (σ_W) decreased when the sample duration increased at a rate of 24% per 5 min. for the short-term measurements and 5% per 60 min. for the TWA measurements. In addition to a different decrease rate for the two categories, the short-term measurements had their residual standard deviation higher than the TWA measurement by a factor of 2. This resulted in global GSDs, across categories of ECA-TASK, varying between 3 and 4.7 for the short-term measurements, and between 2.9 and 3.3 for the TWA measurements. For the area model σ_{w} was different for each category of ECA-TASK (equation 5) and varied exponentially with the sample flow (equation 7). σ_{w} decreased by different amounts depending on the sample duration when the sampling flow increased by 0.1L/min.: <15min. 1.9%, 15-30 min. 6.4%, 30-60 min. 9.2%, 60-120 min. 9.4, >120 min. 9.9%. This resulted in global area GSDs, across categories of ECA-TASK, varying between 3.4 and 8.4 for the short-term measurements, and between 3.4 and 6.7 for the TWA measurements.

The personal and area combined models were selected to predict personal and area concentrations for different combinations of economic sectors and tasks. They were chosen because they compare favourably to the other models and because their interpretation, by allowing taking into account both the industry and task variables, seems more adequate in the framework of occupational exposure assessment. For the personal model, annual population GMs and global and within-facility GSDs were estimated for the short-term measurements for the year 2002, a sampling duration of 10 min. (the median sample duration of short-term measurements), the influence of the type of LEV calculated by weighing the coefficients by their corresponding proportion of measurements in the modelling datasets, and the purpose of the sampling chosen as 'systematic survey'. This category was chosen because it is believed by the authors to be associated with the monitoring of more representative and 'average' exposure conditions than the other categories. The TWA personal measurements were predicted in the same conditions but for a sample duration of 100 min. (the median duration of TWA measurements). For the area model, the GMs and GSDs for the short-term measurements were estimated for the year 2002, a sampling duration of 30 min. (median duration of area short-term samples), a sampling flow rate set at the median of the values of the modelling dataset (0.98 L/min), the influence of the absence/presence of mechanical ventilation calculated by weighing the coefficients by their corresponding proportion of measurements in the modelling datasets. The TWA area measurements were predicted in the same conditions but for sample duration of 100 min. (the median duration of TWA area measurements). Population AMs and probabilities of exceedance of the French short-term exposure limit (1.2 mg/m^3) for short-term measurements and of the 8-h exposure limit (0.6 mg/m^3) for TWA measurements were also calculated. These estimates, along with the sample sizes and the raw GMs and GSDs for each category, are presented in Table 5 and 6 for short-term and TWA personal measurements and in Table 7 and 8 for area measurements, respectively.

For both personal and area measurements 24 and 24 ECA-TASK categories were available for comparison of the predictions for TWA data of the final model with the predictions of models fitted to data restricted to >1h. The personal and area '>1 h' models were fitted, respectively, to 628 and 856 values. With the '>1' prediction as the reference, the comparison for the personal measurement yielded an average bias of 17%

and a rank correlation of 0.95. These values were, respectively, -13% and 0.91 for the area measurements.

Table 5: Short-term personal exposure predictions for the year 2002 in combinations of industries and tasks

Code	Label	N(A)	GM(B)	Est. GM(C)	Est. GSD1(D)	Est. GSD2(E)	F(%)(F)
15	**Food industry**						
B6	Agriculture and food industry	10	0.15	0.15	4.0	2.9	3
202ZB	**Manufacture of wood panels (excluding plywood) and laminates**						
A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	13	0.56	0.09	3.3	2.3	0
A5440	Operation and monitoring of gluing machinery	26	0.61	0.20	3.4	2.3	2
203ZA	**Manufacture of structural wood members, rough turning and tree processing**						
A5440	Operation and monitoring of gluing machinery	23	0.57	0.21	4.4	3.3	7
452LA	**Wood carpentry work**						
A5440	Operation and monitoring of gluing machinery	19	0.81	0.42	3.6	2.5	13
751AA	**Local public administration (includes public hospitals)**						
B8010	Sterilization of examining, surgical equipment	38	0.04	0.02	3.4	2.4	0
B8020	Various biological and bacteriological analyses	17	2.49	0.54	3.5	2.4	19
B8030	Anatomopathological analyses	41	0.44	0.29	4.7	3.6	14
851AA	**Private health care**						
B8099	Other tasks performed in operating rooms	16	0.98	0.07	4.4	3.3	1
851CA	**Medical office**						
B8030	Anatomopathological analyses	19	1.38	0.86	4.0	3.0	38

(A) Sample size (B) Sample GM (mg/m^3) (C) estimated GM (mg/m^3) for year 2002, sampling duration of 10 min., sampling purpose 'systematic survey' and an effect of collective protection chosen as the average of all categories weighted by their proportion in the population (D) Estimated global GSD (E) Estimated within-facility GSD (F) Fraction of exposures estimated to be exceeding the French recommended short-term limit (1.2 mg/m^3)

Table 6: TWA personal exposure predictions for the year 2002 in combinations of industries and tasks

Code	Label	N(A)	GM(B)	Est. GM(C)	Est. GSD1(D)	Est. GSD2(E)	Mu(F)	F(%)(G)
15	**Food industry**							
B6	Agriculture and food industry	36	0.03	0.04	3.1	2.0	0.05	0
202ZA	**Wood slicing and rotary cutting, plywood, manufacture of veneer and blockboard panels**							
A5440	Operation and monitoring of gluing machinery	118	0.21	0.09	3.0	1.9	0.11	0
202ZB	**Manufacture of wood panels (excluding plywood) and laminates**							

	A3120	Operation and monitoring of moulding machinery	39	1.08	0.08	2.9	1.9	0.10	0
	A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	36	0.29	0.07	3.0	1.9	0.09	0
	A5050	Operation and monitoring of abrasion machining equipment	38	0.23	0.08	3.0	2.0	0.11	0
	A5440	Operation and monitoring of gluing machinery	45	0.24	0.06	3.0	1.9	0.08	0
	A8	Control, sterilization, cleaning, maintenance	44	0.49	0.10	3.0	1.9	0.12	0
203ZA		**Manufacture of structural wood members, rough turning and tree processing**							
	A5440	Operation and monitoring of gluing machinery	31	0.62	0.12	3.1	2.1	0.15	1
24		**Chemical industry**							
	A3020	Operation and monitoring of mixer	31	0.29	0.05	3.3	2.3	0.07	0
252AF		**Manufacture plastic plates, sheets, tubes and extruded shapes**							
	A34	Manufacture of composites	29	0.81	0.16	3.1	2.1	0.21	3
252HJ		**Manufacture of engineering plastic parts**							
	A34	Manufacture of composites	33	0.12	0.06	3.0	2.0	0.07	0
252GK		**Manufacture of diverse plastic items**							
	A3020	Operation and monitoring of mixer	28	0.10	0.10	2.9	1.9	0.12	0
275		**Foundries**							
	B2110	Operation and monitoring of equipment producing Croning-type casting moulds	38	0.20	0.06	2.9	1.9	0.08	0
275AA		**Cast iron foundries**							
	B2110	Operation and monitoring of equipment producing Croning-type casting moulds	20	0.24	0.03	3.0	2.0	0.04	0
361GA		**Manufacture wood furniture with machinery**							
	A5440	Operation and monitoring of gluing machinery	32	0.18	0.07	3.0	2.0	0.09	0
452LA		**Wood carpentry work**							
	A5440	Operation and monitoring of gluing machinery	45	0.52	0.39	3.0	2.0	0.48	26
452LB		**Manufacture and installation of framing and associated joinery**							
	A5440	Operation and monitoring of gluing machinery	50	0.34	0.08	3.0	2.0	0.10	0
751AA		**Local public administration (includes public hospitals)**							
	B8010	Sterilization of examining, surgical equipment	29	0.02	0.01	3.2	2.1	0.01	0
	B8020	Various biological and bacteriological analyses	79	0.74	0.17	3.1	2.0	0.22	4
	B8030	Anatomopathological analyses	50	0.52	0.17	3.3	2.2	0.24	6
851AA		**Private health care**							
	B8020	Various biological and bacteriological analyses	68	0.34	0.10	3.1	2.1	0.13	1
	B8030	Anatomopathological analyses	43	0.24	0.07	3.1	2.1	0.09	0
	B8099	Other tasks performed in operating rooms	24	0.10	0.03	3.1	2.0	0.03	0
851CA		**Medical office**							
	B8020	Various biological and bacteriological analyses	32	1.62	0.42	3.1	2.0	0.54	31
	B8030	Anatomopathological analyses	11	1.07	0.32	3.0	2.0	0.41	18
851KA		**Medical laboratories (outside hospitals)**							
	B8020	Various biological and bacteriological analyses	69	1.42	0.41	3.0	2.0	0.51	28
	B8030	Anatomopathological analyses	21	1.39	0.36	3.1	2.0	0.46	24

(A) Sample size (B) Sample GM (mg/m^3) (C) estimated GM (mg/m^3) for year 2002, sampling duration of 100 min., sampling purpose 'systematic survey' and an effect of collective protection chosen as the average of all categories weighted by their proportion in the population (D) Estimated global GSD (E) Estimated within-facility GSD (F) Estimated arithmetic mean of formaldehyde concentrations, calculated from the estimated gm and global gsd (G) Fraction of exposures estimated to be exceeding the French recommended 8 h limit (0.6 mg/m^3)

Table 7: Short-term area exposure predictions for the year 2002 in combinations of industries and tasks

Code	Label	N(A)	GM(B)	Est. GM(C)	Est. GSD1(D)	Est. GSD2(E)	F(%)(F)
15	**Food industry**						
B6	Agriculture and food industry	26	0.24	0.10	3.9	2.9	1
155CA	**Manufacture of cheese**						
A8030	Operation and monitoring of sterilizing machinery	36	0.11	0.10	3.5	2.6	1
17	**Textile industry**						
B4	Textile	22	0.24	0.13	4.8	3.8	5
202ZA	**Wood slicing and rotary cutting, plywood, manufacture of veneer and blockboard panels**						
A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	11	0.32	0.18	4.2	3.2	5
202ZB	**Manufacture of wood panels (excluding plywood) and laminates**						
A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	35	0.67	0.24	3.5	2.6	4
A50	Machining	13	0.61	0.30	3.8	2.9	10
244AC	**Manufacture of basic pharmaceutical products**						
A3020	Operation and monitoring of mixer	14	1.17	1.76	3.6	2.6	65
244CA	**Manufacture of proprietary medicine**						
A8610	Non specific pollution premises (office, meeting room...)	51	0.50	0.23	3.5	2.6	4
252	**Transformation of plastics**						
A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	16	0.02	0.02	3.5	2.5	0
252HJ	**Manufacture of engineering plastic parts**						
A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	23	0.09	0.02	3.9	2.9	0
275	**Foundries**						
B21	Manufacture of casting cores	14	0.11	0.12	3.8	2.9	1
B23	Metal casting	20	0.03	0.02	3.9	3.0	0
275AA	**Cast iron foundries**						
B21	Manufacture of casting cores	11	0.10	0.03	3.8	2.8	0
452	**Civil engineering**						
B33	Operations of extraction and drilling	22	0.05	0.05	3.5	2.5	0
452L	**Carpentry**						
A5440	Operation and monitoring of gluing machinery	16	0.27	0.21	5.1	4.1	11
602AA	**Urban public transportation**						
A1230	Operations of construction vehicles other than forklift, power shovel and jib crane.	43	0.76	0.36	4.3	3.3	16
751AA	**Local public administration (includes public hospitals)**						
A8610	Non specific pollution premises (office, meeting room,...)	14	0.03	0.02	4.1	3.1	0
B8010	Sterilization of examining, surgical equipment	51	0.07	0.06	3.8	2.9	0
B8020	Various biological and bacteriological analyses	46	0.24	0.08	4.4	3.4	1
B8030	Anatomopathological analyses	35	0.36	0.23	3.9	3.0	7
B8040	Anaesthesia in operating room	45	0.09	0.02	4.9	3.9	0
751AC	**External services from local public administration**						
B8020	Various biological and bacteriological analyses	13	1.73	1.79	3.9	3.0	64
851AA	**Private health care**						
B8010	Sterilization of examining, surgical equipments	25	0.10	0.06	3.4	2.5	0

	B8040	Anaesthesia in operating room	35	0.13	0.07	4.3	3.3	1
851CA		**Medical office**						
	B8020	Various biological and bacteriological analyses	15	0.66	0.33	4.2	3.2	13
851KA		**Medical laboratories (outside hospitals)**						
	B8020	Various biological and bacteriological analyses	17	1.35	0.74	5.4	4.4	37

(A) Sample size (B) Sample GM (mg/m^3) (C) estimated GM (mg/m^3) for year 2002, sampling flow of 0.98L/min, duration of 30 min, and effect of ventilation chosen as the average of 'presence' and 'absence' weighted by their respective proportion in the population (D) Estimated global GSD (E) Estimated within-facility GSD (F) Fraction of exposures estimated to be exceeding the French recommended short-term limit (1.2 mg/m^3)

Table 8: TWA area exposure predictions for the year 2002 in combinations of industries and tasks

Code		Label	N$^{(A)}$	GM$^{(B)}$	Est. GM$^{(C)}$	Est. GSD1$^{(D)}$	Est. GSD2$^{(E)}$	Mu$^{(F)}$	F(%)$^{(G)}$
15		**Food industry**							
	B6	Agriculture and food industry	18	0.07	0.05	3.7	2.8	0.08	1
	A4020	Operation and monitoring of ovens and melting pots	18	0.03	0.02	3.4	2.5	0.02	0
202ZA		**Wood slicing and rotary cutting, plywood, manufacture of veneer and blockboard panels**							
	A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	22	0.17	0.10	3.7	2.7	0.16	4
202ZB		**Manufacture of wood panels (excluding plywood) and laminates**							
	A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	17	0.39	0.19	3.4	2.5	0.28	10
	A50	Machining	15	0.57	0.19	3.6	2.7	0.31	12
	A5050	Operation and monitoring of abrasion machining equipment	65	0.31	0.13	3.5	2.6	0.21	5
	A5440	Operation and monitoring of gluing machinery	57	0.31	0.12	3.5	2.5	0.18	4
24		**Chemical industry**							
	A32	reactors	20	0.50	0.26	3.7	2.8	0.43	20
244AC		**Manufacture of basic pharmaceutical products**							
	A3020	Operation and monitoring of mixer	17	0.39	0.19	3.4	2.5	0.28	10
244CA		**Manufacture of proprietary medicine**							
	A8610	Non specific pollution premises (office, meeting room,...)	18	1.04	0.17	6.7	5.5	0.74	23
252		**Transformation of plastics**							
	A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	29	0.04	0.01	3.4	2.5	0.02	0
252HJ		**Manufacture of engineering plastic parts**							
	A3320	Operation and monitoring of press, extruder, injecting and thermoforming machinery	17	0.02	0.01	3.6	2.7	0.01	0
275		**Foundries**							
	B23	Metal casting	13	0.03	0.01	3.7	2.8	0.02	0
452L		**Carpentry**							
	A5440	Operation and monitoring of gluing machinery	21	0.25	0.06	4.0	3.0	0.11	2
602AA		**Urban public transportation**							
	A1230	Operations of construction vehicles other than forklift, power shovel and jib crane.	10	0.64	0.20	3.4	2.5	0.30	11
751AA		**Local public administration (includes public hospitals)**							
	A8610	Non specific pollution premises (office, meeting	54	0.01	0.01	3.8	2.9	0.02	0

	room,...)							
B8010	Sterilization of examining, surgical equipment	52	0.04	0.04	3.6	2.7	0.06	0
B8020	Various biological and bacteriological analyses	49	0.15	0.04	4.0	3.1	0.08	1
B8030	Anatomopathological analyses	17	0.15	0.09	3.7	2.8	0.16	3
B8040	Anaesthesia in operating room	29	0.02	0.02	4.4	3.4	0.03	0
851AA	**Private health care**							
B8010	Sterilization of examining, surgical equipment	12	0.05	0.02	3.4	2.4	0.03	0
B8040	Anaesthesia in operating room	19	0.04	0.02	4.0	3.0	0.03	0
851CA	**Medical office**							
B8020	Various biological and bacteriological analyses	15	0.88	0.38	3.9	2.9	0.67	33
851KA	**Medical laboratories (outside hospitals)**							
B8020	Various biological and bacteriological analyses	22	0.89	0.50	4.8	3.7	1.20	45

(A) Sample size (B) Sample GM (mg/m^3) (C) estimated GM (mg/m^3) for year 2002, sampling flow of 0.98L/min, duration of 30 min, and effect of ventilation chosen as the average of 'presence' and 'absence' weighted by their respective proportion in the population (D) Estimated global GSD (E) Estimated within-facility GSD (F) Estimated arithmetic mean of formaldehyde concentrations, calculated from the estimated gm and global gsd (G) Fraction of exposures estimated to be exceeding the French recommended 8-h limit (0.6 mg/m^3)

4.5 Discussion

4.5.1 Statistical modelling

The industrial classification scheme we used in this study led to the rejection of approximately 50% of the data for the statistical modelling. This was caused by the use of the combination of industry and task as the main 'activity' classification, the exclusion of categories with <30 values, and, to a lesser extent, to the algorithm used to refine the classifications. As an illustration, the use of only a two digit industrial classification would have yielded a sample size of 1925 for the personal data, compared to the 1401 in our models. However, Tables 5 to 8 show several cases (e.g. health sector) where the need for a combination of industry and task and for a fine classification is clearly demonstrated. Moreover, models built by replacing our classification by two digit variables (4 for the combination of industry and task) always showed a worse fit in term of AIC (and in several cases in terms of BIC). Finally, testing modelling datasets not excluding categories with few data resulted in hundreds of additional parameters to estimate, which caused convergence issues for the ML and REML estimation methods.

As expected, Table 2 shows that BIC models are 'included' in the AIC models. We chose the AIC models as the final models because of the absence of the interaction between sample type and industry/task in the BIC models. Hence, although the differences between short-term and TWA samples across industries might not be of sufficient amplitude compared to the number of added parameters to be retained by BIC, we thought their inclusion would provide more accurate predictions. Moreover it is plausible that short-term and TWA measurements do not relate in the same way across occupational settings

The percentages of total variability explained by the fixed effects of personal and area models (ranging from 35 to 57%, Table 2) compare favourably with similar studies in the field of occupational exposure assessment (ranging between 20 and 70%) (Burstyn and Teschke, 1999).

The measurement duration was a major determinant of exposure levels in our study. The test of several ways of classifying measurements according to their duration yielded a dichotomous structure for both personal and area measurements, respectively <15 min versus >15 min. and <60min. versus >60min, with short-term measurements associated in average with exposures two times higher than TWA measurements. The difference between personal and area samples might be related to different strategies used to choose the sampling duration. Formaldehyde concentrations also decreased continuously with the sample duration at variable rates (~10% per h for TWA measurements and ~10% per 5 min. for short-term measurements). A similar influence was observed by Raaschou et al. (a decrease of 50% corresponding to an increase of 108 min.) and Lavoue et al. (a decrease of 5% corresponding to an increase of 60 min.) (Lavoué et al., 2005;Raaschou-Nielsen et al., 2002). Hence, longer sampling times for personal measurements are more likely to include 'no exposure' periods while for area measurements it is plausible that hygienists tend to sample for shorter periods in situations in which they know the ambient levels to be elevated. The higher influence of duration on short-term measurements can be explained by the progressive inclusion of 'no peak' periods in samples lasting from a few min. to more than 10 min. Comparison of the model predictions with those of similar models fitted to data restricted to sample durations > 1 h showed good agreement. This suggests that our models adequately reflect differences between the short-term and TWA measurements in COLCHIC. However, it remains unsure how our observations for TWA measurements relate to full shift exposures since 80% of the data in COLCHIC correspond to durations < 2 h.

A significant decrease of exposure levels over time was present in all models, ranging from 5% per year to 9% per year. This pattern is consistent with the reported generic temporal trends in occupational exposures reported by Symanski et *al.* (Symanski et al., 2001;Symanski et al., 1998).

The variable identifying the type of LEV was influent in the personal models, with exposures associated with LEV higher than those associated with no LEV. These observations are explained by the fact that it is likely that LEV is implemented in contaminated workplaces compared to those without LEV. TWA measurements with enclosing LEV tended to be lower than those with non inclosing LEV but this trend was inversed for short-term measurements. The absence of this variable from the area models is plausible since it is likely that personal exposures rather than ambient levels are prevented by LEV. The variable identifying the presence of mechanical ventilation was present only in the area models with significantly higher exposure associated with the absence of any mechanical ventilation. These results are plausible and tend to show that mechanical ventilation is influent, especially since the effect is observed in spite of the 'selection bias" mentioned for LEV variable. This conclusion is however somewhat hampered by the absence of any observable effect of the general ventilation on personal concentrations.

The reason of sampling was related to formaldehyde personal levels. Visits corresponding to suspicion of health risk, suspicion of exposure and modification of the ventilation were associated with higher exposure levels than those corresponding to systematic surveys, notification of occupational illness and observation of health effects. Suspicion of exposure represented 63% of the analysed data. Higher exposures linked with modification of the ventilation are not unexpected since it is plausible that these changes would take place in contaminated workplaces. A parallel might be made between the higher levels observed for 'suspicion of exposure' in our study and several reports that IMIS concentrations measured after 'complaints' are higher than those obtained during 'planned visits'(Froines et al., 1990;Melville and Lippmann, 2001;Stewart and Rice, 1990). Hence it is likely that exposure deemed potentially significant by the hygienist (in the case of COLCHIC) or the employees (in the case of the IMIS) would be actually higher than concentrations measured during visits for which no preliminary assessment hinted at potentially high exposures. The conclusions that might be drawn from our observations are however limited by the fact that the variable was absent from the area models and would not have been included in the personal models if BIC had been used instead of AIC. Short-term measurement differed notably from TWA measurements only for the 'suspicion of health risk' category, with lower exposure relative to the 'suspicion of exposure' category.

The sampling flow rate was present in the area models, with an increase of 0.1L/min associated with a 6% decrease in TWA concentrations and almost no influence on short-term measurements. Interpretation of these results is not straightforward, especially given the fact that the effect was observed independently of the effect of the sample duration. Indeed, the potential collinearity between these variables was explored in several ways. Hence, The Pearson correlation coefficient calculated respectively for the area and personal measurement were -0.30 and -0.37 between duration and flow. Taking the variable flow out of the area models containing both flow and duration only modified marginally the effect of duration. We therefore conclude that the moderate correlation between the sampling flow and sampling duration did not cause significant confounding between these variables in our models. The sampling flow varied equally between and within laboratories and had an interquartile interval of [0.5-1L/min.]. The recommended sampling flow rates for the analytical method are between 0.2 and 1 L/min. The effects of the sampling flow were not changed when the analysis was restricted to sampling flows within these limits. Since the sampling method for formaldehyde in COLCHIC involves derivatization of formaldehyde with 2,4-dinitrophenylhydrazine (DNPH), a possible explanation might be incompleteness of the derivatization reaction due to a limited reaction rate. However, while this phenomenon has been documented for the sampling method involving derivatizaton with (2-hydroxymethyl) piperidine (NIOSH, 1994;U.S. Department of labor, 1994), no similar observations were found for the DNPH method. The most plausible explanation, according to the authors, is related to the sampling strategy. Hence, hygienists would tend to increase the sampling flow rate when they suspect low concentrations in order to insure that a sufficient quantity of substance is retained on the tube.

The area and personal models yielded very similar nested random effect structures, with a coefficient of correlation between measurements taken during the same sampling campaign of ~0.8, and of 0.4 between measurements taken in the same plant but during different sampling campaigns. For the area models, a weak correlation of 0.08 was detected between measurements taken by the same laboratory but in different sampling campaigns and plants. Lavoué et al. measured within-sampling campaign correlations of 0.17 and 0.56 for area and personal formaldehyde measurements in the wood panel industry in Quebec (Lavoué et al., 2005). Teschke et al. report a correlation of 0.31 between wood dust levels measured during the same inspection in data taken from OSHA's occupational exposure database IMIS (Teschke et al., 1999). The high intra-sampling campaign observed in our study may be due to the fact that most sampling campaigns in our study lasted one day. Moreover, it was impossible in our study to combine 'analytical results' corresponding to tubes used sequentially to evaluate exposure for a longer period. This probably also caused an increase of the observed correlation. The estimated correlation between measurements taken in the same plant show that plant specific determinants of exposure not accounted for by the variables present in our model are influent on formaldehyde exposures. The very low within-laboratory correlation observed in our study permits to conclude that no strong 'region' specific differences exist in the sampling strategies used by the teams collecting data for COLCHIC.

Significant structures of heteroscedasticity of the error term were found in both area and personal models. In all models, the variable identifying the industrial activity was strongly predictive of the residual variance. Such differences in exposure variability across industries have already been reported (Kromhout et al., 1993). Short-term personal measurements were more variable than personal TWA measurements. In addition to this absolute difference, the variability of short-term measurements decreased continuously with the sample duration, this influence being almost non existant for the TWA measurements. The influence of sample duration on the variability of exposure levels has a theoretical basis and has already been observed (Kumagai and Matsunaga, 1995;Preat, 1987). In the area models, an increase of the sampling flow caused a decrease of the residual variability, with a progressively greater influence for longer measurements. No plausible interpretation was found for this observation. The variability of exposure levels determines the number of samples necessary to assess an exposure situation with adequate precision. Hence, the estimated within-facility GSDs presented in Table 5-8 will be useful for industrial hygienists to help determine sample sizes *a priori* when devising sampling strategies in the sectors represented. Furthermore, with regard to statistical modelling, estimates of other parameters in the model depend on the variance-covariance structures in the case of unbalanced data. In our study, not taking into account the correlation structure would have caused an underestimation of the standard errors on the model coefficients by a factor of 2 to 3. Therefore it appears important to take into account and explore such structures of variability when modelling occupational exposures.

4.5.2 Validity of the statistical models

The internal validity of our models appears satisfactory considering the results of the graphical assessments of residuals and estimates of random effects. We thus conclude that the models developed in our study adequately reflect formaldehyde exposure levels found in the COLCHIC database.

The relevance of our results regarding actual exposure conditions in the general workplace is limited by the potential biases present in this OEDB, caused by the differential strategies used to select the workplaces visited and the jobs, workers, and tasks monitored. Very few authors have compared OEDBs exposure data to external exposure sources to assess the extent of their inherent biases. Olsen et al. found that the levels of exposure to solvents in the furniture industry collected with a random sampling strategy were lower than those found in the Danish OEDB ATABAS, which in turns were similar to measurements taken in a random sample of facilities but during specific exposure-generating tasks (Olsen et al., 1991). Vinzents et al. compared xylene exposure data from 5 European OEDBs in three industrial sectors (Vinzents et al., 1995). They report lower median levels in the two databases for which the data is collected for insurance purpose as opposed to compliance to exposure limits for the three other banks. Lavoue et al. observed that formaldehyde exposure data collected by governmental hygienists in the wood panel industry in Quebec were consistently higher than data collected in the same plants by a research team (Lavoué et al., 2005). Several studies about the U.S OEDB IMIS have reported insights on the potential biases contained in this bank, mostly gained from the analysis of the influence on exposure levels of variables identifying the purpose of the sampling or characteristics of the workplace (Coble et al., 2001;Froines et al., 1990;Froines et al., 1986;Gomez, 1997;Oudiz et al., 1983;Stewart and Rice, 1990;Teschke et al., 1999). The use of statistical models in our study allowed estimating the simultaneous influence of several variables on the concentrations stored in COLCHIC, permitting to compensate to some extent for the 'strategy' and 'selection' biases in this OEDB.

4.5.3 Estimation of current exposure levels from the models

Table 5 to 8 illustrate an attempt to draw a global portrait of exposure levels present in COLCHIC by taking into account influent variables (e.g. decreasing temporal trend), and by correcting for the 'strategy' biases inherent to this OEDB. The estimated GMs are consistently lower than raw GMs, by median factors from 2 to 3.5 depending on the table. This decrease is expected since exposures were estimated for year 2002 and a significant decreasing trend over years was observed. Personal predicted GMs are further decreased by our assumption that data collected during systematic surveys is less prone to upward biases than data collected when potential elevated exposure was suspected by the hygienists. Therefore the agreement, or lack thereof, between predicted and raw GMs should not be seen as a way of evaluating the predictive ability of the model (which is measured among other by the R^2) but rather as an illustration of the controlling by the models for the variables identified as determinants.

The predicted exposure levels in tables 5-8 show several industries and activity associated with elevated exposure levels relative to the French limits, mainly anatomopathological and biological analyses in the health sector, and several operation of gluing machinery in the wood carpentry and wood panel industry for personal exposures. In addition ambient exposures were also elevated for the operation and monitoring of mixers in the pharmaceutical and chemical industry, and garages in the urban public transportation. Most of these sectors have been mentioned in reviews on occupational exposure to formaldehyde (International Agency for Research on Cancer, 1995;Niemelä et al., 1997).

While we believe the estimates presented in tables 5 to 8 constitute a step towards the possibility to use exposure levels in OEDBs for exposure assessment, several limits prevent their direct use as a job-exposure matrix. Hence, as discussed previously, the models presented in our study were not validated with exposure data external to the

COLCHIC database. Moreover, the validity of the correction we used to account for a potential 'sampling strategy' bias still has to be confirmed with studies on other substances in COLCHIC. Furthermore, COLCHIC doesn't provide identification (even anonymous) of individuals, which prevents estimation of within- and between-worker variances. Finally, COLCHIC does not cover all workplaces in France and our analysis was restricted to a subset of the economic activities with formaldehyde data (those with the most data) in COLCHIC. Other sources would have to be used to complete the portrait of formaldehyde occupational exposure (Valiante et al., 1992).

Several authors have underlined the potential usefulness of OEDBS for exposure surveillance, exposure assessment for epidemiology or regulatory impact assessment (Botkin and Conway, 1995;Goldman et al., 1992;Gomez, 1993;LaMontagne et al., 2002;Stewart and Rice, 1990). Our results support the idea that analysis of exposure levels stored in OEDBs with refined statistical tools may significantly facilitate their interpretation within these frameworks.

4.6 Conclusion

Through statistical modelling of area and personal exposure measurements contained in the French occupational exposure database COLCHIC, several determinants of occupational exposure to formaldehyde were identified and a multi-industry exposure portrait was elaborated. Short-term measurements were higher than TWA measurements. Personal exposure decreased over time, were higher in workplaces with local exhaust ventilation compared to no LEV, were inversely correlated with the sampling duration, and depended on the reason of sampling, with samples taken during systematic surveys lower than those taken because exposure or health risk were suspected. Area measurements also decreased over time, were inversely correlated with the sampling flow, and were lower when general mechanical ventilation was present. The use of extended linear mixed-effects models allowed the identification of a strong correlation between measurements taken during the same sampling campaign, and a moderate correlation between measurements taken in the same plant. The elaboration of a multi-sector picture of formaldehyde exposure from COLCHIC by correcting for variables potentially linked to the inherent sampling biases present in this OEDB constitute an important step towards a potential wider use of exposure databanks for exposure assessment. Further studies using other substances in COLCHIC or comparing other sources of formaldehyde exposure data to our estimates are needed.

4.7 Acknowledgements

The authors would like to thank the Institut national de recherche et de sécurité (INRS) for providing access to the COLCHIC database. We would also like to thank Brigitte Jeandel and Marilyne L'Huillier for their help in extracting and managing the formaldehyde data in COLCHIC. Finally the authors are grateful to the reviewers of the first draft of this manuscript for their constructive critique. J.L. was supported by the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST).

4.8 References

Botkin A, Conway H. (1995) Relevance of Exposure Data to Regulatory Impact Analyses: Overcoming Availability Problems. Appl. Occup. Environ. Hyg.; 10 383-390

Burstyn I, Teschke K. (1999) Studying the Determinants of Exposure: A Review of Methods. Am. Ind. Hyg. Assoc. J.; 60 57-72

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

Carton B, Goberville V. (1989) La base de données COLCHIC. Cahiers de notes documentaires Securite et hygiene du travail; 134 29-38

Coble JB, Lees PS, Matanoski G. (2001) Time trends in exposure measurements from OSHA compliance inspections of the pulp and paper industry. Appl. Occup. Environ. Hyg.; 16 263-270

European community (EC). (2002) Comission regulation (EC) No 29/2002 of 19 december 2001. Official journal of the European community; L6 3. Available at: http://europa.eu.int/eur-lex/en/consleg/pdf/1990/en_1990R3037_do_001.pdf

Froines JR, Baron S, Wegman DH, S. OR. (1990) Characterization of the Airborne Concentrations of Lead in U.S Industry. Am. J. Ind. Med.; 18 1-17

Froines JR, Wegman DH, Dellenbaugh CA. (1986) An Approach to the Characterization of Silica Exposures in U.S. Industry. Am. J. Ind. Med.; 10 345-361

Goldman LR, Gomez L, Greenfield S, Hall L, Hulka BS, Kaye WE, et al. (1992) Use of exposure databases for status and trend analysis. Arch. Environ. Health; 47 430-438

Gomez MR. (1993) A Proposal to Develop a National Occupational Exposure Databank. Appl. Occup. Environ. Hyg.; 8 768-774

Gomez MR. (1997) Factors associated with exposure in occupational safety and health administration data. Am. Ind. Hyg. Assoc. J.; 58 186-195

Hornung R, Reed LD. (1990) Estimation of Average Concentration in the Presence of Nondetectable Values. Appl. Occup. Environ. Hyg.; 5 46-51

INRS. (2003) Fiche Métropol 001 : Analyse des aldéhydes. Institut national de recherche et de sécurité, Vandoeuvre.

Available from: http://www.inrs.fr/:

International Agency for Research on Cancer. (1995) IARC Monographs on the evaluation of carcinogenic risks to humans Vol.62: Wood dust and formaldehyde. Lyon: World Health Organization.

International Agency for Research on Cancer. (in press) IARC Monograph on the evaluation of carcinogenic risks to humans Vol.88: Formaldehyde, 2-Butoxyethanol and 1-tert-Butoxy-2-propanol. Lyon: World Health Organization.

Joint ACGIH-AIHA Task Group on Occupational Exposure Databases. (1996) Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise. Appl. Occup. Environ. Hyg.; 11 1294-1311

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

Kumagai S, Matsunaga I. (1995) Changes in the distribution of short-term exposure concentration with different averaging time. Am. Ind. Hyg. Assoc. J.; 1 24-31

LaMontagne AD, Herrick RF, Van Dyke MV, Martyny JW, Ruttenber AJ. (2002) Exposure Databases and Exposure Surveillance: Promise and Practice. Am. Ind. Hyg. Assoc. J.; 63 205-212

Lavoué J, Beaudry C, Goyer N, Perrault G, Gérin M. (2005) Investigation of past and current exposures to formaldehyde in the reconstituted wood panels industry in Quebec. Ann. Occup. Hyg.; 49 587-600

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

Ministère de l'économie des finances et de l'industrie. (2003) Décret n° 2002-1622 du 31 décembre 2002 portant approbation des nomenclatures d'activités et de produits. Journal officiel de la république française; Janvier 2003 34

Niemelä RI, Priha E, Heikkila P. (1997) Trends of formaldehyde exposure in industries. Occup. Hyg.; 4 31-46

NIOSH. (1994) Formaldehyde by GC : method 2541. Cincinnati, OH: National Institute for Occupational Safety and Health.

Olsen E, Laursen B, Vinzents PS. (1991) Bias and Random Errors in Historical Data of Exposure to Organic Solvents. Am. Ind. Hyg. Assoc. J.; 52 204-211

Oudiz J, Brown JW, Ayer HE, Samuels S. (1983) A Report on Silica Exposure Levels in United States Foundries. Am. Ind. Hyg. Assoc. J.; 44 374-376

Pinheiro JC, Bates DM. (2000) Mixed-Effects Models in S and S-plus. New York: Springer-Verlag.

Preat B. (1987) Application of Geostatistical Methods for Estimation of the Dispersion Variance of Occupational Exposures. Am. Ind. Hyg. Assoc. J.; 48 877

Raaschou-Nielsen O, Hansen J, Thomsen BL, Johansen I, Lipworth L, McLaughlin JK, et al. (2002) Exposure of Danish Workers to Trichloroethylene, 1947-1989. Appl. Occup. Environ. Hyg.; 17 693-703

Rajan B, Alesbury R, Carton B, Gérin M, Litske H, Marquart E, et al. (1997) European Proposal for Core Information for the Storage and Exchange of Workplace Exposure Measurements on Chemical Agents. Appl. Occup. Environ. Hyg.; 12 31-39

Stewart PA, Rice C. (1990) A source of Exposure Data for Occupational Epidemiology Studies. Appl. Occup. Environ. Hyg.; 5 359-363

Symanski E, Chan W, Chang C-C. (2001) Mixed-Effects Models for the Evaluation of Long-term Trends in Exposure Levels with an Example from the Nickel Industry. Ann. Occup. Hyg.; 45 71-81

Symanski E, Kupper LL, Rappaport SM. (1998) Comprehensive evaluation of long-term trends in occupational exposure: Part 1. Description of the database. Occup. Environ. Med.; 55 300-309

Teschke K, Marion SA, Vaughan TL, Morgan MS, Camp J. (1999) Exposure to Wood Dust in U.S. Industries and Occupation, 1979 to 1997. Am. J. Ind. Med.; 35 581-589

U.S. Department of labor. (1994) Acroleine and/or formaldehyde. Salt Lake City, UH: Occupational safety and health administration, Organic Methods Evaluation Branch, OSHA Analytical Laboratory,.

Ulfvarson U. (1983) Limitations to the Use of Employee Exposure Data on Air Contaminants in Epidemiologic Studies. International Archives of Occupational and Environmental Health; 52 285-300

Valiante D, Richards TB, Kinsley KB. (1992) Silicosis Surveillance in New Jersey: Targeting Worplace using Occupational Disease and Exposure Data. Am. J. Ind. Med.; 21 517-526

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

Vinzents PS, Carton B, Fjeldstad P, Rajan B, Stamm R. (1995) Comparison of Exposure Mesurements Stored In European Databases on Occupational Air Pollutants and Definitions of Core Information. Appl. Occup. Environ. Hyg.; 10 351-354

CHAPITRE V

EXPOSITION PROFESSIONNELLE AU FORMALDEHYDE DANS L'INDUSTRIE ÉTATS-UNIENNE À PARTIR DES DONNÉES D'OSHA ET COMPARAISON AVEC DES DONNÉES DE LA BANQUE FRANÇAISE COLCHIC

Article soumis à "Journal of occupational and environmental hygiene"

Formaldehyde exposure in U.S. industries from OSHA's air sample data and comparison with the COLCHIC databank from France

Jérôme Lavoué[(1)], Raymond Vincent[(2)], and Michel Gérin[(1)*]

(1) Groupe de recherche interdisciplinaire en santé (GRIS)
Département de santé environnemtale et santé au travail
Faculté de médecine
Université de Montréal
P.O Box 6128, Main Station
Montreal (QC) Canada H3C 3J7

(2) Institut national de recherche et de sécurité
Département de métrologie des polluants
Vandoeuvres-les-Nancy
France

Keywords: Occupational exposure databank, mixed-effects models, IMIS, COLCHIC

***Author to whom correspondence should be addressed**

5.1 Abstract

National occupational exposure databanks (OEDBs) have been cited as sources of exposure data for exposure surveillance and exposure assessment for occupational epidemiology. Formaldehyde exposure data recorded in the U.S Integrated Management Information System (IMIS) between 1979 and 2001 were collected in order to elaborate a multi-industry retrospective picture of formaldehyde exposures and identify exposure determinants. The personal measurement results were analysed with linear mixed-effects models while the probability of being not detected (ND) was modelled by logistic regression. Formaldehyde personal data in IMIS were also compared to data recorded in the French OEDB COLCHIC. A total of 5266 IMIS exposure measurements were analysed with linear models which explained 25% of the total variance. Short-term measurement results were higher than time weighted average (TWA) data and decreased 19% per year until 1987 (TWA data 7% per year) and 4% per year (TWA data 4% per year) after that. Exposure varied across industries with maximal estimated TWA geometric means (GM) for 2001 in the Reconstituted wood products, Structural wood members, and Wood dimension and flooring industries (GM=0.2 mg/m^3). Highest short-term GMs estimated for 2001 were in the Funeral service and crematory industry (GM=0.39 mg/m^3). Exposure levels in IMIS were weakly associated to the type of inspection and the season of sampling. Concentrations measured during the same inspection were correlated and varied differently across industries and sample type (TWA, short term). The probability of formaldehyde data in IMIS being ND was related to the location of sampling (state), the year of sampling, the number of employees in the facility and the type of inspection. Measurements made in facilities with many employees, taken around 1995, and measured during complaint inspections were more

likely to be ND. Comparison between data from IMIS and COLCHIC involved, respectively, 4625 and 2971 data measured between 1986 and 2001. Empirical cumulative distribution functions stratified by sampling time showed an approximate 2-fold difference (COLCHIC > IMIS). Adjustment of COLCHIC TWA data to 8 hours and correction for differences across time and industries by using mixed-effects models reduced the difference to a 1.3 COLCHIC to IMIS ratio.

5.2 Introduction

Formaldehyde is an irritant gas found in a wide array of workplaces. Occupational exposure to this substance is mainly due to its presence in amino and phenolic resins used in several products such as varnishes, glues and plastics. Formaldehyde's other major uses include as a component in sanitizing products, histological fixative products and embalming fluids and as an intermediate in chemical synthesis. The recent change in the International Agency for Research on Cancer (IARC) classification of formaldehyde from group 2A (probably carcinogenic to humans) to group 1 (carcinogenic to humans) constitutes an incentive for improved exposure assessment in exposure surveillance and occupational epidemiology [1].

Occupational exposure databanks (OEDBs) have been described previously as potential sources of data for exposure surveillance or occupational epidemiology [2-4]. Several issues have been raised regarding whether data in OEDBs are representative of exposures experienced by the general working population. They include potential biases caused by non random selection of industries, companies, jobs, workers, and working conditions (e.g. "worst case" scenario) [5] and the lack of comprehensive ancillary information accompanying measurements [6;7]. While several authors have explored the influence on exposure levels of variables available in OEDBs in order identify biases [4;8-10], only two studies were found that reported a comparison of data from an OEDB to external exposure data [5;11].

In the U.S., the Integrated Management Information System (IMIS) is an OEDB maintained since 1972 by the Occupational Safety and Health Administration (OSHA). It has been described in some detail by Stewart and Rice [(4)]. Briefly, the IMIS is a centralized computer database that contains monitoring results obtained during industrial visits by OSHA's compliance officers in a wide array of industries. Several authors have reported the use of IMIS data for exposure surveillance. While Valiante et al. used the IMIS to identify industries with potential silica overexposure [(12)] and Linch et al. estimated the number of employees exposed to different levels of silica in the U.S. [(13)], most authors performed descriptive analyses of IMIS data for a specific contaminant [(14-17)] or industry [(18)]. Teschke et al. used the IMIS as a supplementary source of exposure data for an epidemiological study on the effects of occupational exposure to wood dust [(19)].

The objectives of this study were to explore the extent to which formaldehyde levels in IMIS are explained by the available variables, to draw a multi-sector retrospective picture of formaldehyde exposures in IMIS, and to compare formaldehyde data in IMIS to those available in the French OEDB COLCHIC, which we recently analysed [(10)]

5.3 Methods

5.3.1 The Integrated Management Information System (IMIS)

Each record in IMIS includes information about the company in which the inspection was made, including name and address, number of employees, and union status.

Industries are identified by a 4-digit SIC code from the 1987 or 1972 Standard Industrial Classification [20]. Information on the job title monitored is provided in a free text form. The date, type and scope of the inspection, as well as notification of any violation are also recorded. Each record is associated with the sample characteristics, including type of exposure monitored (e.g. TWA, STEL) and type of measurement (e.g. Personal, Area). Quantitative exposure levels are available in IMIS since 1979.

The IMIS extract made available for this study included all formaldehyde data from 1979 to 2001. It did not include variables identifying companies, the scope of the inspection and union status. The extract was refined by excluding screening data (corresponding to detector tubes) and data with a unit different from mg/m^3 or ppm.

There is no variable identifying the analytical method in IMIS. The current OSHA method uses an XAD-2 adsorbent tube which has been impregnated with 2-(hydroxymethyl)piperidine. The samples are desorbed with toluene and then analyzed by gas chromatography using a nitrogen selective detector [21]. This method was first implemented in 1985 and updated in 1989. The former OSHA method used a bubbler filled with a 10% methanol aqueous solution and determination by differential pulse polarography or colorimetry [22].

5.3.2 Statistical modelling

Since personal measurements represented more than 90% of our dataset, statistical modeling was restricted to these measurements. Two different modelling techniques were used in the analysis of the personal data in IMIS. First, the measurements reported above a limit of detection were modelled with extended linear mixed-effects models. We could not use imputation methods to include the non detects (NDs) in the linear models because the structure of the dataset did not identify whether a particular ND corresponded to a time-weighted average (TWA), short-term exposure level (STEL), peak, or ceiling sample [8]. In order to draw some information from the NDs in IMIS, a second analysis was performed, in which the probability of personal data being ND was modelled by logistic regression.

Partial aggregation of data across categories of 4-digit SIC codes was conducted in order to improve analyses with the industry variable, as follows: when less than 40 measurement results were available for a 4-digit SIC category, the more specific digit was dropped to create a broader category. The process was repeated until the category contained 40 values or the code was reduced to a SIC major (2-digit) division. With this classification algorithm, when results are presented for a broadened category (i.e. code reduced to 2 digits, for example), the results exclude all data within this broad category that belong to more specific categories with more than 40 values. This process preserved specific industry codes in well populated categories while avoiding an elevated number of categories with only a few values.

The job variable was not tested in the statistical models due to its un-standardized text format. Because of the many different industries in which formaldehyde is found and since it would have required knowledge about each process involved to create adequate standardized lists, manual standardization of the job titles was not attempted. The variables tested in the statistical models are listed in table I.

Table I: Variables tested in the empirical statistical models

Variable	Type	Description
	Fixed effects	
INSPECTYPE	Nominal (10 categories)	Reason that caused the inspection
INSPECTYPE.1	Nominal (4 categories) 1. Non programmed-other 2. Non programmed-referral 3. Non programmed-complaint 4. Programmed	Obtained by aggregation of categories of INSPECTYPE
INSPECTYPE.2	Nominal (3 categories) 1. Non programmed-other 2. Non programmed-complaint 3. Programmed	Obtained by aggregation of categories of INSPECTYPE
INSPECTYPE.3	Nominal (2 categories) 1. Non programmed 2. Programmed	Obtained by aggregation of categories of INSPECTYPE
XPTYP	Nominal (5 category) 1.TWA 2.STEL 3.Peak 4.Ceiling 5.Not detected	Type of measurement. Category 5 was excluded from the linear model analysis. This variable was not analysed in the logistic regression.
XPTYP.1	Nominal (3 category) 1.TWA 2.Short term 3.Not detected	Obtained by aggregation of categories of XPTYP. Category 3 was excluded from the linear model analysis. This variable was not analysed in the logistic regression.
SEASON	Nominal (4 categories)	Season of the sampling as defined by the following cut-off dates: winter (12/22 to 3/20), spring (3/21 to 6/21), summer (6/22 to 9/22), autumn (9/23 to 12/21)
CLASS	Nominal (respectively 59 and 76 categories in the 'linear' and 'logistic' datasets)	Constructed from the aggregation of the 4-digit SIC category (see methods).
YEAR	Continuous (integer) 1979 to 2001	Year of sampling
NPlANT	Continuous (integer)	Number of employees in the facility. Was also tested after logtransformation
	Random effects	
STATE	Nominal (respectively 50 and 48 categories in the 'linear' and 'logistic' datasets)	Identification of the state where the inspection took place. This variable was set as a fixed effect in the logistic regression.
ACTIVITY	Nominal (respectively 1769 and 2438 categories in the 'linear' and 'logistic' datasets)	Identifies an inspection

Linear mixed effects models

Personal formaldehyde concentrations were analysed with extended linear mixed-effects models after logarithmic transformation [10;23]. The corresponding model framework can be described by the following equation:

$$Ln(C)_{ijk} = \sum (Fixed.effects) + (Random.effectA)_i + (Random.effectB)_{ij} + (Error)_{ijk} \quad (1)$$

$i = 1,\ldots,M$, $j = 1,\ldots,M_i$, $k = 1,\ldots,M_{ij}$

where there are M groups for variable A, M_i groups for variable B in the i^{th} group of variable A, and M_{ij} observations in the j^{th} group of variable B in the i^{th} group of variable A. The total number of observations is $\sum_{i=1}^{M}\sum_{j=1}^{M_i} M_{ij}$. $Ln(C)_{ijk}$ is the natural logarithm of the k^{th} observation in the j^{th} group of variable B in the i^{th} group of variable A. The model assumptions are: 1) $(Random.effectA)$ and $(Random.effectB)$ are distributed normally with mean 0; 2) $(Random.effectA)$, $(Random.effectB)$, and $(Error)$ are statistically independent; and 3) $(Error)$ follows a multinormal distribution with mean 0 and different possible variance-covariance structures.

The influence of YEAR was tested with potential changes in the slope in 1987, which corresponded to the implementation of the OSHA 1 ppm Permissible Exposure Limit (PEL) and 2 ppm STEL, and 1992, which corresponds to the implementation of the 0.75 ppm PEL. In addition, the data were visually evaluated to see if any other periods were important. For all nominal variables, estimates of the model for coefficients corresponding to categories with less than 10 data are not reported. A hierarchical random effect structure ACTIVITY in STATE was tested in the models. We also explored a potential influence of the response, or of the predictor variables on the residual variance (heteroscedastic model). All first order interactions between the fixed effects were tested.

Intra-group correlation was measured as a fraction of the total variance represented by the variance of the random effects as described in Lavoué et al. [10].

The models were fitted using the function *lme* of the statistical package S-plus 7.0. Restricted maximum likelihood (REML) optimization was used to choose the random effects and residual variance structures and estimate the final model parameters. Maximum likelihood (ML) optimization was used to compare models with different fixed effect structures. Model building was performed by means of a manual forward stepwise procedure. The Akaike information criterion (AIC) was used as a discrimination criterion to build the model used for the exposure predictions . We chose the AIC because, being more 'inclusive' than the Bayesian information criterion (BIC), it yields models with better predictive ability. A model was also built using the BIC to identify variables selected by this more 'stringent' criterion and illustrate the variability in the model building process.

In order to illustrate the quantitative influence on exposures of the categorical variables (i.e season of sampling), relative indices of exposure (RIE, equation 2) were calculated.

$$RIE_{levelA}(\%) = 100 * \exp\left(Coeff_{levelA} - Coeff_{level\,Ref}\right) \quad (2)$$

where RIE_{levelA} is the relative index of exposure for level A of the variable in question, $Coeff_{level\ A}$ is the estimated coefficient corresponding to the category A and $Coeff_{levelRef}$ is the estimated coefficient corresponding to the reference category. $Coeff_{levelRef}$ is 0 when the reference category is included in the intercept. For each variable, the category corresponding to the highest number of observations (the reference category) was assigned the value 100% exposure. The RIE of every other category of the variable was then calculated. As an illustration, a 30% RIE for a category means that, on average, exposures in this category are ~3 times lower than those in the reference category.

In order to create a picture of formaldehyde exposures from the IMIS data, the AIC model was used to predict exposure levels for each industry. The model variables other than industrial category were used to adjust the predictions to account for their influence on exposure levels.

Internal validation was primarily conducted by graphical assessment of residuals and estimates of random effects regarding the assumptions underlying the estimation. There was no external validation of the final models.

Logistic regression

The probability for a personal formaldehyde concentration being ND was modelled in a standard logistic regression analysis as described in Hosmer and Lemeshow [24]. The influence of YEAR was tested in the same way as in the linear models.

The model was built in a manual forward stepwise procedure using BIC as a discriminating criterion and fitted using the *glm* function of S-plus 7.0. Coefficients were estimated using iteratively reweighted least squares. Potential correlation between data taken during the same inspection was evaluated by fitting the chosen fixed effect structure to a population average model using General estimating equations (GEEs) with an exchangeable correlation structure with the *gee* function of S-plus 7.0. Since it was not possible with this function to fit a nested correlation structure as in the linear models, the STATE variable was tested as a fixed effect in the logistic regression.

Internal validation was primarily conducted by graphical assessment of residuals. There was no external validation of the final model. As a general measure of goodness of fit, the squared Pearson correlation coefficient of the observed outcome (ND or not ND) with the predicted probability of being ND was calculated as described in equation 5.6 in Hosmer and Lemeshow [(24)].

In order to measure the agreement of the ranking of industrial sectors by the linear model (by increasing predicted concentrations) and by the logistic regression (by decreasing probability of being ND), the variable CLASS was forced into the logistic model and the Spearman rank correlation between the coefficients corresponding to a category of CLASS for the linear model and the modified logistic model was calculated.

5.3.3 Comparison of IMIS and COLCHIC

The COLCHIC databank is the French national OEDB [25]. Set up in 1987, it contains the results of measurements taken since 1986 by eight French Regional Health Insurance Fund interregional chemical laboratories and the laboratories of the Institut National de Recherche et de Sécurité (INRS) [26]. An analysis of formaldehyde measurements in COLCHIC was presented by Lavoué et al. [10]. The analytical method used for the determine formaldehyde concentrations stored in COLCHIC was described previously [27].

The IMIS and COLCHIC extracts available for this analysis contained data from, respectively, 1979 to 2001 and 1986 to 2003. Personal measurements accounted, respectively, for 90 and 56% of IMIS and COLCHIC data. The comparison was therefore restricted to personal measurements from both databanks taken between 1986 and 2001. Empirical cumulative distribution functions (ECDF) of data from both OEDBs were plotted, with stratification by the measurement duration (i.e. short-term vs. TWA). Since the TWA data in COLCHIC correspond to sample times varying between 15min. and several hours, and Lavoué et al. observed that TWA exposure levels continuously decreased with increasing sample time, the COLCHIC TWA data were further stratified by sample time (15-60 min., 60-120 min. and >120 min.) [10].

In order to identify differences between IMIS and COLCHIC exposure levels within specific industries, data from both databanks were recoded according the International Standard Industrial Classification (ISIC Rev.3). The 4-digit ISIC codes were obtained for COLCHIC data by correspondence with the 4-digit NAF codes (The French classification). In the case of IMIS, some data with 4-digit SIC codes corresponding to several different ISIC codes were excluded from the comparison. The comparison was also limited to combinations of ISIC categories and measurement type (i.e. short-term vs. TWA) for which both databanks contained at least 10 data. The last criterion was chosen arbitrarily as a compromise between a minimum number of data in each category for a meaningful comparison and a sizeable dataset. In order to perform the comparison by accounting for determinants of exposure identified in both databanks, the IMIS and COLCHIC data were merged and analysed with linear mixed-effects model as described above. The variables common to both databanks and available for the analysis comprised the source of data (IMIS or COLCHIC), YEAR, SEASON, TYPE (short-term, TWA), ISIC code and a variable identifying the sampling visit (ACTIVITY for IMIS, and CAMPAIGN in COLCHIC [(10)]). The sampling visit was tested as a random effect and the other variables as fixed effects. Again the potential influence of the fixed effects on the residual variance was explored. BIC was chosen as the discrimination criterion. To correct for the fact that 50% of COLCHIC TWA data correspond to sample times between 40 and 100 min. compared to the IMIS full shift samples, a model was also constructed after adjusting the COLCHIC data to a sample time of 8 hours. This was performed by using the results of Lavoué et al., who estimated a 6% per 60 min. decrease rate for TWA levels [(10)].

All analyses were conducted with the statistical software S-plus 7.0 Professional Edition for Windows (Insightful corp., Seattle, WA).

5.4 Results

5.4.1 Descriptive analysis

The IMIS extract contained 7910 formaldehyde exposure measurements. Table II shows the number of personal and area data for each type of measurement. These data were taken in 479 4-digit SIC codes (median of 5 data per code, interquartile interval [2-15]), during 2750 industrial visits (median of 2 measurements per visit, interquartile interval [1-4]). The 5 most visited industries (as represented by a 4-digit SIC code) were: Plastic products, not elsewhere classified (code 3089, n=419), Gray and ductile iron foundries (code 3321, n=410), General medical and surgical hospitals (code 8062, n=393), Funeral service and crematories (code 7261, n=360), and Motor vehicle parts and accessories (code 3714. n=300). The formaldehyde measurements in IMIS were spread amongst the 50 states and Puerto Rico, with a median of 113 data per state (interquartile interval [27-200]). Figure 1 shows the number of formaldehyde measurements entered each year in IMIS.

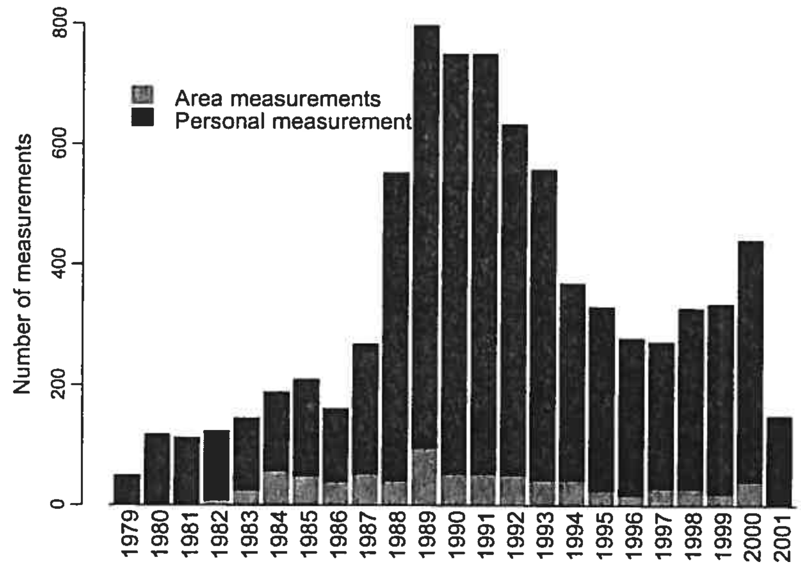

Figure 1: Number of personal and area measurement entered each year in IMIS

Table II: Number of area and personal data for each measurement type in the IMIS databank

	Non detected	Ceiling	Peak	STEL	TWA
Area	340	84	14	20	304
Personal	1825	1080	158	745	3340

5.4.2 Statistical modelling

Linear mixed-effects model

The modelling dataset, initially defined as including personal detected formaldehyde levels in IMIS, was further reduced by excluding 28 data reported as <0.002 mg/m^3 (one tenth of the limit of detection of current OSHA's analytical method for formaldehyde [21]) and 18 data >25 mg/m^3 (the concentration of formaldehyde immediately dangerous for life and health). Eleven more data were excluded because the reported SIC code did not correspond to any code in the SIC manual. This yielded a modelling dataset of 5266.

The fixed effects of the final model explained 25% of the total variance. The variables INSPECTYPE.3, CLASS, YEAR, SEASON, XPTYP.1 and interactions of XPTYP.1 with YEAR and CLASS were included in the model. The influence of YEAR was best modelled with a change in the slope only in 1987. Graphical assessment of a smoothed curve of formaldehyde levels by year did not suggest any other change in the slope. The estimated influence on exposures of YEAR along with RIEs for INSPECTYPE.3 and SEASON are presented in table III. Estimated for 1979, the median ratio of short-term (including PEAK, CEILING and STEL data) to TWA concentrations across industries was 5.7 with a minimum of 2.7 and a maximum of 27.2. Estimated for 2001 the median ratio was 1.7, with a minimum of 0.8 and a maximum of 8.4.

Table III: Effects on exposure of the year of sampling, type of inspection, and season as estimated by the linear model

Variable / Category	Short term measurements	TWA measurements
Year of sampling		
Yearly decrease in exposure (<1987)	19% [16;23][A]	7% [3;10]
Yearly decrease in exposure (>1987)	4% [2;6]	4% [2;6]
Type of inspection		
Not programmed	100%	
Programmed	94% [81;108]	
Season of sampling		
Winter (Reference)	100%	
Autumn	104% [91;137]	
Spring	92% [80;106]	
Summer	117% [100;137]	

(A) Approximate 95% confidence interval

A nested random effect structure ACTIVITY in STATE was included in the model. A correlation of 0.71 was estimated between measurements made during the same inspection. The correlation between measurements made in the same state across inspections was estimated at 0.05. The residual variance of the model varied between categories of CLASS and categories of XPTYP. With the residual standard deviation for TWA data set at 100%, it was estimated at 107% for the STEL data, 85% for the Peak data, and 95% for the Ceiling data. This variance-covariance structure resulted in GSDs varying between 3.2 and 5.8 across industries and types of sample.

The model built with the BIC criterion included the fixed effects YEAR and XPTYP.1 with an interaction between these variables. Only ACTIVITY was included as a random effect. The residual standard deviation varied with CLASS. The fixed effects of the model explained 14% of the total variance. The coefficients corresponding to variables also in the AIC model were much similar to those of the AIC model and are not presented here.

In order to construct a picture of formaldehyde exposure levels useful for exposure surveillance, GMs and GSDs for short-term and TWA measurements were predicted for the year 2001, stratified by industry. The predictions were made by using an averaged effect for season and by using the coefficient for programmed inspections. Tables IV and V present the GMs and GSDs, uncorrected for exposure determinants, and the predicted GMs and GSDs, respectively, for TWA and short-term measurements. In addition Table IV also presents the estimated arithmetic means of formaldehyde concentrations and exceedance fractions for the OSHA PEL (0.92 mg/m^3). Table V presents the exceedance fractions for the ACGIH TLV-Ceiling (0.37 mg/m^3).

Table IV: Raw and predicted geometric means of TWA formaldehyde concentrations in IMIS

Code	Label	N(A)	GM(B)	Est. GM(C)	Est. GSD(D)	Mu(E)	F(%)(F)
Division C	**Construction**	**27**	**0.12**	**0.08**	**3.8**	**0.20**	**4**
Division D	**Manufacturing**	**53**	**0.11**	**0.07**	**3.8**	**0.16**	**2**
22	Textile Mill Products	80	0.10	0.06	4.1	0.16	3
226	Dyeing And Finishing Textiles, Except Wool Fabrics	49	0.09	0.05	3.8	0.13	2
23	Apparel And Other Finished Products Made From Fabrics And Similar Materials	62	0.10	0.05	3.8	0.12	1
2326	Men's and Boys' Work Clothing	29	0.24	0.11	3.4	0.23	4
2329	Men's and Boys' Clothing, Not Elsewhere Classified	29	0.13	0.07	3.3	0.13	1
233	Women's, Misses', And Juniors' Outerwear	26	0.14	0.06	3.4	0.13	1
239	Miscellaneous Fabricated Textile Products	42	0.13	0.06	3.5	0.13	1
24	Lumber And Wood Products, Except Furniture	36	0.24	0.20	3.8	0.49	13
2431	Millwork	37	0.12	0.11	3.4	0.23	4
2434	Wood Kitchen Cabinets	77	0.07	0.10	3.6	0.23	4
2435	Hardwood Veneer and Plywood	66	0.16	0.10	4.0	0.26	5
2436	Softwood Veneer and Plywood	64	0.10	0.08	3.8	0.19	3
2493	Reconstituted Wood Products	48	0.28	0.19	4.0	0.50	13
2499	Wood Products, Not Elsewhere Classified	38	0.15	0.12	3.6	0.29	6
25	Furniture And Fixtures	62	0.14	0.09	3.4	0.20	3
2511	Wood Household Furniture, Except Upholstered	36	0.15	0.09	4.0	0.23	5
2521	Wood Office Furniture	36	0.13	0.07	4.5	0.22	4
254	Partitions, Shelving, Lockers, And Office	33	0.18	0.12	5.1	0.46	11
2621	Paper Mills	39	0.05	0.03	4.1	0.08	1
2653	Corrugated and Solid Fiber Boxes	38	0.23	0.13	4.1	0.35	8
267	Converted Paper And Paperboard Products	46	0.11	0.07	4.0	0.18	3
28	Chemicals And Allied Products	112	0.14	0.09	5.1	0.35	8
2821	Plastics Materials, Synthetic Resins, and Nonvulcanizable Elastomers	24	0.16	0.07	4.5	0.22	4
30	Rubber And Miscellaneous Plastics Products	60	0.07	0.05	4.5	0.14	2
3089	Plastics Products, Not Elsewhere Classified	168	0.13	0.08	4.1	0.22	4
32	Stone, Clay, Glass, And Concrete Products	57	0.11	0.08	4.5	0.24	5
3296	Mineral Wool	32	0.11	0.06	3.6	0.14	2
33	Primary Metal Industries	46	0.16	0.12	4.5	0.36	8
3321	Gray and Ductile Iron Foundries	209	0.16	0.08	4.0	0.22	4
3325	Steel Foundries, Not Elsewhere Classified	40	0.14	0.11	3.8	0.27	5
3365	Aluminum Foundries	40	0.17	0.12	3.8	0.29	6
3366	Copper Foundries	31	0.13	0.09	4.1	0.24	5
34	Fabricated Metal Products, Except Machinery And Transportation Equipment	118	0.09	0.07	3.5	0.16	2
35	Industrial And Commercial Machinery And Computer Equipment	66	0.18	0.09	3.8	0.21	4
356	General Industrial Machinery And Equipment	30	0.19	0.09	4.5	0.28	6
36	Electronic And Other Electrical Equipment And Components, Except Computer Equipment	80	0.07	0.04	4.1	0.11	1
3672	Printed Circuit Boards	30	0.10	0.07	3.8	0.16	2
37	Transportation Equipment	30	0.10	0.08	4.5	0.23	5
3714	Motor Vehicle Parts and Accessories	111	0.08	0.05	4.0	0.13	2
38	Measuring, Analyzing, And Controlling Instruments; Photographic, Medical And Optical	30	0.06	0.04	4.5	0.13	2

	Goods; Watches And Clocks						
39	Miscellaneous Manufacturing Industries	30	0.12	0.08	3.6	0.18	3
Division E	**Transportation, Communications, Electric, Gas, And Sanitary Services**	**54**	**0.07**	**0.05**	**3.8**	**0.12**	**1**
Division F	**Wholesale Trade**	**22**	**0.21**	**0.11**	**3.8**	**0.27**	**6**
51	Wholesale Trade-non-durable Goods	62	0.17	0.09	4.8	0.32	7
Division G	**Retail Trade**	**43**	**0.18**	**0.07**	**3.8**	**0.16**	**2**
Division I	**Services**	**103**	**0.07**	**0.05**	**3.8**	**0.12**	**1**
7261	Funeral Service and Crematories	179	0.16	0.10	4.5	0.30	7
7384	Photofinishing Laboratories	25	0.10	0.06	4.5	0.17	3
80	Health Services	60	0.10	0.07	4.2	0.20	4
806	Hospitals	31	0.22	0.11	3.8	0.28	6
8062	General Medical and Surgical Hospitals	142	0.17	0.10	5.4	0.43	10
8071	Medical Laboratories	41	0.14	0.09	4.5	0.28	6
8211	Elementary and Secondary Schools	37	0.02	0.02	3.8	0.06	0
8221	Colleges, Universities, and Professional Schools	22	0.16	0.11	3.5	0.24	4
Division J	**Public Administration**	**42**	**0.05**	**0.04**	**3.8**	**0.09**	**1**
92	Justice, Public Order, And Safety	32	0.05	0.04	4.2	0.12	2

(A) Sample size (B) Sample GM (mg/m^3) (C) estimated GM (mg/m^3) for year 2001, programmed visit, average effect of season (D) Estimated GSD (E) Estimated arithmetic mean of formaldehyde concentrations, calculated from the estimated gm and GSD (F) Fraction of exposures estimated to exceed OSHA's PEL(0.92 mg/m^3)

Table V: Raw and predicted geometric means of short term formaldehyde concentrations in IMIS

Code	Label	N(A)	GM(B)	Est. GM(C)	Est. GSD(D)	F(%)(E)
Division C	**Construction**	**16**	**0.39**	**0.17**	**3.7**	**28**
Division D	**Manufacturing**	**40**	**0.72**	**0.15**	**3.7**	**25**
22	Textile Mill Products	39	0.17	0.11	4.0	19
226	Dyeing And Finishing Textiles, Except Wool Fabrics	11	0.34	0.10	3.7	15
23	Apparel And Other Finished Products Made From Fabrics And Similar Materials	27	0.16	0.07	3.7	11
2326	Men's and Boys' Work Clothing	13	0.66	0.15	3.3	23
2329	Men's and Boys' Clothing, Not Elsewhere Classified	37	0.20	0.08	3.3	10
233	Women's, Misses', And Juniors' Outerwear	27	0.11	0.06	3.4	7
239	Miscellaneous Fabricated Textile Products	12	1.02	0.11	3.5	16
24	Lumber And Wood Products, Except Furniture	19	0.67	0.32	3.7	46
2431	Millwork	23	0.29	0.17	3.4	26
2434	Wood Kitchen Cabinets	23	0.15	0.13	3.6	20
2435	Hardwood Veneer and Plywood	55	0.30	0.21	3.9	34
2436	Softwood Veneer and Plywood	36	0.10	0.08	3.7	13
2493	Reconstituted Wood Products	34	0.58	0.35	3.9	49
2499	Wood Products, Not Elsewhere Classified	21	0.26	0.13	3.6	21
25	Furniture And Fixtures	31	0.35	0.13	3.4	20
254	Partitions, Shelving, Lockers, And Office	27	0.29	0.17	5.0	32
2621	Paper Mills	19	0.10	0.05	4.0	8
2653	Corrugated and Solid Fiber Boxes	42	0.32	0.15	4.0	26
267	Converted Paper And Paperboard Products	12	0.13	0.12	3.9	20
28	Chemicals And Allied Products	74	0.43	0.19	4.9	34
2821	Plastics Materials, Synthetic Resins, and Nonvulcanizable Elastomers	26	0.52	0.19	4.3	32

30	Rubber And Miscellaneous Plastics Products	26	0.17	0.13	4.3	23
3089	Plastics Products, Not Elsewhere Classified	109	0.33	0.14	4.0	24
32	Stone, Clay, Glass, And Concrete Products	43	0.19	0.11	4.3	20
3296	Mineral Wool	18	0.24	0.10	3.6	15
33	Primary Metal Industries	26	0.41	0.21	4.3	35
3321	Gray and Ductile Iron Foundries	100	0.29	0.13	3.9	23
3325	Steel Foundries, Not Elsewhere Classified	26	0.29	0.15	3.7	24
3365	Aluminum Foundries	19	0.31	0.16	3.7	27
3366	Copper Foundries	26	0.32	0.14	4.0	24
34	Fabricated Metal Products, Except Machinery And Transportation Equipment	62	0.17	0.09	3.5	14
35	Industrial And Commercial Machinery And Computer Equipment	44	0.39	0.12	3.7	19
356	General Industrial Machinery And Equipment	28	0.47	0.19	4.3	32
36	Electronic And Other Electrical Equipment And Components, Except Computer Equipment	57	0.12	0.07	4.0	11
3672	Printed Circuit Boards	22	0.22	0.12	3.7	20
37	Transportation Equipment	28	0.11	0.10	4.3	18
3714	Motor Vehicle Parts and Accessories	24	0.27	0.11	3.9	18
38	Measuring, Analyzing, And Controlling Instruments; Photographic, Medical And Optical Goods; Watches And Clocks	15	0.23	0.11	4.3	21
39	Miscellaneous Manufacturing Industries	13	0.14	0.12	3.6	20
Division E	**Transportation, Communications, Electric, Gas, And Sanitary Services**	**27**	**0.10**	**0.08**	**3.7**	**13**
Division F	**Wholesale Trade**	**11**	**1.05**	**0.19**	**3.7**	**31**
51	Wholesale Trade-non-durable Goods	27	0.67	0.16	4.7	30
Division G	**Retail Trade**	**14**	**0.29**	**0.12**	**3.7**	**20**
Division I	**Services**	**37**	**0.11**	**0.07**	**3.7**	**10**
7261	Funeral Service and Crematories	132	0.67	0.39	4.3	52
7384	Photofinishing Laboratories	30	0.22	0.11	4.3	20
80	Health Services	43	0.41	0.21	4.1	35
806	Hospitals	25	0.57	0.20	3.7	33
8062	General Medical and Surgical Hospitals	136	0.45	0.23	5.2	38
8071	Medical Laboratories	43	0.34	0.19	4.3	32
8211	Elementary and Secondary Schools	13	0.44	0.19	3.7	31
8221	Colleges, Universities, and Professional Schools	30	1.32	0.30	3.5	43
Division J	**Public Administration**	**14**	**0.27**	**0.13**	**3.7**	**21**
92	Justice, Public Order, And Safety	18	0.27	0.17	4.1	29

(A) Sample size (B) Sample GM (mg/m^3) (C) estimated GM (mg/m^3) for year 2001, programmed visit, average effect of season (D) Estimated GSD (E) Fraction of exposures estimated to exceed the ACGIH TLV(0.37 mg/m^3)

In order to provide the reader with the ability to predict exposure levels in other conditions, the coefficients of the model are provided in the appendix. They allow predicting the GM of formaldehyde levels for a specific set of conditions. For the prediction of exposure metrics requiring variability parameters (e.g. exceedance fraction or arithmetic mean) the estimated GSDs provided in tables IV and V can be used.

Logistic regression

In addition to the restriction of the analysis to personal data, 15 data were excluded from the logistic regression analysis because they corresponded to 2 states without ND, yielding a modelling dataset of 7133 exposure measurements.

The variables included in the logistic model comprised STATE, INSPECTYPE.2, YEAR, and the log-transformed NPLANT. Fitting the fixed effect structure as a population average model with GEEs yielded a within-ACTIVITY correlation coefficient of 0.39. The coefficients for the fixed effects were similar to those of the simple logistic model.

The effect of YEAR on the probability of being ND was best modelled with changes in the slope in 1987 and 1996 (the change in 1996 was suggested by a graphical assessment of the evolution of the fraction of NDs with time). Figure 2 presents actual and predicted proportion of NDs over time. Inclusion in the model of the variable STATE represented the greatest improvement of the fit compared to the other variables (measured as a decrease of the BIC). The actual proportions of NDs varied across state between 7 and 91%, whereas the predicted proportion varied between 12 and 91%. The measured influence of the number of employees corresponded to an increase of 14% (95% CI [3-26]) of the odds of being ND when the number of employees increased 10 times. Compared to the non-programmed complaint category of INSPECTYPE.2, measurements corresponding to non-programmed other and programmed had respective odd ratios of 0.69 (95% CI [0.56-0.84]) and 0.70 (95% CI [0.60-0.83]).

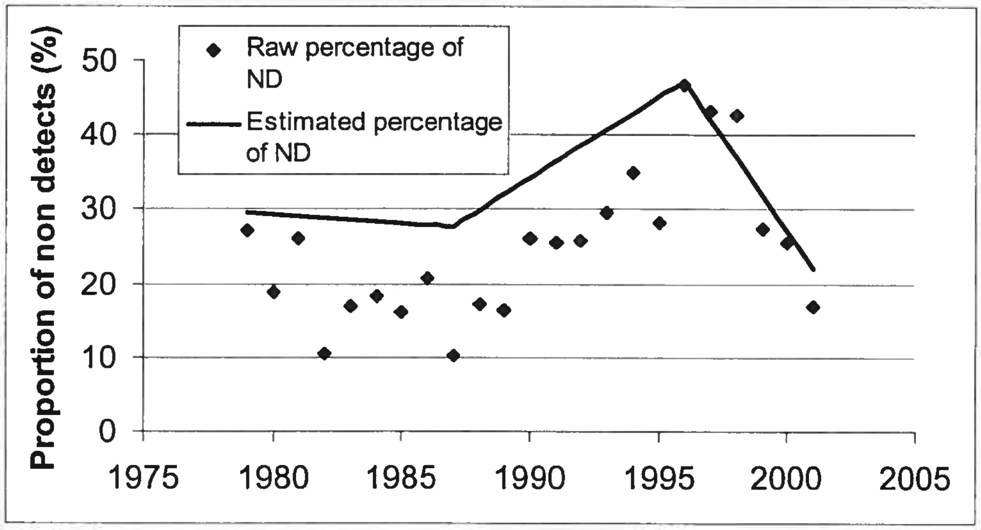

Figure 2: Time trend in the proportion of non detects.

The squared Pearson correlation coefficient of the observed outcome with the predicted probability of being ND was 0.08.

The Spearman rank correlation between the coefficients corresponding to a category of CLASS for the linear model and the logistic model was -0.39.

5.4.3 Comparison of IMIS and COLCHIC

Comparison of cumulative probability curves

The comparison dataset comprised 2971 and 4625 COLCHIC and IMIS data, respectively. ECDFs of data from both OEDBs stratified as a function of the measurement duration (i.e. short-term vs. TWA), are presented in Figure 3 and 4.

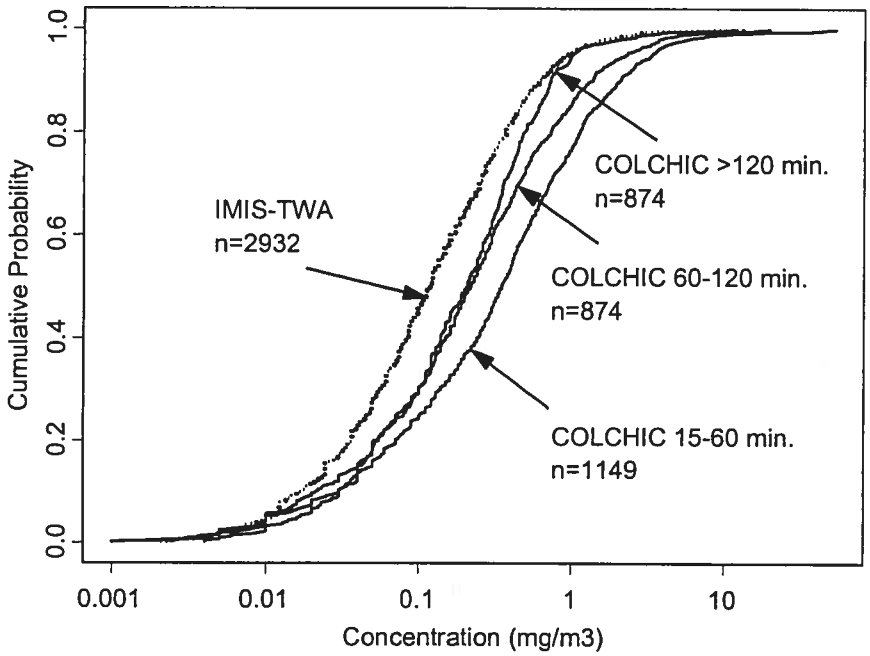

Figure 3: Empirical cumulative distribution functions of the IMIS and COLCHIC TWA data

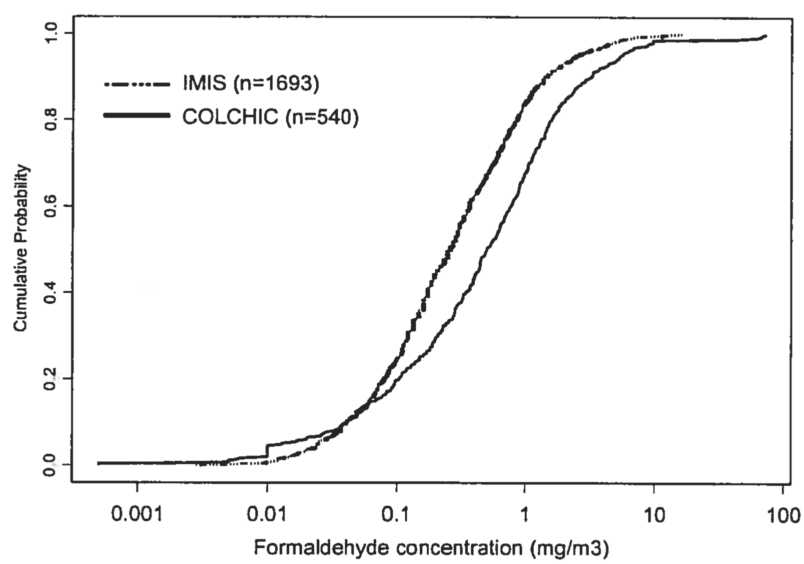

Figure 4: Empirical cumulative distribution functions of IMIS and COLCHIC short-term data

Comparison within industries

Restricting the comparison to common ISIC codes and to categories with at least 10 short-term and TWA values yielded a dataset of 2327 measurement results, of which 54% came from COLCHIC. The IMIS and COLCHIC data contained, respectively, 38 and 14% of short-term data. There were 7 ISIC categories common to both databanks: 2021, Manufacture of veneer sheets, plywood, laminated board, particle board and other panels and boards; 2520, Manufacture of plastics products; 2731, Casting of iron and steel; 3610, Manufacture of furniture; 8511, Hospital activities; 8512, Medical and dental practice activities; 8519, Other human health activities. There were 26 ISIC categories present only in the COLHIC dataset while 15 ISIC categories were present only in the IMIS dataset.

The model using the COLCHIC data unadjusted for sample time explained 8% of the total variance and contained only SOURCE and TYPE, interacting with each other, as fixed effects. With the COLCHIC TWA data being set at 100% exposure, COLCHIC short-term data were estimated at 126% (95% CI [111-143]), IMIS TWA data were estimated at 56% (95% CI [44-72]), and IMIS short-term data were estimated at 101% (95% CI [78-131]). A correlation of 0.75 (common to both datasets) was estimated between measurements taken during the same sampling visit. The residual variance varied with SOURCE and ISIC. It resulted in GSDs between 3.9 and 4.6 across industries for COLCHIC and between 4.2 and 5.6 for IMIS.

Adjusting the COLCHIC TWA data to an 8 h sampling time caused an average decrease of 27% compared to the unadjusted levels. The model using the adjusted COLCHIC data explained 10% of the total variance and contained SOURCE, TYPE, and YEAR as fixed effects. With the TWA data being set at 100% exposure, the short-term data were estimated at 185% (95% CI [166-205]). With the COLCHIC data being set at 100% exposure, the IMIS data were estimated at 71% (95% CI [57-89]). The yearly decrease in exposure levels was estimated at 4% (95% CI [1-6]). A correlation of 0.46 was estimated between measurements taken during the same sampling visit. The residual variance varied with SOURCE and ISIC. It resulted in GSDs across industries between 3.8 and 6.8 for COLCHIC and between 3.3 and 4.8 for IMIS.

5.5 Discussion

5.5.1 Statistical models

Linear mixed-effects model

The percentage of variability explained by the fixed effects in our model (25%) is rather low compared to similar analyses performed on data from other sources (28) but is comparable to other studies on IMIS data (8;9;18;19). The generally low performance of such analyses on IMIS data is probably caused by the small number of variables documented in IMIS. As an illustration, much higher values of R^2 were obtained by Lavoué et al. with data from the French OEDB COLCHIC (57%), which, among other information, contains data about tasks and characteristics of general and local ventilation. The lack of ancillary information in IMIS somewhat limits its use for occupational exposure assessment, and the addition of new variables in IMIS should be considered. Vincent and Jeandel recently described the redesign of the COLCHIC databank, involving optimization of codification schemes and addition of new variables [26].

The difference between TWA, STEL, Peak and Ceiling measurements was best modelled in our study as a simple dichotomy between TWA and short-term measurements. STEL measurements exist in IMIS only since the 1987 standard and are defined as 15 min. samples. Peak and Ceiling samples correspond to former OSHA standards. Ceiling limits should not be exceeded at any time during the shift but can be measured as 15 min. samples if instantaneous samples are not feasible. The peak value should not be exceeded for more than 30 min. during the work-shift, and it is plausible peak values would correspond to sample times of approximately 30 min. The absence of significant differences between the STEL, PEAK and Ceiling measurements is therefore not surprising. The ratios of short-term to TWA exposures in IMIS varied across industries and decreased from a median of 5.7 in 1979 to a median of 1.7 in 2001. The temporal trends estimated in our model for short-term and TWA measurement show that this evolution took place mainly between 1979 and 1987 after which both types of measurements decreased at a similar rate. These results suggest that the most extreme formaldehyde exposures were controlled before the implementation of the 1987 standard. The 7% per year decrease rate in TWA concentrations prior to 1987 was reduced to 4% per year until 2001. These time trends are similar to those reported in other studies on IMIS [9;19] or on formaldehyde exposures [10].

The type of inspection was a predictor of formaldehyde exposure levels in IMIS and was best modelled as a dichotomy between 'programmed' and 'non-programmed' inspections. Programmed inspections were associated with exposures sligthly lower (6%) than those measured during non-programmed inspection. Programmed inspections correspond to visits made in a random sample of companies within specific industrial sectors, selected because of particular safety/health concerns, and categories of company size [29]. The majority of non-programmed inspections were visits triggered by employee complaints ("complaint" inspections) or referral by a safety officer ("referral" inspection). The influence of the type of inspection on exposure levels in IMIS has been explored in several studies since the 1980s to gain insight into suspected upward biases present in this OEDB [4;8;9;15;16;19]. It has been postulated that visits triggered by complaints or referral are more likely to correspond to companies with high exposure within an industry. Altogether, these studies, performed for different substances, in different industries, and with different methodologies, provide no consistent evidence of the presence or absence of such bias. Our observations add to the weight of evidence that if complaint and referral inspections actually bias data in IMIS toward high exposure companies within industries, the bias is of limited amplitude.

The season of sampling appeared to be a marginal predictor of formaldehyde data in our model, with summer being counter-intuitively associated with slightly higher exposures than winter. In other studies this variable was found to be related to exposure [30;31] with higher exposures during cold seasons. It is believed to act as a proxy of factors influencing air-exchange rates and recirculation in facilities. Our observations can be explained by the high variability of climatic conditions across the U.S. The outside temperature would have provided a better predictor in this regard. Moreover, while the studies of Van Tongeren et al. and Lavoué et al. included only one industrial sector, our study involved many industries for which the season might not be related to exposures in the same way.

The within-inspection correlation (0.71) observed in our study is higher than that observed by Teschke et al. [(19)] for the analysis of wood dust measurements in IMIS (0.31). It is possible that the between-inspection variance in the model of Teschke et al. was reduced by additional variables (such as job), which would have reduced the observed within-inspection correlation. Part of the difference might also be attributed to differences in the variance-covariance structure modelled and in the fitting techniques. Our results are similar, however, to those observed by Lavoué et al for the correlation between formaldehyde measurements in COLCHIC taken during the same sampling campaign [(32)]. Our observations confirm the existence of correlation patterns within exposure data, which need to be taken into account when analysing these data or during the design of sampling strategies.

The existence of differences in the variability of exposure levels across industries observed in our study is plausible and has already been reported [(33)]. TWA formaldehyde data in IMIS appeared less variable than STEL data but more variable than Peak or Ceiling data, albeit by a small margin. Theoretically short-term measurements should be more variable than TWA measurements [(34;35)]. Our observations are difficult to interpret in this regard and might be due to chance or confounding by factors not accounted for in our analysis.

Predicted exposure levels

Table IV and V provide a picture of formaldehyde exposure levels in IMIS corrected for determinants. The predictions were made for 2001, an average effect of the season of sampling, and for programmed inspection. The latter two variables had a relatively small influence on exposure levels, and therefore the correction for year of sampling is likely to account for the majority of the differences between the raw and estimated GMs.

The picture drawn from Table IV shows that for the majority of sectors covered in this analysis, recent TWA exposure levels show good conformity to the OSHA PEL of 0.92 mg/m^3 (0.75 ppm). Hence, most estimated GMs are close to or less than 0.10 mg/m^3, corresponding to exceedance fractions rarely over 5-10%. Two categories have markedly higher estimated GMs. Category 2493 (reconstituted wood products) had an estimated GM of 0.19 mg/m3 and the category 24 had a GM of 0.20 mg/m^3. The latter category corresponds to lumber and wood products, except furniture, excluding industries with codes 2431, 2434, 2435, 2436, 2493, and 2499 (Table IV). Further analysis of the data in category 24 showed that most measurements came from the SIC codes 2426: Hardwood dimension and flooring mills, and 2439: Structural wood members.

Table V shows that in most industrial categories, estimated short-term GMs are between 0.10 and 0.30 mg/m^3, corresponding to exceedance fractions of the ACGIH TLV-Ceiling rarely under 20% and reaching 52%. In addition to categories 24 and 2493 mentioned above, SIC 7261 (Funeral services and crematories) and SIC 8221(Colleges, universities, and professionals schools) correspond to exceedance fractions greater than 40%. Exposure to formaldehyde in categories 24 and 2499 are related to its use in amino resins. In categories 7261 and 8221, formaldehyde is present in embalming and fixative formulations.

The coefficients presented in the appendix, by allowing retrospective prediction of formaldehyde exposure levels in IMIS, can serve as supplementary exposure assessment tools for studies using retrospective designs in occupational epidemiology.

Logistic regression

The logistic regression in our study showed that conditions associated with higher probability of a result being ND (higher number of employees, complaint inspections, specific states) seem poorly related to those associated with low exposure levels (later periods, programmed inspections). Moreover, only moderate inverse concordance between exposure levels and fraction of NDs stratified by industry was observed. The time trends in figure 2 also show poor concordance, except for the period 1987-1995, between exposure levels and the fraction of ND. While part of our observations might have been confounded by changes in the limits of detection over time (OSHA changed its method from bubblers to sorbent tubes in 1985 and updated the new method in 1989), they globally argue against the hypothesis that ND data in IMIS are measured and recorded using similar practices as the detected data. Mendeloff reported the existence of significant under-recording of the monitoring data in IMIS and his observations suggest that low and undetected concentrations are less likely to be recorded in IMIS than higher concentrations [36]. The author also reported that the phenomenon is variable across OSHA offices, which is compatible with our observations of variable fractions of NDs across states. On the other hand, Jones et al., based on the analysis of data from two OSHA area offices in one region, report that the under-reporting appeared random with regard to exposure levels [29].

The complaint inspections in our dataset were 30% more likely to be ND compared to referral or programmed inspections. It is plausible that during complaint inspections, the compliance officer would be inclined to monitor exposures they know are very low or even not present just to respond to the employees' concern. Melville and Lippmann emphasize that the ND codification in IMIS does not allow discriminating 'not detected' (ND in an environment likely to contain formaldehyde) and 'not found' (ND in an environment unlikely to contain formaldehyde) results [8]. Whereas excluding 'not detected' results from the analysis would bias the exposure estimates upward, including 'not found' results on the other hand, would introduce a downward bias.

Within-inspection correlation was also identified in the logistic model, confirming the observations made for the linear models.

5.5.2 Limitations of the industry-exposure matrix

Several limitations hamper the direct use of the prediction model we developed to assess occupational exposure to formaldehyde in the U.S. First the value of the model itself is limited by the absence in our analysis of potential influential variables that are in the databank. Identification of individual companies, union status and scope of the inspection were not available in our extract. Jobs, being only available in free text and often not reported in a consistent manner, were not tested in the model. Second, several important exposure-related elements (e.g presence / type of ventilation) are not recorded in IMIS. Third, while the exclusion of NDs is likely to have caused an upward bias in our estimates, this issue is complicated by the impossibility to discriminate 'not detected' from 'not found' data in IMIS. We believe a significant proportion of NDs in IMIS correspond to "not found" since the percentage of ND (~30%) is very high when considering the exposure levels (~between 0.1 and 0.5 mg/m3) compared to the limit of detection (0.02 mg/m^3). As an illustration, a lognormal distribution with GSD= 2.7 and 30% of values <0.020 mg/m^3 would have a GM of 0.034 mg/m^3. Fourth, it has been long suspected that data in the IMIS does not adequately reflect exposures in the general U.S. working population [36]. The various industries found in IMIS mainly result from visits caused by complaints, referral by safety officers or programs targeted at specific industrial sectors. As such they might not provide an adequate picture of the presence of formaldehyde in U.S. workplaces. In this regard, Jones et al. advised against the use of IMIS to detect new 'problem areas' [29]. Within industries, a bias toward 'dirty' companies can be related to the reason of the inspection. Our observations nevertheless suggest that such bias should be of small amplitude. Within a company exposure monitoring is probably oriented toward jobs or task causing the highest exposures. This can be accounted for if jobs are part of the analysis or if knowledge of the process permits to identify such jobs. It is also possible that for a given job an inspector would try to monitor worst case conditions. However, such conditions are not always easy to identify, and Jones et al. further suggest that, based on interviews, such "worst case" strategy was not used systematically[29]. As previously mentioned by Stewart et al. these limitations are not different from those found in other sources of exposure data and IMIS has the advantage of prospectively increasing its dataset [4]. They merely require careful interpretation from the exposure assessor.

5.5.3 Comparison with COLCHIC data

In our study, graphical comparison of ECDFs showed that COLCHIC formaldehyde data are approximately twice as high as IMIS data, for both short-term and TWA data (with COLCHIC TWA data restricted to sample times >120min, see figure 3). Refining the comparison by considering common industries and using mixed-effects models yielded a similar ratio for TWA data but a smaller one (~1.3) for short-term data. Adjustment of the COLCHIC TWA data (of which 80% are <2 hrs) to 8-h sample times prior to modelling yielded a ratio of 1.3, common to both short-term and TWA data. The adjustment of COLCHIC data probably caused artefacts in the estimation of the variance-covariance structure of the statistical model and prevented any rigorous inference-based conclusion to be drawn. Nevertheless, we believe that this analysis provided the most valid comparison. We therefore conclude that formaldehyde data in COLCHIC are higher than in IMIS by a factor of approximately 1.3 after taking into account differences in sampling times.

The observation of higher levels in COLCHIC than in IMIS is somewhat unexpected, because, with regards to the results of Vinzents et al. [11], COLCHIC is run for insurance purpose whereas IMIS is run for compliance purpose. Moreover, NDs were not included in the IMIS dataset. The observed difference might be due to several reasons. Firstly differences in processes or controls associated with formaldehyde in France and in the U.S. can't be excluded. Secondly, the exposure limits for formaldehyde in France (0. 6 mg/m^3 as 8-hour TWA and 1.2 mg/m^3 as 15-min. STEL [37]) are only recommendations. As such they might represent a less potent incentive for exposure control than legal OELs such as the OSHA PELs. Thirdly sampling strategies used to collect data might differ between the two OEDBs. Hence, of the data in our COLCHIC dataset, 70% were measured by an occupational health physician or hygienist during visits triggered by suspicion of exposure. Lavoué et al. observed that levels corresponding to such visits tended to be higher than those corresponding to visits made in the framework of a systematic survey [10], whereas we found little differences in IMIS detected levels taken during programmed or non programmed (e.g. complaint, referral) inspections.

Overall, the quite moderate difference observed between the two databanks argues in favour of the possibility of using OEDBs from different countries to supplement one another as tools for occupational exposure assessment. In particular, the increasing number of occupational epidemiological studies involving multiple countries provides an incentive for the investigation of differences amongst national exposure databanks.

5.6 Conclusion

Through statistical modelling of formaldehyde personal exposure levels in the U.S. IMIS a multi-industry retrospective picture of short-term and TWA exposures was drawn. Recent TWA concentrations in IMIS were highest in the Reconstituted wood products, Structural wood members, and Wood dimension and flooring industries while short-term concentrations were also elevated in the funeral homes and crematories, and education sectors. Formaldehyde concentrations were marginally higher when measured during visits caused by employee complaints or referral by a safety officer compared to programmed visits. Logistic regression analyses showed that the probability of formaldehyde data being labelled "not detected" was not influenced by the same variables as the detected exposure levels, suggesting that low and high formaldehyde exposures were not measured and/or recorded in IMIS using the same strategies. After correction for industry, time trend and sampling time, formaldehyde concentrations in IMIS were slightly lower than those available in the French OEDB COLCHIC. Despite the lack of ancillary information accompanying measurements in IMIS and the presence of several sources of bias, our study suggests, as have several others, that this databank still represents an invaluable source of insight into occupational exposures in the U.S. We recommend that more information on exposure determinants be recorded in IMIS and that more studies be conducted that compare national occupational exposure databanks.

5.7 Acknowledgements

The authors thank Bruce Beveridge from OSHA and Brigitte Jeandel from the INRS for providing access, respectively, to IMIS and COLCHIC formaldehyde data. We would also like to thank Jan Erik Deadman for his helpful editorial comments. J.L. was supported by the Quebec association for occupational hygiene, health and safety (AQHSST) and the Institut de Recherche Robert-Sauvé en Santé et en Sécurité du Travail (IRSST).

5.8 References

1. International Agency for Research on Cancer: *IARC Monograph on the evaluation of carcinogenic risks to humans Vol.88: Formaldehyde, 2-Butoxyethanol and 1-tert-Butoxy-2-propanol*. Lyon: World Health Organization,in press.

2. **Goldman, L.R., L. Gomez, S. Greenfield, *et al.*:** Use of exposure databases for status and trend analysis. *Arch. Environ. Health* 47(6): 430-438 (1992).

3. **LaMontagne, A.D., R.F. Herrick, M.V. Van Dyke, J.W. Martyny and A.J. Ruttenber:** Exposure Databases and Exposure Surveillance: Promise and Practice. *Am. Ind. Hyg. Assoc. J.* 63(2): 205-212 (2002).

4. **Stewart, P.A. and C. Rice:** A source of Exposure Data for Occupational Epidemiology Studies. *Appl. Occup. Environ. Hyg.* 5(6): 359-363 (1990).

5. **Olsen, E., B. Laursen and P.S. Vinzents:** Bias and Random Errors in Historical Data of Exposure to Organic Solvents. *Am. Ind. Hyg. Assoc. J.* 52(5): 204-211 (1991).

6. **Rajan, B., R. Alesbury, B. Carton, *et al.*:** 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 (1997).

7. **Joint ACGIH-AIHA Task Group on Occupational Exposure Databases:** Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise. *Appl. Occup. Environ. Hyg.* 11(11): 1294-1311 (1996).

8. **Melville, R. and M. Lippmann:** Influence of data elements in OSHA air sampling database on occupational exposure levels. *Appl. Occup. Environ. Hyg.* 16(9): 884-899 (2001).

9. **Gomez, M.R.:** Factors associated with exposure in occupational safety and health administration data. *Am. Ind. Hyg. Assoc. J.* 58(3): 186-195 (1997).

10. **Lavoué, J., R. Vincent and M. Gérin:** Statistical modelling of formaldehyde occupational exposure levels in French industries 1986-2003. *Annals of Occupational Hygiene Advance Access (March 1, 2006)*: (2006).

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

12. **Valiante, D., T.B. Richards and K.B. Kinsley:** Silicosis Surveillance in New Jersey: Targeting Worplace using Occupational Disease and Exposure Data. *Am. J. Ind. Med.* 21: 517-526 (1992).

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

14. **Oudiz, J., J.W. Brown, H.E. Ayer and S. Samuels:** A Report on Silica Exposure Levels in United States Foundries. *Am. Ind. Hyg. Assoc. J.* 44(5): 374-376 (1983).

15. **Froines, J.R., S. Baron, D.H. Wegman and O.R. S.:** Characterization of the Airborne Concentrations of Lead in U.S Industry. *Am. J. Ind. Med.* 18: 1-17 (1990).

16. **Froines, J.R., D.H. Wegman and C.A. Dellenbaugh:** An Approach to the Characterization of Silica Exposures in U.S. Industry. *Am. J. Ind. Med.* 10: 345-361 (1986).

17. **Freeman, C.S. and E.A. Grossman:** Silica exposures in workplaces in the united states between 1980 and 1992. *Scand. J. Work Environ. Health* 21 suppl 2: 47-49 (1995).

18. **Coble, J.B., P.S. Lees and G. Matanoski:** Time trends in exposure measurements from OSHA compliance inspections of the pulp and paper industry. *Appl. Occup. Environ. Hyg.* 16(2): 263-270 (2001).

19. **Teschke, K., S.A. Marion, T.L. Vaughan, M.S. Morgan and J. Camp:** Exposure to Wood Dust in U.S. Industries and Occupation, 1979 to 1997. *Am. J. Ind. Med.* 35: 581-589 (1999).

20. Executive Office of the President-Office of Management and Budget: *Standard Industrial Classification Manual*. Springfield, VA: National Technical Information Service,1987.

21. Hendricks, W.: *OSHA Sampling and Analytical Method Number 52: Acroleine and/or formaldehyde*. Salt Lake City, UT: United States Department of Labor, Occupational Safety & Health Administration, OSHA analytical laboratories, Organic methods evaluation branch,1989.

22. OSHA: *OSHA Industrial hygiene technical manual*. Salt Lake City, UT: United States Department of Labor, Occupational Safety and Health Administration,1980.

23. Pinheiro, J.C. and D.M. Bates: *Mixed-Effects Models in S and S-plus*. New York: Springer-Verlag,2000.

24. Hosmer, D. and S. Lemeshow: *Applied logistic regression, 2nd edition*. Hoboken, NJ: John Wiley & Sons, Inc.,2000.

25. **Carton, B. and V. Goberville:** La base de données COLCHIC. *Cahiers de notes documentaires Securite et hygiene du travail* 134: 29-38 (1989).

26. **Vincent, R. and B. Jeandel:** COLCHIC-Occupational Exposure to Chemical Agents database: Current Content and Development Perspectives. *Appl. Occup. Environ. Hyg.* 16(2): 115-121 (2001).

27. INRS: *Fiche Métropol 001 : Analyse des aldéhydes*. Institut national de recherche et de sécurité, Vandoeuvre,2003.

28. **Burstyn, I. and K. Teschke:** Studying the Determinants of Exposure: A Review of Methods. *Am. Ind. Hyg. Assoc. J.* 60: 57-72 (1999).

29. Jones, C.A., L. Weld, W. Gray, *et al.*: *The Sampling and Reporting Processes in OSHA MIS Data*. Cincinnati, OH: United States National Institute for Occupational Safety and Health, Grant No. R03-OH-002135 (NTIS No. PB2003-104588),1986.

30. **Lavoué, J., C. Beaudry, N. Goyer, G. Perrault and M. Gérin:** Investigation of past and current exposures to formaldehyde in the reconstituted wood panels industry in Quebec. *Ann. Occup. Hyg.* 49(7): 587-600 (2005).

31. **Van Tongeren, M. and K.G. Gardiner:** Determinants of inhalable Dust Exposure in the European Carbon Black Manufacturing Industry. *Appl. Occup. Environ. Hyg.* 16(2): 237-245 (2001).

32. **Lavoué, J., R. Vincent and M. Gérin:** Statistical modelling of formaldehyde occupational exposure levels in French industries 1986-2003. *Annals of Occupational Hygiene 50(3) 305-321.*

33. **Kromhout, H., E. Symansky and S.M. Rappaport:** A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. *Ann. Occup. Hyg.* 37(3): 253-270 (1993).

34. **Preat, B.:** Application of Geostatistical Methods for Estimation of the Dispersion Variance of Occupational Exposures. *Am. Ind. Hyg. Assoc. J.* 48(10): 877 (1987).

35. **Kumagai, S. and I. Matsunaga:** Changes in the distribution of short-term exposure concentration with different averaging time. *Am. Ind. Hyg. Assoc. J.* 1: 24-31 (1995).

36. Mendeloff, J.: *A New Strategy for Estimating Occupational Exposures to Toxic Substances.* Cincinnati, OH: National Institute for Occupational Safety and Health (microfiche number NIOSH-00182240),1984.

37. INRS: *Valeurs limites d'exposition professionnelle aux agents chimiques en France.* Vandoeuvres-les-Nancy: Institut national de recherche et de sécurité (Note documentaire 2098),2005.

5.9 Appendix

Coefficients of the linear mixed effect model for the prediction of IMIS formaldehyde exposure levels

Intercept	-1.15			**Winter**	0	**Programmed**	-0.07
(Year-1979)-TWA measurements		**(Year-1979)-Short-term measurements**		**Spring**	-0.08	**Non programmed**	0
<1987	-0.07	**<1987**	-0.21	**Summer**	0.16		
>1987	-0.04	**>1987**	-0.04	**Autumn**	0.04		
SIC codes - TWA measurements				**SIC codes – Short-term measurements**			
7261	0	**3325**	0.12	**7261**	2.57	**3325**	1.59
22	-0.49	**3365**	0.22	**22**	1.26	**3365**	1.68
226	-0.57	**3366**	-0.09	**226**	1.16	**3366**	1.50
23	-0.66	**34**	-0.30	**23**	0.88	**34**	1.13
2326	0.14	**35**	-0.11	**2326**	1.63	**35**	1.36
2329	-0.38	**356**	-0.06	**2329**	0.99	**356**	1.82
233	-0.46	**36**	-0.85	**233**	0.69	**36**	0.82
239	-0.52	**3672**	-0.37	**239**	1.25	**3672**	1.41
24	0.72	**37**	-0.26	**24**	2.37	**37**	1.15
2431	0.11	**3714**	-0.65	**2431**	1.71	**3714**	1.27
2434	0.03	**38**	-0.87	**2434**	1.43	**38**	1.34
2435	0.04	**39**	-0.23	**2435**	1.94	**39**	1.41
2436	-0.21	**51**	-0.04	**2436**	1.02	**51**	1.70
2493	0.69	**7384**	-0.56	**2493**	2.45	**7384**	1.27
2499	0.25	**80**	-0.31	**2499**	1.48	**80**	1.95
25	-0.03	**806**	0.16	**25**	1.48	**806**	1.91
2511	-0.08	**8062**	0.06	**2511**	(A)	**8062**	2.02
2521	-0.31	**8071**	-0.05	**2521**	(A)	**8071**	1.82
254	0.20	**8211**	-1.44	**254**	1.74	**8211**	1.86
2621	-1.23	**8221**	0.10	**2621**	0.57	**8221**	2.29
2653	0.29	**92**	-0.77	**2653**	1.59	**92**	1.74
267	-0.35	**D.C**	-0.14	**267**	1.35	**D.C**	1.75
28	-0.04	**D.D**	-0.40	**28**	1.84	**D.D**	1.62
2821	-0.30	**D.E**	-0.68	**2821**	1.84	**D.E**	1.02
30	-0.75	**D.F**	0.14	**30**	1.43	**D.F**	1.85
3089	-0.19	**D.G**	-0.37	**3089**	1.52	**D.G**	1.40
32	-0.24	**D.H**	(A)	**32**	1.27	**D.H**	(A)
3296	-0.45	**D.I**	-0.68	**3296**	1.15	**D.I**	0.81
33	0.19	**D.J**	-0.95	**33**	1.93	**D.J**	1.47
3321	-0.15			**3321**	1.50		

(A): Less than 10 measurements available

<u>Formulae to make predictions:</u>

For measurements before 1987:

$$\ln(GM) = \mathrm{int}\,ercept + Coeff_{SIC.code} + Coeff_{season} + Coeff_{Inspection} + (year - 1979) * Coeff_{<1987}$$

For measurements after 1987:

$$\ln(GM) = \mathrm{int}\,ercept + Coeff_{SIC.code} + Coeff_{season} + Coeff_{Inspection} + 8 * Coeff_{<1987}$$
$$+ (year - 1987) * coeff_{>1987}$$

Note: Due to rounding errors, a difference up to 15% might appear when predicting GMs compared to values presented in Tables IV and V.

CHAPITRE VI

DISCUSSION GÉNÉRALE

6. DISCUSSION GÉNÉRALE

L'analyse individuelle et comparative des cinq sources de données décrites dans les précédents chapitres a généré de nombreux résultats en regard des problématiques abordées dans ce travail. Le tableau N°5 présente un résumé succinct de ces résultats classés par chapitre. Ils sont discutés dans les sections suivantes, d'abord sur la question de l'utilisation des données préexistantes pour l'évaluation de l'exposition puis sur le sujet de l'exposition professionnelle au formaldéhyde.

6.1 Utilisation des données préexistantes pour l'établissement de portraits de l'exposition professionnelle

6.1.1 Information entourant les données d'exposition

Les analyses conduites durant ce travail confirment les observations retrouvées dans la littérature selon lesquelles les informations entourant les données d'exposition, nécessaires pour en évaluer la pertinence, sont généralement documentées de façon insuffisante (Caldwell et coll., 2001;Marquart et coll., 2001). Ainsi, en ce qui concerne la revue de littérature présentée au chapitre III, une proportion importante des données n'étaient associées qu'à un procédé et une période historique imprécise (fourchette de 10 ans). Selon les critères proposés par Tielemans et coll., la plupart de ces articles auraient dû être exclus de l'analyse (Tielemans et coll., 2002).

Tableau N°5 : Synthèse des principaux résultats obtenus lors de l'analyse de cinq sources de données d'exposition professionnelle au formaldéhyde

Chapitre II : Étude des déterminants de l'exposition au formaldéhyde dans l'industrie des panneaux de bois aggloméré au Québec

Données

- Québec
- Panneaux de bois aggloméré
- 12 usines
- Mesures par prise par une équipe de recherche et des hygiénistes du gouvernement
- 865 mesures en ambiance et individuelles
- Durée des mesures entre 2 et 6 heures
- Période 1984-2002

Méthodes d'analyse

Modèles linéaires mixtes

Résultats principaux

- 57-61% de la variance expliquée
- Procédé, saison, emploi et zone sont déterminants
- Tendance historique : diminution de 1984 à 1995 puis augmentation (~-10% / an puis +10%/ans)
- Diminution des niveaux quand la durée de mesure augmente (~-5% / 60 min.)
- Données gouvernementales > données de recherche (facteur 2,5 à 3)
- Corrélation intra-campagne de mesure (0,16 ambiance, 0,56 individuel)
- Tableaux de prédiction des moyennes géométriques par emploi / zone et procédé pour 2002

Chapitre III : Utilisation de la simulation Monte Carlo pour reconstruire des niveaux d'exposition au formaldéhyde à partir de paramètres de synthèse rapportés dans la littérature

Données

- Revue de littérature sur l'exposition au formaldéhyde dans industrie des panneaux de bois aggloméré
- 13 articles
- 874 mesures d'ambiance et individuelles (dont 142 mesures uniques)
- Période 1965-1995

Méthodes d'analyse

- Simulation des niveaux d'exposition synthétisés dans les articles
- Modèles linéaires mixtes

Résultats principaux

- 38% de la variance des mesures d'ambiances expliquée
- Procédé, emploi et zone sont déterminants
- Tendance historique: réduction de l'exposition de 1965 à 1995 d'un facteur 6
- Précision et biais des équations utilisées pour la simulation entre 5 et 30%
- Tableaux de prédiction des moyennes géométriques par zone, procédé et période historique
- Influences des zones, emplois et procédés similaires à celles observées dans l'étude Québécoise
- Niveaux de la période 1985-1994 semblable à ceux prédits pour 1990 pour les données gouvernementales dans l'étude Québécoise

Chapitre IV : Modélisation statistique des niveaux d'exposition professionnelle dans l'industrie française, 1986-2003

Données

- Banque de données COLCHIC
- Données multisectorielles
- 692 établissements
- Mesures prise par les équipes des caisses régionales d'assurance maladie ou de l'INRS(a)
- 7392 mesures d'ambiance et individuelles
- Durée des mesures entre 1 min et 12 heures
- Période 1986-2003

Méthodes d'analyse

Modèles linéaires mixtes

Résultats principaux

- 35-57% de la variance expliquée
- Industrie et poste de travail sont déterminants
- Tendance historique : -7 à -9% / an de 1984 à 2003
- Données court-terme > données VEMP(b) (facteur~2)
- Diminution des niveaux quand la durée de mesure augmente: ~-10% / 60 min. pour les données VEMP(b), ~10% / 5 min. pour les données court-terme
- Données individuelles prise à cause de suspicion de risque ou d'exposition > visites planifiées (facteur~2)
- Corrélation intra-campagne de mesure (0,80 ambiance et individuel)
- Tableaux de prédiction des niveaux d'exposition par industrie / poste de travail pour 2002

Chapitre V : Exposition professionnelle au formaldehyde dans l'industrie états-unienne à partir des données d'OSHA et comparaison avec des données de la banque française COLCHIC

Données

- Banque de données IMIS
- Données multisectorielles
- Mesures prises par les inspecteurs d'OSHA
- 7148 mesures individuelles
- Mesures de type court-terme et VEMP(b)
- Période 1979-2001

Méthodes d'analyse

- Modèles linéaires mixtes pour les mesures > limite de détection
- Régression logistique de la probabilité d'être non détecté
- Comparaison avec 2971 données de COLCHIC (1986-2001)

Résultats principaux

- 25% de la variance expliquée par le modèle linéaire
- Industrie et raison de visite sont déterminants
- Données court-terme > données VEMP (facteur évoluant de 6 à 2 entre 1979 et 2001)
- Tendance historique . -19% /an puis -4% / an à partir de de 1987 pour les données court terme (-7% / an puis -4% / an pour les données VEMP,b)
- Corrélation intra-campagne de mesure (0.7)
- Tableaux de prédiction des niveaux d'exposition par industrie pour 2001
- Probabilité d'être non détecté varie avec taille de l'établissement, année, État et raison de visite
- Après ajustement pour la durée de mesure et l'année, données de COLCHIC > IMIS (facteur ~1,3)

(a) Institut national de recherche et de sécurité ; (b) Valeur d'exposition moyenne pondérée

Pour les données québécoises, plusieurs variables documentées par l'équipe de recherche de l'IRSST durant les visites d'usines n'ont pas été incluses dans l'analyse car elles manquaient en général dans les rapports d'hygiène des CLSC (p.ex. type de presse, épaisseur des panneaux fabriqués, température ambiante, indices de productivité). L'analyse par modélisation des données des BDEP IMIS et COLCHIC a clairement donné de moins bons résultats pour la banque IMIS, notamment en terme de proportion de la variance expliquée par les variables prédictives. COLCHIC possède en particulier l'avantage sur IMIS de contenir une classification standard de poste de travail. De plus COLCHIC comporte plusieurs variables additionnelles comme la température ambiante, la méthode analytique, ou les systèmes de ventilation.

Plusieurs facteurs sont susceptibles d'expliquer nos observations en regard de cette problématique. En ce qui concerne la littérature, la revue présentée au chapitre III incluait des mesures antérieures à 1995, dont la majorité dataient des années 80, bien avant les propositions de Tielemans et coll. (2002), des groupes de travail européen et états-unien (1995) (Joint ACGIH-AIHA Task Group on Occupational Exposure Databases, 1996;Rajan et coll., 1997;Tielemans et coll., 2002), ou encore d'études soulignant l'importance de l'identification des déterminants de l'exposition (Burstyn et Teschke, 1999). Il est donc possible que les auteurs d'alors aient jugé suffisantes les informations fournies. Cependant, l'explication la plus probable selon nous est que la présentation de données d'exposition, en particulier en vue de leur utilisation a posteriori, ne constitue pas l'objectif principal de la majorité des études. Ainsi, les études les moins bien documentées dans la revue présentée au chapitre III étaient des études épidémiologiques. L'évaluation de l'exposition ne constitue qu'une fraction, historiquement négligée (Stewart et Stenzel, 1999;Stewart et Stewart, 1994) des résultats qui doivent être présentés et discutés dans de telles études. De surcroît, l'incitation de plus en plus forte des publications scientifiques à la soumission de manuscrits le plus courts possible n'incite pas les auteurs à fournir des détails qui ne soient pas de nécessité directe pour l'évaluation de leur travail. En outre, les rapports d'évaluation des équipes d'hygiène des CLSC, au Québec, sont destinés principalement aux entreprises visitées. Il en résulte que les auteurs de ces rapports n'ont pas tendance à « encombrer » leurs résultats d'informations sur les procédés, les tâches, ou les emplois, qui sont parfaitement connues par les destinataires des rapports. Il existe aujourd'hui un besoin grandissant en données d'exposition, notamment pour l'épidémiologie professionnelle qui s'emploie de plus en plus à évaluer des risques de nature chronique causés par de faibles expositions (Kromhout, 2002;Stewart et Stenzel, 1999). Les coûts reliés à la mesure de l'exposition étant élevés, il devient selon nous nécessaire, quel que soit l'objectif principal relié à la prise de mesures d'exposition professionnelle, que leur utilisation à posteriori dans des cadres plus généraux devienne un objectif secondaire systématique. Cet objectif requiert la mise à disposition des futurs utilisateurs d'informations leur permettant d'interpréter au mieux les mesures disponibles.

La problématique associée aux informations entourant les mesures d'exposition est différente dans le cas des BDEP. Ainsi, l'utilisation des données enregistrées dans ces banques pour la surveillance de l'exposition, l'épidémiologie professionnelle ou l'établissement de valeurs limites d'exposition fait partie la plupart du temps des objectifs visés par les organismes qui les ont mis en place (Burns et Beaumont, 1989;Carton et Goberville, 1989;Stamm, 2000). L'utilisation a posteriori des données enregistrées fait donc partie explicitement des objectifs des BDEP. Les problèmes viennent dans ce cas de la quantité d'information enregistrée avec les mesures et de la façon dont cette information est enregistrée. Les recommandations des groupes de travail européen et états-unien, résumées au tableau N°1, englobent un nombre important de caractéristiques, qui peuvent impliquer un travail laborieux de collection et d'enregistrement. Lamontagne et coll. soulignent la nécessité de faire des compromis entre l'abondance d'information et la simplicité d'enregistrement d'une mesure, notamment en regard de la difficulté d'obtention de certaines informations et de la quantité de travail requise de la part du personnel saisissant les données (LaMontagne et coll., 2002a;LaMontagne et coll., 2002b). Le risque est ainsi d'obtenir un faible taux d'enregistrement des informations si leur quantité est trop importante. Vincent et Jeandel on décrit la réorganisation récente de la banque COLCHIC. Les auteurs ont souligné la nécessité de simplifier la saisie des informations et de permettre au personnel générant les données d'obtenir un retour sur l'information qu'ils ont enregistrée pour améliorer les taux d'enregistrement et minimiser les erreurs de saisie (Vincent et Jeandel, 2001). La difficulté d'utilisation de la variable indiquant l'emploi dans la banque IMIS (champs de texte non standardisé) illustre bien la problématique de la façon dont les informations sont enregistrées. L'existence d'une codification standard n'est pas sans écueils puisqu'il faut faire des compromis entre une codification précise mais vulnérable aux classements erronés et une classification trop large pour être utilisable. La récente réorganisation de la banque COLCHIC représente selon nous un exemple d'amélioration continue à suivre dans le domaine des informations entourant les mesures dans les BDEP (Vincent et Jeandel, 2001).

6.1.2 Identification des biais dans les sources de données préexistantes

Si l'on se réfère au schéma conceptuel de la figure 1, nos analyses ne fournissent pas directement d'information sur la représentativité des secteurs industriels inclus dans les données sur le formaldéhyde dans COLCHIC et IMIS par rapport aux secteurs où le formaldéhyde est réellement présent (Case 1 de la figure 1). En effet, ne disposant pas de liste de référence à laquelle comparer la liste des secteurs identifiés dans ces banques, toute validation est impossible. Les variables indiquant les raisons des visites industrielles ayant donné lieu à des mesures, qui varient de plaintes d'employés, soupçon de la présence d'exposition ou de risque, à la sélection quasi aléatoire d'établissements dans des secteurs identifiés comme prioritaires (voir les chapitres IV et V), pointent vers la possibilité à la fois d'inclusion de secteurs où il n'y aurait pas d'exposition au formaldéhyde (p.ex. plainte d'employés non formés en hygiène) et de l'inclusion de secteurs pour lesquels les niveaux d'exposition sont élevés (indication par un hygiéniste, secteur préalablement identifié comme étant à risque). De plus certains secteurs sont explicitement hors de la juridiction des organismes responsables de ces banques (par ex. Défense nationale pour COLCHIC, et le secteur des mines pour IMIS). L'étude la plus informative à ce sujet est probablement celle de Valiante et coll., qui ont observé que les secteurs industriels associés à des niveaux d'exposition à la silice dans IMIS représentaient 54% des emplois retrouvés chez les cas de silicoses du registre de cette maladie au New Jersey aux Etats-Unis (Valiante et coll., 1992). Les auteurs concluent que l'utilisation de plusieurs sources de données est nécessaire pour identifier des industries à risque.

Notre étude fournit également une bonne illustration de la problématique de la case 2 de la figure 1 (Sous-procédés associés à des expositions différentes dans un même secteur d'activité). Ainsi, à la fois dans COLCHIC et IMIS, le secteur des panneaux de bois aggloméré est représenté par une seule catégorie. Or les données de ce secteur présentées aux chapitres II et III montrent que l'un des trois sous-procédés de cette industrie (les panneaux OSB), est associé à des niveaux plus faibles que les deux autres (PB et MDF). Dans la mesure ou tous ces procédés sont représentés par une même catégorie dans une BDEP, le niveau moyen calculé sur l'ensemble du secteur des panneaux de bois fournira une information biaisée sur les niveaux d'exposition (dans ce cas, sous-estimation des niveaux pour PB et MDF et surestimation des niveaux pour OSB). Il est également possible que les données présentes dans la BDEP ne concernent que les sous-procédés associés à de fortes expositions. Dans ce cas, l'application des données à l'ensemble du secteur d'activité causera une surestimation des expositions réelles dans l'ensemble du secteur. Cette problématique peut être partiellement contournée par une connaissance approfondie des procédés à l'intérieur d'un secteur d'activité, en particulier par l'identification de ceux qui sont associés à des expositions très faibles ou nulles, si l'on fait l'hypothèse qu'ils ne sont pas ou peu représentés dans la BDEP.

Les biais causés par la sélection non aléatoire des usines d'un secteur (case 3 dans la figure 1) ont fait l'objet de la majorité des études publiées sur IMIS (voir chapitre I tableau 1). Étant donné qu'une proportion importante des mesures dans IMIS provient de visites causées par des plaintes d'employés, l'hypothèse la plus souvent énoncée suppose que les usines visitées dans ce cadre correspondent à des milieux où l'exposition est moins bien maîtrisée. Ces données seraient donc biaisées vers le haut par rapport à celles prises lors de visites planifiées à l'avance, qui correspondent à une sélection aléatoire d'usines à l'intérieur d'un secteur d'activité. De façon générale, les études utilisant des mesures descriptives (p.ex. médiane des expositions par secteur d'activité) ont rapporté de plus fortes expositions pour les visites causées par des plaintes (voir chapitre I tableau N°2). En revanche les études ayant cherché à tenir compte de l'influence d'autres variables (emplois, tendances temporelles) n'ont mis en évidence que des effets faibles voire non détectables. Nos résultats d'analyse des données d'exposition au formaldéhyde dans IMIS ont montré une influence marginale (différence de ~5%) du type de visite sur les niveaux d'exposition détectés. Cependant les mesures causées par une plainte avaient significativement plus de chance de donner un résultat non détecté que les autres types de visite. Dans le cas de COLCHIC, nos analyses montrent des différences plus importantes reliées à la raison de la visite. Ainsi, les visites causées par le soupçon d'exposition ou de risque, ou encore l'implantation de systèmes de ventilation, correspondaient à des expositions plus élevées d'un facteur 2 que celles associées à des campagnes systématiques. Ces résultats sont cependant mitigés par le fait que nous n'avons observé une influence de la raison de la visite que pour les mesures individuelles. Finalement, il apparaît important d'étudier de façon systématique les variables reliées au stratégies de mesure dans une BDEP afin de pouvoir corriger les estimations de niveaux d'exposition en conséquence. À la fois dans le cas d'IMIS et de COLCHIC, l'identification de la raison de la visite comme déterminant et son inclusion dans les modèles statistiques ont permis d'effectuer de telles corrections.

Dans cette étude, une tendance nette à la concentration des mesures disponibles dans un nombre limité d'emplois et de postes de travail (case 4 de la figure 1) a été mise en évidence dans les données issues de la revue de littérature dans le secteur des panneaux de bois et dans COLCHIC. Cette tendance n'était pas présente dans les données issues des dossiers des CLSC et ne pouvait pas être évaluée dans IMIS puisque cette dernière ne contient pas de variable identifiant un emploi standardisé. De plus, les emplois cités dans la revue de littérature correspondaient aux deux groupes d'exposition identifiés dans les données québécoises comme associés aux plus fortes expositions. Ces résultats confirment le danger d'interprétation erronée des données d'exposition pour lesquelles aucune information n'est disponible concernant l'emploi ou le poste de travail évalué, notamment si l'on applique les niveaux d'exposition à tous les corps de métiers. Le problème est cependant contourné si l'information est disponible.

Dans le schéma conceptuel des biais présenté à la figure 1, les biais correspondant aux cases 5 et 6, qui englobent le choix des employés et des périodes échantillonnées, sont probablement les plus complexes à évaluer car ils dépendent à la fois de la stratégie d'échantillonnage utilisée et de l'expertise du mesureur. Ainsi, l'hygiéniste peut avoir visé à mesurer les conditions normales d'exposition, à identifier l'employé le plus exposé, ou encore à évaluer une période particulière associée à une forte exposition. Il peut également avoir réalisé une sélection aléatoire de l'employé et de la période mesurée. De plus, la capacité de l'intervenant en hygiène à identifier les conditions qu'il vise à mesurer (p.ex. les conditions «normales » d'exposition, ou encore les pires conditions) détermine l'amplitude du biais par rapport à un échantillonnage purement statistique (Olsen et coll., 1997). La problématique est en outre compliquée par le fait que les stratégies d'échantillonnage sont rarement définies de façon explicite. Au Québec, selon le Règlement sur la santé et la sécurité au travail (Gouvernement du Québec, 18 juillet 2001), le Guide d'échantillonnage des contaminants de l'air en milieu de travail (Direction des opérations, 2000) doit servir de référence aux équipes d'hygiène des CLSC pour les stratégies de mesure. Cependant ce document, bien qu'il décrive les différentes approches possibles, ne recommande pas de stratégie spécifique. Aux États-Unis, il est plausible de penser que les inspecteurs de l'Occupational Safety and Health Administration (OSHA) utilisent la stratégie proposée par NIOSH, qui recommande la sélection des employés les plus exposés pour évaluer la conformité aux normes (Leidel et coll., 1977). Cependant Jones et coll., sur la base d'entrevues avec des inspecteurs d'OSHA, rapportent que cette stratégie n'est pas systématiquement employée (Jones et coll., 1986). Vincent et Jeandel ont précisé dans chacun de leurs articles sur COLCHIC que les données correspondent probablement à des stratégies de type « pire des cas » (Vincent et Jeandel, 2001;Vincent et Jeandel, 2002). Nos observations sur les données québécoises, sur COLCHIC et sur IMIS indiquent que les moyennes géométriques estimées des concentrations de formaldéhyde diminuent lorsque la durée de mesure augmente, les données court terme étant plus élevées que les données de type VEMP. En théorie, un échantillonnage aléatoire de périodes plus courtes à l'intérieur d'un quart de travail devrait donner lieu à des moyennes arithmétiques identiques pour les mesures, quelles que soient leur durée (Rappaport et coll., 1988). Les mesures à court terme étant en général plus variables que les mesures plus longues (Kumagai et Matsunaga, 1995;Kumagai et Matsunaga, 1999), ceci implique que les moyennes géométriques des mesures devraient diminuer lorsque la durée de mesure est plus courte. Nos résultats, contraires à ce principe, suggèrent donc que les mesures que nous avons analysées représentent des tâches ou des moments de la journée où l'exposition est présente ou plus élevée que la moyenne de la journée plutôt qu'une fraction représentative de cette journée. Ils rejoignent les constatations d'Olsen et coll., qui ont observés des résultats plus élevés pour des mesures prises durant des tâches générant de l'exposition par rapport à des mesures prises de façon aléatoire durant la journée (Olsen et coll., 1991). De plus, les mesures effectuées par les équipes d'hygiène des CLSC (voir chapitre II), étaient plus élevées que celles prises par l'équipe de recherche de l'IRSST après correction pour la durée d'échantillonnage, indiquant, pour une même durée d'échantillonnage, la sélection de périodes d'exposition élevées. Globalement, notre étude appuie l'hypothèse énoncée notamment par Olsen et coll. (Olsen et coll., 1997), selon laquelle, dans le cadre de la vérification de l'acceptabilité des conditions d'exposition dans un milieu de travail visité, les intervenants en hygiène privilégient la mesure des situations spécifiques, en particulier par l'évaluation de tâches générant de fortes expositions, dans le but d'optimiser l'utilisation des ressources disponibles pour le mesurage.

Il existe des indices suggérant l'existence dans la banque IMIS d'un sous-enregistrement des résultats mesurés par les inspecteurs d'OSHA (Jones et coll., 1986;Mendeloff, 1984), qui correspond à la problématique de la case 7 dans la figure 1. Les deux études disponibles rapportent cependant des résultats contradictoires sur l'existence d'un sous-enregistrement préférentiel des résultats faibles ou non détectés. Notre analyse des mesures de formaldéhyde dans IMIS ne permet pas de conclure dans ce sens. Nous avons plutôt observé une proportion importante de non détectés par rapport aux niveaux moyens enregistrés, ce qui indiquerait plutôt une surreprésentation de mesures correspondant à des milieux dont l'atmosphère n'est pas contaminée par le formaldéhyde. Une partie importante de ces mesures ont probablement été prises lors de visites durant lesquelles les inspecteurs mesurent des « zéro » à la suite de préoccupations d'employés (voir chapitre V). Nos résultats tendent cependant à confirmer qu'il existe une variabilité dans les pratiques d'enregistrement, notamment en fonction des différentes unités géographiques de OSHA.

En résumé, grâce à l'étude conjointe de plusieurs sources de données d'exposition professionnelle au formaldéhyde, cette étude a montré l'existence de biais reliés à la sélection des entreprise où des mesures sont effectuées, aux systèmes de classification employés pour les secteurs d'activité, les emplois / postes de travail, et aux stratégies de mesure employées. De plus nos résultats suggèrent que ces biais peuvent être au moins partiellement compensés si les informations accompagnant les mesures sont adéquates et si des études de validation sont conduite avec des données mesurées dans des contextes contrôlés.

6.1.3 Comparaison entre les différentes sources de données

Lors de la comparaison des données d'exposition au formaldéhyde québécoises et issues de la revue de littérature dans le secteur des panneaux de bois aggloméré, les déterminants commun aux deux sources de données (emploi, poste de travail, procédé) avaient une influence très semblable sur l'exposition. Les niveaux d'exposition estimés à partir de la littérature, après correction pour la période temporelle et l'emploi / zone de travail étaient similaires à ceux estimés à partir des données des CLSC, eux-même étant plus élevés que les données mesurées par l'IRSST d'un facteur 2,5 à 3. La comparaison entre IMIS et COLCHIC a également révélé plusieurs déterminants communs de l'exposition et de sa variabilité. Ainsi, les tendances temporelles, les différences court-terme / VEMP, l'influence des secteurs d'activité et la corrélation estimée entre les mesures effectuées durant la même visite étaient similaires dans les deux banques. En terme absolu, nous avons estimé, après correction pour plusieurs facteurs confondants, que les niveaux d'exposition au formaldéhyde étaient 30% plus élevés dans COLCHIC que dans IMIS. Vinzents et coll. ont observé de plus grandes différences entre plusieurs banques de données européennes (rapport maximal de 4), mais les analyses étaient restreintes à la comparaison de moyennes géométriques de mesures agglomérées sur une période de 10 ans sans correction pour l'emploi ou la durée d'échantillonnage. Les différences observées dans notre étude entre IMIS et COLCHIC, ainsi qu'entre les données de la littérature et québécoises, sont comparables à celles mesurées lors d'autres exercices de comparaison rapportés dans la littérature (Burstyn et coll., 2002;Stewart et coll., 2003). Globalement, nos résultats indiquent un bon accord entre les différentes sources analysées, d'autant plus satisfaisant que les données proviennent de pays différents, ont été mesurées dans des contextes réglementaires distincts en utilisant des stratégies de mesures variables. S'ils sont confirmés par d'autres études de comparaison, impliquant notamment d'autres substances, ces résultats suggèrent la possibilité d'utiliser conjointement différentes sources de données pour augmenter le nombre de résultats disponibles pour une évaluation.

6.2 Exposition professionnelle au formaldéhyde

6.2.1 Niveaux d'exposition estimés à partir des différentes sources de données

Grâce à l'emploi de la modélisation statistique, des portraits de l'exposition professionnelle au formaldéhyde prenant en compte l'influence de variables déterminantes ont pu être élaborés à partir de chacune des sources de données disponibles, fournissant des informations à la fois sur les niveaux d'exposition, leurs déterminants, et leur variabilité. Les portraits élaborés comportaient de plus une dimension historique puisque des tendances temporelles dans les niveaux d'exposition ont été également identifiées et quantifiées.

Les analyses réalisées dans le secteur des panneaux de bois aggloméré, présentées aux chapitres II et III ont mis en évidence une diminution importante des niveaux d'exposition des années 60 à nos jours. L'analyse des données québécoises, couvrant la période 1984-2002, a néanmoins montré une stabilisation, et une ré-augmentation des niveaux mesurés à partir de 1995 (niveaux les plus récent équivalents à ceux des années 1987-1988). À l'intérieur du secteur, les procédés MDF et PB sont associés à des niveaux plus élevés que le procédé OSB (d'un facteur 2 à 5), pour lequel les résines employées sont plus résistantes à l'hydrolyse. Nous avons également pu établir des groupes d'exposition similaires à partir des emplois de ce secteur. Ainsi les emplois impliquant une proportion importante du quart de travail passée hors des salles de contrôle dans la zone de production principale sont associés aux plus fortes expositions. Estimée pour l'année 2002, la moyenne géométrique correspondant à ce groupe d'exposition est 0,22 mg/m^3 pour les données mesurées par l'équipe de recherche de l'IRSST. À titre de comparaison, les moyennes géométriques estimées à partir de l'analyse de COLCHIC, pour les données individuelles de type VEMP et l'année 2003, variaient de 0,06 à 0,10 mg/m^3 entre les différents postes de travail dans le secteur de la fabrication de panneaux de bois et de laminés. Pour IMIS la moyenne géométrique estimée pour 2001 dans le même secteur était 0,19 mg/m^3 pour les données individuelles de type VEMP. Dans les deux derniers cas, la codification industrielle ne permettait pas d'identifier le sous-procédé (OSB, PB, MDF ou autre).

Les analyses multisectorielles présentées aux chapitres IV et V ont également mis en évidence une diminution importante des niveaux d'exposition au formaldéhyde au fil des années dans les banques IMIS et COLCHIC. Ainsi pour COLCHIC les niveaux individuels et d'ambiance, à la fois pour les mesures court-terme et de type VEMP, ont diminué en moyenne de 75% entre 1986 et 2003. Pour IMIS, entre 1979 et 2001, les concentrations individuelles ont diminué respectivement de 68% et 90% pour les données de type VEMP et court terme. Ces tendances sont compatibles avec celles observées par Symanski et coll. pour plusieurs ensembles de données à partir d'une revue de littérature (Symanski et coll., 1998a;Symanski et coll., 1998b). Elles indiquent globalement une amélioration significative des conditions d'exposition au formaldéhyde du début des années 80 à nos jours.

Les deux analyses multisectorielles ont également permis de mettre en évidence des niveaux correspondant aux mesures court-terme supérieurs à ceux correspondant à des mesures de type VEMP, d'un facteur moyen de 2, variable selon les secteurs d'activité. Ces résultats montrent que des expositions de courte durée significativement plus élevées que la moyenne de la journée ont lieu dans de nombreux secteurs d'activité.

Le tableau N°6 présente une synthèse des niveaux d'exposition estimés à partir de l'analyse de COLCHIC et IMIS dans les secteurs d'activité correspondant aux plus fortes expositions. Pour chacune des banques les moyennes géométriques des mesures individuelles de type court-terme et VEMP sont estimées pour l'année la plus récente, pour les 3 secteurs d'activité correspondant aux estimations les plus élevées.

Tableau N°6 : Moyennes géométriques estimées des concentrations individuelles de formaldéhyde pour les secteurs d'activité associés aux plus fortes expositions dans IMIS et COLCHIC.

Secteur d'activité	MG[A] (mg/m^3)
IMIS – mesures de type VEMP	
Fabrication de planchers et de pièces dimensionnées de bois de feuillus et de pièces de charpenterie	0,20
Produits de bois reconstitué	0,19
Fabrication de caisses en fibres solides ou en carton ondulé	0.13
IMIS – mesures de type court-terme	
Salons funéraires et crématorium	0,39
Produits de bois reconstitué	0,35
Fabrication de planchers et de pièces dimensionnées de bois de feuillus et de pièces de charpenterie	0,32
Collèges, Universités et écoles professionnelles	0,30
COLCHIC[B] – mesures de type VEMP (ajustées pour une durée de 8 heures)	
Analyses biologiques et bactériologiques	0,28
Travail de charpenterie	0,26
Analyses anatomopathologiques	0,22
COLCHIC[B] – mesures de type court-terme	
Analyses anatomopathologiques	0,86
Analyses biologiques et bactériologiques	0,54
Travail de charpente	0,42

(A) Moyenne géométrique
(B) Estimation pour le poste de travail correspondant à la moyenne géométrique la plus élevée

Le tableau N°6 révèle les concentrations les plus élevées autour de 0,3 mg/m^3 pour les mesures de type VEMP et entre 0,5 et 0,9 mg/m^3 pour les mesures court-terme. Si l'on se réfère aux analyses de risques évoquées au chapitre I, ces résultats suggèrent un très faible risque cancérogène pour les niveaux d'exposition existant actuellement dans l'industrie. En revanche ils n'excluent pas, en particulier durant de courtes durées, la survenue d'irritation.

Parmi les sources de données analysées dans ce travail, les données québécoises, issues de la littérature et de COLCHIC incluaient des mesures d'ambiance en plus des mesures individuelles. En règle générale, les mesures d'ambiance ne constituent pas une alternative adéquate aux mesures individuelles pour estimer l'exposition de travailleurs (Perkins, 1997). Nos analyses ont permis d'identifier des déterminants communs aux deux types de mesures (en particulier des tendances historiques similaires) et d'autre spécifiques à l'un ou l'autre (p.ex. la ventilation locale ou générale). Les mesures d'ambiance étaient plus variables que les mesures individuelles. De façon générale, elles étaient également plus élevées que les mesures individuelles dans des secteurs d'activité où l'exposition est causée par la contamination de l'atmosphère de la zone de production entière (en particulier dans le secteur des panneaux de bois aggloméré). Les mesures individuelles, en revanche, étaient plus élevées dans des secteurs où l'exposition est plutôt reliée à des tâches spécifiques (p.ex. secteur des analyses biologiques et anatomopathologiques). Nos résultats confirment le besoin de connaître en détail les mécanismes de génération de l'exposition pour pouvoir utiliser les mesures d'ambiance comme substitut de données individuelles.

6.2.2 Limites des portraits de l'exposition au formaldéhyde présentés

Les estimations des niveaux d'exposition au formaldéhyde, présentées à la section 6.2.1 et détaillées dans les précédents chapitres, comportent plusieurs limites quant à leur interprétation directe comme matrice emploi/secteur-exposition pour l'analyse du risque à la santé posé par le formaldéhyde ou pour une étude épidémiologique.

Les données que nous avons analysées ne permettaient pas d'évaluer des différences entre les travailleurs, qui sont plausibles et ont été mises en évidence dans plusieurs études (Kromhout et coll., 1993;Symanski et coll., 2001;Symanski et coll., 2006;Van Tongeren et coll., 2006). L'ampleur des différences entre travailleurs détermine notamment l'importance de l'atténuation de la courbe dose-réponse causée par l'erreur de classement de l'exposition dans les études épidémiologiques qui utilisent un système de classification par groupe d'exposition (par ex. par emploi) (Heederick et coll., 1991;Werner et Attfield, 2000). De plus, ainsi que discuté dans la section 6.1, les sources disponibles ne représentent pas un échantillon aléatoire des populations exposées et un certain nombre de biais sont susceptibles d'avoir faussé nos estimations. Finalement, la plupart de nos résultats concernent des données de type VEMP. En terme d'indice d'exposition, si l'on considère le risque d'irritation posé par le formaldéhyde, c'est la distribution à long terme des concentrations instantanées subies par un travailleur, en particulier la fraction de ces expositions dépassant le seuil d'irritation qui apparaît comme le meilleur indicateur du risque. De plus, bien que l'exposition cumulative soit considérée en général comme le meilleur indice d'exposition pour le risque cancérogène, il semble que pour le formaldéhyde ce risque soit plutôt associé à sa cytotoxicité et donc que la fréquence de dépassement du seuil de cytotoxicité soit l'indice d'exposition le plus appropriée (Wibowo, 2003). Or les données de type VEMP ne permettent pas de caractériser directement les distributions d'exposition instantanées. À ce titre, même s'il est logique de considérer comme de meilleurs indicateurs de risque les estimations pour les mesures court-terme, disponibles dans le cas de COLCHIC et IMIS, nos analyses suggèrent que les mesures court-terme dans ces banques représentent des évènements particuliers associés à de fortes expositions. Ils ne sont donc pas représentatifs de la distribution réelle des mesures court-terme mais plus probablement de la partie droite de cette distribution. Rappaport et Roach ont proposé un outil permettant d'estimer les distributions de mesures court-terme à partir de la distribution des mesures de type VEMP, basé sur le fait que ces distributions possèdent théoriquement la même moyenne arithmétique (Rappaport et coll., 1988). À titre d'illustration, d'après ces résultats, au plus 5% des valeurs d'une distribution lognormale dépassent 4 fois sa moyenne arithmétique, quelque soit l'écart type géométrique de cette distribution. En l'absence de meilleur indicateur, cet outil pourrait être utilisé à partir des estimations de moyennes arithmétiques présentés dans les tableaux N°6 et N°8 du chapitre IV et dans le tableau N°4 du chapitre V pour estimer les fractions maximales de dépassement d'une valeur seuil par les expositions court-terme.

Les estimations de niveaux d'exposition présentées dans ce travail ont été réalisées à partir de modèles statistiques de régression multiple, plus précisément les modèles linéaires mixtes (Pinheiro et Bates, 2000). Cette méthode représente un outil de choix pour l'analyse des données d'exposition puisqu'elle permet d'évaluer l'influence conjointe de multiples variables sur les niveaux d'exposition tout en tenant compte de l'existence de structures de variance-covariance complexes dans les données (Burdorf et Van Tongeren, 2003). En particulier, nos analyses ont permis de mettre en évidence la présence d'une corrélation significative entre les mesures prises durant une même campagne d'échantillonnage. Elles indiquent l'importance d'évaluer ce type de corrélation lors de l'analyse de données d'exposition puisque ne pas en tenir compte conduit généralement à une sous-estimation de l'incertitude et, dans le cas de données sévèrement non compensées (« unbalanced data » en anglais), peuvent causer un biais dans l'estimation. Ces résultats confirment de plus les recommandations d'autres auteurs de réaliser des évaluations d'hygiène industrielle sur des période de temps supérieures à quelques jours consécutifs (Buringh et Lanting, 1991;Deadman et coll., 1996;Francis et coll., 1989;Symanski et Rappaport, 1994). Cependant, l'analyse par modélisation statistique est limitée par la quantité d'information disponible entourant les données et par l'influence des variables disponibles sur les données d'exposition. Les modèles développés à partir des données québécoises (chapitre II) et des données de COLCHIC (Chapitre IV) ont permis d'expliquer plus de 50% de la variance des concentrations de formaldéhyde, ce qui laissent penser que les principaux déterminants de l'exposition au formaldéhyde dans ces sources de données ont été identifiés. En revanche, les modèles construits à partir des données de la littérature et des données de IMIS n'ont expliqué respectivement que 38 et 25% de la variabilité des niveaux d'exposition, suggérant que d'importants déterminants n'ont pu être inclus dans l'analyse, ce qui réduit la valeur des prédictions présentées. De plus, malgré le nombre important de données dans la plupart des sources (de plusieurs centaines à plusieurs milliers), la variabilité importante des niveaux d'exposition a limité la puissance statistique disponible pour détecter les effets de certaines variables. À titre d'illustration, les tendances temporelles estimées par les modèles présentés aux chapitres IV et V correspondent à des tendances moyennes sur l'ensemble des secteurs industriels. Il est plausible et même probable que ces tendances varient en réalité d'un secteur d'activité à l'autre. Cependant les différences n'étaient pas d'ampleur suffisante et les données en nombre suffisant pour que cette interaction entre le secteur d'activité et la variable identifiant l'année de mesurage soit incluse dans les modèles. Finalement, les modèles utilisés, bien que validés de façon interne au moyen d'outils diagnostiques graphiques, n'ont pu être validés de façon externe. La généralisation directe des résultats obtenus à l'extérieur des ensembles de données analysées n'est donc pas possible.

Les limites présentées dans cette section ne sont pas différentes de celles évoquées dans d'autres analyses (voir p.ex. (Burstyn et coll., 2002)). En particulier elles apparaissent modérées par comparaison à l'incertitude importante associée à l'évaluation de l'exposition professionnelle dans les études épidémiologiques (Ahrens et Stewart, 2003). Dans notre étude, les pourcentages de variance expliquée par les variables prédictives étaient comparables à ceux obtenus en général dans ce type d'analyse (Burstyn et Teschke, 1999), les estimations ont été corrigées pour atténuer des biais reliés aux stratégies de mesure et les différences observées entre les cinq sources de mesures disponibles étaient modérées.

Nous concluons donc que le travail présenté fournit un portrait utile de l'évolution des conditions d'exposition au formaldéhyde durant les deux dernières décennies. Il pourrait être utilisé en surveillance de l'exposition pour identifier des secteurs où les niveaux d'exposition sont encore élevés, où d'autres pour lesquels l'exposition semble avoir été maîtrisée à un niveau acceptable. En épidémiologie du travail, les portraits historiques élaborés pourraient être utilisés dans des études cas-témoin de population (Gérin et coll., 1985;Siemiatycki, 1984). Dans ces études les sujets proviennent de la population générale et ont pu être employés dans de multiples secteurs d'activité. Nos données pourraient être utilisées dans ce contexte pour aider à l'élaboration d'une matrice emploi-exposition historique. D'autre part, l'identification de groupes d'exposition similaire dans le secteur des panneaux de bois aggloméré à partir des données québécoises et de la littérature pourrait servir à l'élaboration de questionnaires spécifiques à cette industrie utilisés par les experts chargé d'évaluer l'exposition (Stewart et coll., 1998). Dans les études de cohorte les sujets sont sélectionnés permis un nombre limité d'établissement dans un secteur industriel spécifique. Les méthodes d'évaluation de l'exposition dans ce type d'étude sont en général constituées de plusieurs procédures utilisées en fonction de la disponibilité de mesures d'exposition (Stewart et coll., 1996). Les estimations présentées dans notre travail pourraient servir dans les circonstances où aucune mesure n'est disponible pour assister les experts.

6.3 Conclusion générale

À partir de l'analyse approfondie de cinq sources de données d'exposition au formaldéhyde, nous avons démontré que plusieurs biais sont présents dans les différentes sources si l'on cherche à caractériser l'exposition moyenne dans la population active. Nos résultats suggèrent cependant que si les mesures sont accompagnées d'informations adéquates, il est possible de quantifier certains biais et de corriger les estimations en conséquence. Nous recommandons donc que l'utilisation future des données soit envisagée systématiquement lorsque l'exposition professionnelle est mesurée, pour garantir l'enregistrement de l'information nécessaire à leur interprétation correcte. La concordance entre les différentes sources de données analysées, en particulier au niveau des déterminants de l'exposition, est rassurante en regard de l'utilisation de sources de données mesurées dans des contextes variables. D'autres études devraient êtres réalisées pour explorer les différences entres les BDEP accessibles au public et pour mieux caractériser le biais relié aux pratiques usuelles en hygiène par rapport au design aléatoire. Les différents portraits d'exposition présentés dans ce travail fournissent une mise à jour importante des connaissances sur l'exposition professionnelle au formaldéhyde, en particulier de son évolution durant les deux dernières décennies. Bien que non interprétables directement comme des matrices emploi-exposition, ils constituent néanmoins des outils utilisables à la fois pour la surveillance de l'exposition et l'épidémiologie professionnelle.

À notre connaissance, le présent travail est le premier dans lequel les modèles empiriques dits linéaires mixtes ont été employés pour établir des portraits historiques et multisectoriels de l'exposition professionnelle à une substance chimique à partir de BDEP, et pour modéliser non seulement les concentrations mesurées mais aussi leur variabilité et leurs structure de corrélation. L'utilisation de l'analyse détaillée de cinq sources de données d'exposition à une même substance puis de leur comparaison pour identifier et quantifier d'éventuels biais confère également un caractère d'originalité à notre travail. Finalement, une nouvelle méthode d'interprétation des données d'exposition issues de la littérature a été élaborée dans le cadre cette étude.

CHAPITRE VII

BIBLIOGRAPHIE

7. BIBLIOGRAPHIE

Ahrens W, Stewart PA. (2003) Retrospective exposure assessment. In: Nieuwenhuijsen M, Editor. Exposure assessment in occupational and environmental epidemiology. New York, NY: Oxford University Press. p. 103-118.

Bégin D, Gérin M, Adib, Fournier C, DeGuire L. (1995) Development of an Occupational Exposure Data Bank on the Territory of a Department of Community Health in Montreal. Applied Occupational and Environmental Hygiene; 10 355-360.

Botkin A, Conway H. (1995) Relevance of Exposure Data to Regulatory Impact Analyses: Overcoming Availability Problems. Applied Occupational and Environmental Hygiene; 10 383-390.

Bredendiek-Kämper S. (2001) Do EASE Scenario fit workplace reality ? A validation study of the EASE model. Applied Occupational and Environmental Hygiene; 16 182-187.

Burdorf A, Van Tongeren M. (2003) Variability in Workplace Exposures and the Design of Efficient Measurement and Control Strategies. Annals of Occupational Hygiene; 47 95-99.

Buringh E, Lanting R. (1991) Exposure variability in the workplace: Its implications for the assessment of compliance. American Industrial Hygiene Association Journal; 52 6-13.

Burns DK, Beaumont PL. (1989) The HSE National Exposure Database - (NEDB). Annals of Occupational Hygiene; 33 1-14.

Burstyn I, Boffetta P, Burr GA, Cenni A, Knecht U, Sciarra G, et al. (2002) Validity of empirical models of exposure in asphalt paving. Occupational and Environmental Medicine; 59 620-624.

Burstyn I, Kromhout H, Boffetta P. (2000) Literature Review of Levels and Determinants of Exposure to Potential Carcinogens and Other Agents in the Road Construction Industry. American Industrial Hygiene Association Journal; 61 715-726.

Burstyn I, Teschke K. (1999) Studying the Determinants of Exposure: A Review of Methods. American Industrial Hygiene Association Journal; 60 57-72.

Caldwell DJ, Armstrong TW, Barone MJ, Suder JA, Evans MJ. (2001) Lessons Learned While Compiling a Quantitative Exposure Database from the Published Litterature. Applied Occupational and Environmental Hygiene; 16 174-177.

Caldwell DJ, Armstrong TW, Barone NJ, Suder JA, Evans MJ. (2000) Hydrocarbon Solvent Exposure Data: Compilation and Analysis of the Litterature. American Industrial Hygiene Association Journal; 61 881-894.

Carton B. (1995) COLCHIC Chemical Exposure Database: Information on Lead and Formaldehyde. Applied Occupational and Environmental Hygiene; 10 345-350.

Carton B, Goberville V. (1989) La base de données COLCHIC. Cahiers de notes documentaires Securité et hygiène du travail; 134 29-38.

Carton B, Jeandel B. (1993) The lead exposure hazard - Information supplied by the COLCHIC chemical exposure data base. Cahiers de notes documentaires Securité et hygiène du travail 2nd quarter; 229-236.

CIIT. (1999) Formaldehyde: Hazard Characterization and Dose-Response Assessment for Carcinogenicity by the Route of Inhalation. 1999

CIRC. (1995) Monographies du CIRC sur l'Évaluation des Risques de Cancérogénicité pour l'Homme Vol.62: Poussières de bois et formaldéhyde. Lyon: Centre International de Recherche sur le Cancer, Organisation Mondiale de la Santé.

CIRC. (Sous presse) Monographies du CIRC sur l'Évaluation des Risques de Cancérogénicité pour l'Homme Vol.88: Formaldéhyde, 2-Butoxyéthanol et 1-tert-Butoxy-2-propanol. Lyon: Centre International de Recherche sur le Cancer, Organisation Mondiale de la Santé.

Coble JB, Lees PS, Matanoski G. (2001) Time trends in exposure measurements from OSHA compliance inspections of the pulp and paper industry. Applied Occupational and Environmental Hygiene; 16 263-270.

Conolly RB, Kimbell JS, Janszen D, Schlosser PM, Kalisak D, Preston J, et al. (2004) Human Respiratory Tract Cancer Risks of Inhaled Formaldehyde: Dose-Response Predictions Derived From Biologically-Motivated Computational Modeling of a Combined Rodent and Human Dataset. Toxicological Science; 82 279-296.

Deadman JE, Armstrong BG, Thériault GP. (1996) Exposure to 60-Hz magnetic and electric fields at a Canadian electric utility. Scandinavian Journal of Work, Environment and Health; 22 415-424.

Direction des opérations. (2000) Guide d'échantillonnage des contaminants de l'air en milieu de travail (7ième édition). Montréal, QC: Institut de recherche en santé et en sécurité du travail.

Dutch Expert Committee on Occupational Standards. (2003) Formaldehyde - Health-based recommended occupational exposure limit. The Hague: Health Council of the Netherlands (publication no. 2003/02OSH)

Francis M, Selvin S, Spear R, Rappaport SM. (1989) The Effect of Autocorrelation on the Estimation of Workers' Daily Exposures. American Industrial Hygiene Association Journal; 50 37-43.

Freeman CS, Grossman EA. (1995) Silica exposures in workplaces in the United States between 1980 and 1992. Scandinavian Journal of Work, Environment and Health; 21 suppl 2 47-49.

Froines JR, Baron S, Wegman DH, O'Rourke S. (1990) Characterization of the Airborne Concentrations of Lead in U.S. Industry. American Journal of Industrial Medicine; 18 1-17.

Froines JR, Wegman DH, Dellenbaugh CA. (1986) An Approach to the Characterization of Silica Exposures in U.S. Industry. American Journal of Industrial Medicine; 10 345-361.

Gérin M, Siemiatycki J, Kemper H, Bégin D. (1985) Obtaining Occupational Exposure Histories in Epidemiologic Case-Control Studies. Journal of Occupational Medicine; 27 420-426.

Goldberg M, Kromhout H, Guenel P, Fletcher AC, Gerin M, Glass DC, et al. (1993) Job exposure matrices in industry. International Journal of Epidemiology; 22 S10-S15.

Goldman LR, Gomez L, Greenfield S, Hall L, Hulka BS, Kaye WE, et al. (1992) Use of exposure databases for status and trend analysis. Archives of Environmental Health; 47 430-438.

Gomez MR. (1993) A Proposal to Develop a National Occupational Exposure Databank. Applied Occupational and Environmental Hygiene; 8 768-774.

Gomez MR. (1997) Factors associated with exposure in occupational safety and health administration data. American Industrial Hygiene Association Journal; 58 186-195.

Gouvernement du Québec. (18 juillet 2001) Règlement sur la santé et la sécurité du travail. Gazette officielle du Québec, Partie 2, Lois et règlements; 133 5020-5133.

Goyer N, Perrault G, Beaudry C, Bégin D, Bouchard M, Carrier G, et al. (2004) Impact d'un abaissement de la valeur d'exposition admissible au formaldéhyde. Montréal: Institut de recherche Robert-Sauvé en santé et en sécurité du travail (R-386).

Heederick D, Boleij JSM, Kromhout H, Smid T. (1991) Use and analysis of exposure monitoring data in occupational epidemiology - An example of an epidemiological study in the Dutch animal food industry. Applied Occupational and Environmental Hygiene June; 6 458-464.

Joint ACGIH-AIHA Task Group on Occupational Exposure Databases. (1996) Data Elements for Occupational Exposure Databases: Guidelines and Recommendations for Airborne Hazards and Noise. Applied Occupational and Environmental Hygiene; 11 1294-1311.

Jones CA, Weld L, Gray W, Greenlee P, Quinn M, Wiarda E. (1986) The Sampling and Reporting Processes in OSHA MIS Data. Cincinnati, OH: United States National Institute for Occupational Safety and Health, Grant No. R03-OH-002135 (NTIS No. PB2003-104588).

Kauppinen TP, Toikkanen J, Pukkala E. (1998) From Cross-Tabulations to Multipurpose Exposure Information Systems: A New Job-Exposure Matrix. American Journal of Industrial Medicine; 33 409-417.

Kromhout H. (2002) Design of measurements strategies for workplace exposures. Occupational and Environmental Medicine; 59 349-354.

Kromhout H, Symansky E, Rappaport SM. (1993) A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Annals of Occupational Hygiene; 37 253-270.

Kumagai S, Matsunaga I. (1995) Changes in the distribution of short-term exposure concentration with different averaging time. American Industrial Hygiene Association Journal; 1 24-31.

Kumagai S, Matsunaga I. (1999) Within-shift variability of shirt-term exposure to organic solvent in indoor workplaces. American Industrial Hygiene Association Journal; 60 16-21.

LaMontagne AD, Herrick RF, Van Dyke MV, Martyny JW, Ruttenber AJ. (2002a) Exposure Databases and Exposure Surveillance: Promise and Practice. American Industrial Hygiene Association Journal; 63 205-212.

LaMontagne AD, Van Dyke MV, Martyny JW, Simpson MW, Holwager LA, Clausen BM, et al. (2002b) Developpment and Piloting of an Exposure Databse and Surveillance System for DOE Cleanup operations. American Industrial Hygiene Association Journal; 63 213-224.

Lauwerys RR. (1999) Toxicologie industrielle et intoxications professionnelles. Paris: Masson.

Leidel NA, Busch KA, Lynch JR. (1977) Occupational Exposure Sampling Strategy Manual. Cincinnati, OH: United States Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, National Institute for Occupational Safety and Health (DHEW (NIOSH) Publication No. 77-173).

Page web: *http://www.cdc.gov/niosh/pdfs/77-173.pdf:*

Leroyer C, Dewitte J-D. (1999) Asthme au formaldéhyde. In: Bessot J-C, Pauli G, Editors. L'asthme professionnel. Paris: Éditions Margaux Orange. p. 353-364.

Linch KD, Miller WE, Althouse RB, Groce DW, Hale JM. (1998) Surveillance of Respirable Crystalline Silica Dust Using OSHA Compliance Data (1979 - 1995). American Journal of Industrial Medicine; 34 547-558.

Marquart H, Van Drooge H, Groenewold M, Van Hemmen J. (2001) Assessing Reasonable Worst-Case Full Shift Exposure. Applied Occupational and Environmental Hygiene; 16 210-217.

Melville R, Lippmann M. (2001) Influence of data elements in OSHA air sampling database on occupational exposure levels. Applied Occupational and Environmental Hygiene; 16 884-899.

Mendeloff J. (1984) A New Strategy for Estimating Occupational Exposures to Toxic Substances. Cincinnati, OH: National Institute for Occupational Safety and Health (microfiche number NIOSH-00182240).

Money CD, Margary SA. (2002) Improved Use of Workplace Exposure Data in the Regulatory Risk Assessment of Chemicals within Europe. Annals of Occupational Hygiene; 46 279-285.

Morandi M, Maberti S. (2001) Aldehydes. In: Bingham E, Cohrssen B, Powell CH, Editors. Patty's Toxicology (Fifth edition). New York, NY: John Wiley & Sons, Inc. p. 963-1088.

Olsen E, Anglov T, Bach E, Breum NO, Christensen JM, Fallentin N, et al. (1997) Methods for the assessment of exposure to chemical agents at the workplace.

Luxembourg: European Commission, Employment & social affairs, Health and safety at work (EUR 15964 EN).

Olsen E, Jensen B. (1994) On the concept of the "normal" day - Quality control of occupational hygiene measurements. Applied Occupational and Environmental Hygiene Apr; 9 245-255.

Olsen E, Laursen B, Vinzents PS. (1991) Bias and Random Errors in Historical Data of Exposure to Organic Solvents. American Industrial Hygiene Association Journal; 52 204-211.

Oudiz J, Brown JW, Ayer HE, Samuels S. (1983) A Report on Silica Exposure Levels in United States Foundries. American Industrial Hygiene Association Journal; 44 374-376.

Perkins JL. (1997) Modern Industrial Hygiene vol.1 - Recognition and Evaluation of Chemical Agents. New York, NY: Van Nostrand Reinhold.

Pichard A. (2005) I NE R I S - Fiche de données toxicologiques et environnementales des substances chimiques : Formaldéhyde. Verneuil-en-Halatte: Institut National de l'environnement Industriel et des risques.

Page web: www.ineris.com/print.php?module=doc&action=getFile&id=160:

Pinheiro JC, Bates DM. (2000) Mixed-Effects Models in S and S-plus. New York: Springer-Verlag.

Rajan B, Alesbury R, Carton B, Gérin M, Litske H, Marquart E, et al. (1997) European Proposal for Core Information for the Storage and Exchange of Workplace Exposure Measurements on Chemical Agents. Applied Occupational and Environmental Hygiene; 12 31-39.

Rappaport SM, Selvin S, Roach SA. (1988) A Strategy for Assessing Exposures with Reference to Multiple Limits. Applied Industrial Hygiene; 3 310-315.

Répertoire toxicologique de la CSST. Formaldéhyde. *http://www.reptox.csst.qc.ca/RechercheProduits.asp*, page Web visitée le 15 Janvier 2006.

Répertoire toxicologique de la CSST. Formol. *http://www.reptox.csst.qc.ca/RechercheProduits.asp*, page Web visitée le 15 Janvier 2006.

Siemiatycki J. (1984) An Epidemiological Approach to Discovering Occupational Carcinogens by Obtaining Better Information on Occupational Exposures. In: Recent Advances in Occupational Health. 143-157.

Stamm R. (2000) MEGA-Database: One Million Data Since 1972. Applied Occupational and Environmental Hygiene; 16 159-163.

Stewart PA, Lees PS, Correa A, Breysse PA, Gail M, Graubard BI. (2003) Evaluation of Three Retrospective Exposure Assessment Methods. Annals of Occupational Hygiene; 47 399-411.

Stewart PA, Lees PS, Francis M. (1996) Quantification of historical exposures in occupational cohort studies. Scandinavian Journal of Work, Environment and Health; 22 405-414.

Stewart PA, Rice C. (1990) A source of Exposure Data for Occupational Epidemiology Studies. Applied Occupational and Environmental Hygiene; 5 359-363.

Stewart PA, Stenzel M. (1999) Data Needs for Occupational Epidemiologic Studies. Journal of Environmental Monitoring; 1 75N-82N.

Stewart PA, Stewart WF. (1994) Occupational case-control studies: II. Recommendations for exposure assessment. American Journal of Industrial Medicine; 26 313-326.

Stewart PA, Stewart WF, Siemiatycki J, Heineman EF, Dosemeci M. (1998) Questionnaires for collecting detailed occupational information for community-based case control studies. American Industrial Hygiene Association Journal; 59 39-44.

Symanski E, Chan W, Chang C-C. (2001) Mixed-Effects Models for the Evaluation of Long-term Trends in Exposure Levels with an Example from the Nickel Industry. Annals of Occupational Hygiene; 45 71-81.

Symanski E, Kupper LL, Hertz-Picciotto I, Rappaport SM. (1998a) Comprehensive evaluation of long-term trends in occupational exposure: Part 2. Predictive models for declining exposures. Occupational and Environmental Medicine; 55 310-316.

Symanski E, Kupper LL, Rappaport SM. (1998b) Comprehensive evaluation of long-term trends in occupational exposure: Part 1. Description of the database. Occupational and Environmental Medicine; 55 300-309.

Symanski E, Maberti S, Chan W. (2006) A Meta-Analytic Apporach for Characterizing the Within-Worker and Between-Worker Sources of Variation in Occupational Exposure. Annals of Occupational Hygiene Advance Access (March 2, 2006).

Symanski E, Rappaport SM. (1994) An investigation of the dependence of exposure variability on the interval between measurements. Annals of Occupational Hygiene; 38 361-372.

Teschke K, Marion SA, Vaughan TL, Morgan MS, Camp J. (1999) Exposure to Wood Dust in U.S. Industries and Occupation, 1979 to 1997. American Journal of Industrial Medicine; 35 581-589.

Tielemans E, Marquart H, de Cock J, Groenewold M, Van Hemmen J. (2002) A proposal for Evaluation of Exposure Data. Annals of Occupational Hygiene; 46 287-297.

U.S. Environmental Protection Agency. (1991) Formaldehyde risk assessment update. Washington, D.C.: U.S. EPA, Office of toxic substances.

Ulfvarson U. (1983) Limitations to the Use of Employee Exposure Data on Air Contaminants in Epidemiologic Studies. International Archives of Occupational and Environmental Health; 52 285-300.

Valiante D, Richards TB, Kinsley KB. (1992) Silicosis Surveillance in New Jersey: Targeting Worplace using Occupational Disease and Exposure Data. American Journal of Industrial Medicine; 21 517-526.

Van Tongeren M, Burstyn I, Kromhout D, Gardiner KG. (2006) Are Variance Components of Exposure Heterogeneous Between Time Periods and Factories in the European Carbon Black Industry? Annals of Occupational Hygiene; 50 55-64.

Vincent R, Jeandel B. (2001) COLCHIC-Occupational Exposure to Chemical Agents database: Current Content and Development Perspectives. Applied Occupational and Environmental Hygiene; 16 115-121.

Vincent R, Jeandel B. (2002) Exposition professionnelle au plomb : Analyse des résultats archivés dans la base de données COLCHIC. Cahiers de notes documentaires - Hygiène et sécurité du travail; 63-72.

Vinzents PS, Carton B, Fjeldstad P, Rajan B, Stamm R. (1995) Comparison of Exposure Mesurements Stored In European Databases on Occupational Air Pollutants and Definitions of Core Information. Applied Occupational and Environmental Hygiene; 10 351-354.

Walker JF. (1975) Formaldehyde. Huntington, NY: Robert E. Krieger Publishing Company.

Werner M, Attfield MD. (2000) Effect of Different Grouping Strategies in Developing Estimates of Personnal Exposures: Specificity Versus Precision. Applied Occupational and Environmental Hygiene; 15 21-25.

Wibowo A. (2003) Formaldehyde. Stockholm: National Institute for Working Life, Nordic coucil of Ministers, The Nordic Expert Group for Criteria Documentation on Health Risks from Chemicals and the Dutch Expert Committee on Occupational Standards.