Master of Public Health

Master de Santé Publique

Using administrative data to study social inequalities in health: sources of biases, and application to work injuries in France

Grisela PRENGA

Class and year of the Master:
EPH 2023-2025
Advanced Epidemiology and
Biostatistics

Location of the practicum:

INED, Aubervilliers, France

Professional advisor:

Emilie Counil, INED

Academic advisor:

Emilie Counil, INED

Acknowledgements

List of acronyms

APE: Principal Activity Code

AT-MP: Work Accidents and Occupational Diseases

BTS: All Employees Database

CASD: Secure Data Access Center

CI: Confidence Interval

CNAM: National Health Insurance Fund

CNAM-TS: Work and Health Branch of CNAM

CRAM: Regional Health Insurance Fund

DADS: Annual Declarations of Social Data

DARES: Directorate for Research, Studies, and Statistics

EU: European Union

ILO: International Labour Organization

INED: French Institute for Demographic Studies

INSEE: National Institute of Statistics and Economic Studies

IPW: Inverse Probability Weighting

ISCO: International Standard Classification of Occupations

NAF: French Classification of Economic Activities

OECD: Organisation for Economic Co-operation and Development

OHS: Occupational Health and Safety

PCS: Professions and Socio-professional Categories

RECORD: Reporting of studies using Observational Routinely-collected health Data

RGPD: General Data Protection Regulation

RWD: Real World Data

SAS: Statistical Analysis System

SIRET: Business Establishment Identifier

SUMER: Medical Surveillance of Occupational Risk Exposures

SIRx: Standardized Injury Rate Ratio (weighted)

List of tables and figures

Table of Contents

Abstract

Background: Social inequalities in occupational injury risk remain a major concern in public health, yet in France, few studies have quantitatively examined disparities between managers and non-managers. This thesis addresses that gap using administrative data from the private sector.

Objectives: First, this thesis aims to quantify the degree of inequality in work injury risk between managers and non-managers among private sector employees in France, across industries, sexes, and age groups, using the SIRx approach. Second, it examines how the choice of time-at-risk measure, person-years vs. person-hours, affects the magnitude and ranking of these inequalities. Lastly, it describes some dimensions of administrative data quality that are important to clarify the remaining blind spots in knowledge.

Methods: The analysis is based on administrative data sources and uses standardized injury rate ratios (SIRx), stratified by sex, sector, and age group. Two denominators of exposure, person-years and person-hours are compared to evaluate their impact on inequality measurement.

Results: The findings from this study contribute both substantively and methodologically. Substantively, they offer a high-resolution map of where and for whom the burden of occupational injury falls most heavily in the French labor market. Non-managers are at significantly higher risk of severe injury across all strata, with rate ratios reaching up to 14 among men and 6 among women. Inequalities are especially high among younger workers, and greater in male-dominated, physically demanding industries. Methodologically, they demonstrate how analytic choices, such as the exposure denominator, affect the interpretation of health inequalities and highlight the value of integrating hours-based exposure metrics in occupational health surveillance.

Conclusion: By combining rigorous statistical methods with detailed stratification and sensitivity to exposure time, this study contributes to a deeper understanding of how workplace injury risk is socially patterned and how it can be more accurately and equitably measured. The analysis is also informed by a broader methodological and theoretical framework on the use of administrative data in health inequality research. In conclusion, the findings reveal stark and persistent inequalities in work-related injury risks between managers and non-managers in France. These disparities are structured by gender, sector, and life stage, and are magnified when considering the actual time workers spend on the job.

Resumé

Contexte : Les inégalités sociales face au risque d'accidents du travail demeurent une préoccupation majeure en santé publique. Pourtant, en France, peu d'études ont examiné de manière quantitative les écarts entre cadres et non-cadres. Ce mémoire comble cette lacune en s'appuyant sur des données administratives issues du secteur privé.

Objectifs : Premièrement, ce travail vise à quantifier le degré d'inégalité du risque d'accidents du travail entre cadres et non-cadres dans le secteur privé en France, en tenant compte des secteurs d'activité, du sexe et des groupes d'âge, à l'aide de l'indicateur SIRx. Deuxièmement, il examine comment le choix de la mesure du temps d'exposition au risque, années-personnes vs heures travaillées, influence l'ampleur et le classement de ces inégalités. Enfin, il décrit certaines dimensions de la qualité des données administratives nécessaires pour éclairer les angles morts persistants dans les connaissances.

Méthodes : L'analyse repose sur des sources de données administratives et utilise des ratios standardisés d'incidence des accidents (SIRx), stratifiés selon le sexe, le secteur d'activité et le groupe d'âge. Deux dénominateurs d'exposition, les années-personnes et les heures personnes, sont comparés pour évaluer leur impact sur la mesure des inégalités.

Résultats : Les résultats de cette étude apportent une contribution à la fois empirique et méthodologique. Sur le plan empirique, ils offrent une cartographie fine des secteurs et des populations les plus exposés aux accidents du travail dans le marché du travail français. Les non-cadres présentent un risque significativement plus élevé de blessure grave dans l'ensemble des catégories analysées, avec des ratios pouvant atteindre 14 chez les hommes et 6 chez les femmes. Les inégalités sont particulièrement marquées chez les jeunes travailleurs et dans les secteurs physiquement exigeants à prédominance masculine. Sur le plan méthodologique, ils montrent que des choix analytiques tels que le dénominateur d'exposition influencent l'interprétation des inégalités de santé.

Conclusion : En combinant des méthodes statistiques rigoureuses, une stratification détaillée et une prise en compte du temps d'exposition, ce travail permet de mieux comprendre la structuration sociale du risque d'accidents du travail et propose des outils pour le mesurer de façon plus équitable. L'analyse s'appuie également sur un cadre théorique et méthodologique plus large relatif à l'usage des données administratives dans l'étude des inégalités sociales de santé. En conclusion, les résultats mettent en évidence des inégalités marquées et persistantes entre cadres et non-cadres, structurées par le sexe, le secteur d'activité et le moment de la vie, et amplifiées lorsqu'on considère le temps effectivement passé au travail.

1. Introduction

Understanding how social position translates into unequal health risks is a central question in public health research. The workplace is a key arena where these risks manifest, particularly through exposure to injury, illness, and long-term disability.

1.1 Current trends in occupational injuries

Work accidents, encompassing both fatal and serious non-fatal injuries, remain a major global public health and economic challenge (Takala et al., 2024). According to the International Labour Organization (ILO), nearly 3 million workers die each year due to work-related accidents and diseases, a figure that has increased by over 5% since 2015 (International Labour Organization, 2023). Of these, approximately 330,000 deaths are directly caused by workplace accidents, while the remaining fatalities stem from occupational diseases such as circulatory disorders, cancers, and respiratory illnesses. Non-fatal injuries are even more widespread: ILO estimates that 395 million workers worldwide sustain non-fatal work injuries annually, many of which result in long-term disability or significant time off work. Europe has made notable progress in reducing workplace fatalities and serious injuries[1], but the numbers remain substantial. In their latest report, the European Union (EU28) reported 3,286 fatal

	Non-fatal accidents at work involving at least 4 calendar days of absence from work			Fatal accidents at work
	Total	Men	Women	Total
EU	2 973 646	1 969 779	1 003 046	3 286
Belgium	61 164	42 703	18 460	45
Bulgaria	2 044	1 333	711	83
Czechia	36 029	24 182	11 830	88
Denmark	110 668	41 923	68 429	41
Germany	791 319	582 822	208 211	397
Estonia	5 301	3 727	1 574	15
Ireland	20 404	13 024	7 264	25
Greece	4 824	3 430	1 394	25
Spain	497 832	340 327	157 504	411
France	622 538	381 528	241 009	775
Croatia	10 068	6 307	3 756	48
Italy	330 131	209 908	120 223	469
Cyprus	1 326	992	334	9
Latvia	2 319	1 482	837	29
Lithuania	4 699	2 923	1 697	32
Luxembourg	6 447	4 972	1 475	12
Hungary	25 289	16 030	9 259	70
Malta	1 564	1 244	320	15
Netherlands	84 831	54 543	30 288	25
Austria	55 152	42 942	12 210	109
Poland	66 397	40 591	25 806	180
Portugal	125 607	87 708	37 899	141
Romania	3 173	2 185	988	127
Slovenia	16 023	9 227	6 796	17
Slovakia	7 925	5 078	2 847	31
Finland	35 743	23 748	11 995	27
Sweden	44 829	24 901	19 928	40
Iceland	1 328	897	430	1
Norway	10 854	6 392	4 462	31
Switzerland	94 739	72 688	22 051	58

Figure 1 *Number of non-fatal and fatal accidents at work, 2022 (people)*
Source : Eurostat

[1] Non-fatal accidents reported in the framework of ESAW are accidents that imply at least four full calendar days of absence from work (serious accidents)

accidents at work in 2022 (Figure 1) (Eurostat, 2024)[2]. Non-fatal accidents numbered nearly 3 million cases in the EU that year (Eurostat, 2024). The sectors most affected are construction, manufacturing, and transport, reflecting similar patterns worldwide (Figure **2**). Accidents are not evenly distributed across the continent. Countries like France, Portugal, and Spain report some of the highest rates of both fatal and non-fatal accidents, with incidence rates exceeding 2,700 non-fatal accidents per 100,000 employed persons (Eurostat, 2024). France stands out in Europe for its persistently high rates of severe work accidents. In 2022, France recorded 738 recognized fatal work accidents, averaging about two deaths per day, nearly double the EU average fatality rate (Eurostat, 2024). This figure excludes additional fatalities from commuting accidents and occupational diseases, which further increase the total burden.

Non-fatal accidents are also frequent. In 2022, France reported a standardized incidence rate of 2,733 non-fatal accidents per 100,000 employed persons, among the highest in the EU (Eurostat, 2024). The construction, manufacturing, and transport sectors are consistently the most hazardous.

Industry sector	Non-fatal accidents (%)
Construction	12,2%
Transportation and storage	8,8%
Manufacturing	18,0%
Agriculture, forestry and fishing	3,9%
Distributive trades	11,7%
Administrative and support service activities	8,8%
Public administration and defence; compulsory social security	5,1%
Accommodation and food service activities	5,0%
Human health and social work activities	15,8%
Water supply; sewerage, waste management and remediation activities	1,6%
Professional, scientific and technical activities	1,3%
Mining and quarrying	0,3%
Information and communication	0,5%
Other service activities	1,1%
Education	2,7%
Arts, entertainment and recreation	1,6%
Financial and insurance activities	0,3%
Electricity, gas, steam and air conditioning supply	0,2%
Real estate activities	0,4%
Activities of extraterritorial organisations and bodies	0,0%
Activities of households as employers	0,1%

Figure 2 *Non-fatal accidents at work by industry sector, 2022*
Source : Eurostat

---

[2] Eurostat is the statistical office of the European Union. https://ec.europa.eu/eurostat/web/main/home

1.2 Social and occupational disparities in work injuries

Growing evidence points to enduring and sometimes widening within-countries disparities in occupational health outcomes, with lower-status workers facing consistently higher exposure to physical risks and a heavier burden of injury. Among the most visible expressions of this inequality is the difference in work injury rates between managers and non-managers, a divide that captures both hierarchical position and, indirectly, other structural determinants of health such as education, autonomy, and job stability (Havet and Penot, 2022; Niedhammer et al., 2018; Richardson et al., 2023). Temporary workers, and more broadly precarious workers, have also been shown as particularly vulnerable, facing double the risk of fatal accidents compared to permanent employees, often due to less comprehensive safety training and more precarious working conditions (Kreshpaj et al., 2022).

However, despite the recognized importance of occupational class and employment relationships, few studies have comprehensively examined this relationship using national, population-level data. Even fewer have explored how the magnitude of this inequality varies by sector, sex, or age, or how different measures of exposure (e.g., person-hours vs. person-years) may affect estimates of relative risk. The task of measuring disparities between and within countries is actually made difficult by the type of routinely available data, which have limitations.

1.3 Under-reporting and disparities in reporting systems

While a large part of the afore-mentioned disparities is likely due to differences in labor market structure and in enforcement of OHS regulations, heterogeneity of reporting systems across countries is also at stake. This is particularly related to differences in legal definitions underlying the reporting of work accidents. For instance, in France, the legal definition of a work accident is broad, covering any physical or psychological harm linked to work. Accidents during working hours and at the workplace are presumed work-related unless proven otherwise by the employer (Inan, 2022). Such expansive definitions differ from narrower ones in other countries, complicating international comparisons and highlighting the need to consider legal contexts in reporting disparities. What's more, while the French reporting and insurance system is mandatory for all private sector workers (Inan, 2022), information on public sector and non-salaried workers is much sparser. Even for private sector salaried workers, relying on reported and compensated injuries alone to measure the burden of work injuries has some limitations. Under-reporting has in particular been pointed and described across EU countries (EUROGIP, 2023) in particular in France, all the more so for milder injuries that are likely under-counted.

Under-reporting of occupational injuries disproportionately affects lower-status and precariously employed workers, masking the true burden of workplace harm. In Norway, fatal work-related injuries were under-reported by up to 55%, with official statistics missing many deaths in high-risk, low-visibility sectors such as agriculture, transport, and informal employment (Gjertsen et al., 2022). Similarly, a Swedish study found that non-fatal injuries were significantly more under-reported among precarious workers, 22.6% compared to 15.0% for those in stable employment, highlighting a clear gradient of invisibility tied to job security (Kreshpaj et al., 2022). These findings underscore how precarious employment conditions and systemic data gaps contribute to the neglect of vulnerable workers in occupational health surveillance and policy-making.

These observations are important to highlight in a context where administrative datasets, such as injury registries and employment records, are increasingly used for epidemiological studies due to their scale, detail, and longitudinal coverage (Connelly et al., 2016; Gavrielov-Yusim and Friger, 2014). In France, sources like the National Health Insurance claims and Annual Declaration of Social Data (DADS) provide information on both injury events and employment characteristics, including occupational code, sector, and demographic variables. These data enable large-scale, detailed analyses that are often infeasible in survey-based research. Yet they also pose important challenges, including under-reporting, potential misclassification, data linkage error, and lack of contextual variables, factors that must be critically addressed in any administrative data study (Hashimoto et al., 2014; Katz and Freier, 2022; Shaw et al., 2022; Van Walraven and Austin, 2012).

1.4 Study rationale and research strategy

Our study intends to address some of the knowledge gaps about work injury disparities in France and the usefulness of administrative data by using the National Health Insurance claims covering private sector employees.

To do so, we develop a research strategy based on the SMRw, a well-established epidemiological measure (Cole and Hernan, 2008; Sato and Matsuyama, 2003; Brookhart et al., 2013) that enables standardized comparisons between groups with different level of occupational exposure, such as job type and sector, by adjusting for confounding factors like age and sex. Following the work by Richardson et al., (2023), we seek to quantify disparities in work injury rates between managers and non-managers, by estimating Standardized Injury Rate Ratios (SIRx) stratified by industry sector, sex, and age group. An innovation of our study is the use of two distinct time-at-risk denominators: person-years and person-hours. While person-years are a common exposure metric in occupational health surveillance (Bena et al., 2013; Richardson et al., 2023), they may underestimate risk for groups who spend disproportionately more time at work. This is especially relevant in lower-status occupations,

where workers may face longer hours, more frequent shift work, and greater pressure to accept overtime (Wagstaff and Sigstad Lie, 2011). By comparing SIRx values derived from person-years and person-hours, this study hopes to provide a more nuanced picture of inequality in work injury risk, one that considers not just who is exposed, but how much exposure they accumulate. The underlying hypothesis is that disparities in injury rates will be consistently large across groups, but more pronounced when actual working time is accounted for. This is based on the assumption that non-managers, especially in manual or semi-skilled roles, not only have more hazardous working conditions but also accumulate greater time in those conditions (Havet and Penot, 2022; Niedhammer et al., 2018; Richardson et al., 2023). Conversely, women are known to work part-time more often than men do (Eurostat, 2023) and this could also bias estimates towards the null.

1.5 Aim and objectives

The overarching aim of this thesis is to examine and quantify social inequalities in work-related injuries between managers and non-managers working in the private sector in France, using large-scale administrative data. Within this broader aim, the study pursues the following specific objectives:

1. To describe the demographic and occupational composition of the French workforce by sex, age group, and managerial status, and to assess the distribution of reported work-related injuries across these groups.

2. To estimate the Standardized Injury Rate Ratios (SIRx) comparing the injury risk of non-managers to that of managers, stratified by:
   - Sex (men and women),
   - Age group (five age categories),
   - Industry sector (NAF classification).

3. To compare SIRx values using two different exposure measures, person-years and person-hours, in order to assess how the choice of time-at-risk denominator affects the magnitude and interpretation of injury risk inequalities.

4. To identify sectors and demographic groups with the greatest disparities, highlighting combinations of sex, sector, and age group where non-managers face the highest relative injury risks.

5. To contribute to methodological discussions on the use of administrative data in occupational health research, by demonstrating how large-scale, routinely collected data can be used to analyze health inequalities in a stratified, and policy-relevant way, without omitting their remaining blind spots.

2. Methods

2.1 Literature search

A crucial part of this study involves reviewing existing conceptual frameworks and methodological tools designed to assess challenges and sources of bias in epidemiological research using administrative data. This includes frameworks addressing selection bias, misclassification, confounding, and data linkage errors. The insights gained from this review will be directly applied to analyze and interpret the French administrative data available for this study.

In conducting this comprehensive literature review, a multi-faceted methodological approach was employed to ensure a thorough examination of the extant literature. The process started with an in-depth analysis of seminal works provided by the research supervisor, which served as the foundational corpus for this investigation. These initial texts provided crucial context and direction, informing the subsequent stages of the review process.

Building upon this foundation, a snowballing technique was utilized to expand the literature base. This involved examination of the bibliographies of the initial papers, a method referred to as backward/forward citation chaining (Wohlin, 2014). This approach facilitated the identification of additional relevant sources and work that had significantly contributed to the field's theoretical and empirical landscape.

To further enhance the comprehensiveness of the review, contemporary bibliometric tools were employed. Specifically, ResearchRabbit[3] was utilized to visualize and explore related research by mapping the articles as network and timeline (Figure 9 in the Appendix).

Screening of the articles was done primarily by title and abstract in accordance with the relevance to the core themes of the investigation.

The subsequent phase involved analysis and synthesis of the accumulated literature. Studies were systematically categorized using a thematic framework including "administrative data and biases", "occupational injuries", and "statistical methods" allowing for the identification of patterns, trends, and relationships. This organizational approach facilitated the understanding of key concepts and methodologies within the field.

Synthesis matrices were constructed to facilitate the comparison of key ideas, methodologies, and findings from multiple sources.

---

[3] ResearchRabbit is an innovative citation-based literature mapping tool available online :
https://www.researchrabbit.ai/

2.2 Data sources and selection of study population

This project is based on two administrative data sources that we combine in order to derive SIRx. On the one hand, data on work accidents and occupational diseases (AT-MP) provided by the National Health Insurance Fund for Salaried Workers (CNAM-TS), and on the other hand, Annual declaration of Social Data (DADS) which is an administrative formality that any company employing staff must complete. Both of the used data sources in this study are accessible after approval by the statistical confidentiality committee (Comité du Secret Statistique) through a Secure Data Access Center (CASD - Secure Data Access Centre).

2.2.1 Numerator (number of injuries)

The CNAM-TS data made available to researcher over the 2003-2018 period is structured into two main tables: Claims and Section (AT-MP-CNAM, CASD, DARES).

The Claims table compiles information on workplace accidents and occupational illnesses, as well as the characteristics of the victims. Each accident or illness that contributed to the calculation of each year is recorded on a single line, identified by an anonymized number for the victim and an accident number. An accident is recorded in the year in which it is compensated, which may differ from the year in which it occurred.

This table includes additional data, such as the sociodemographic characteristics of the victims, occupational code (ISCO), as well as the SIRET (Establishment Directory Identification System) of the establishment where the accident occurred. The data also includes details on compensation paid to victims, such as financial compensation, daily allowances, number of lost days, as well as annuities or lump sum compensation.

The Section table contains information on establishments, subject to the general social security scheme that employed the victims listed in the Claims table. The data includes elements such as the SIRET, the code of the Regional Health Insurance Fund (CRAM) to which the victim belongs, the Social Security risk code, the postal code, and the economic activity code (APE). The link between the Section and Claims tables is made by combining the SIRET variables and the CRAM number. This method is essential to ensure precise correspondence between accident information and the establishments concerned, thus allowing a detailed analysis of occupational risks within different companies and sectors.

2.2.2 Denominator (person-time at risk)

The Base Tous Salariés based on DADS (BTS) is a required declaration that all French employers with salaried staff must submit (BTS-Salariés, INSEE, CASD). It compiles employment and wage information used by various government agencies to track workforce and

salary trends. Each year, employers must report, for every company site, a set of standardized details about their employees. This includes data such as age, sex, region, company, job title, employment period, number of paid hours, type of contract, and total compensation. This dataset focuses on the employee's main job held during a year. Our choice was driven by a first round of analysis using census data to estimate person-year at risk. The conclusion was that of a substantial mismatch between the numerator and denominator that likely altered the validity of estimates (Traore, 2024).

To construct the analytic sample, a series of exclusion criteria were applied to ensure consistency between the two administrative sources (numerator and denominator), following work by DARES (DARES, 2009), as well as relevance to the study objectives. First, individuals younger than the official minimum working age of 15 were excluded, aligning with the legal threshold for entering apprenticeships in France. The dataset was then refined by removing interns and individuals receiving employment-related benefits, as these groups are not considered part of the salaried workforce. Further exclusions were made for individuals whose occupations fell outside the general social security scheme, which ensures comparability in labor protections and injury reporting systems. Records with missing data on key demographic or employment variables were also removed to maintain analytic reliability, and documented as such to feed the data quality assessment (Figure 5 and 6 in the Appendix). This approach is consistent with standard epidemiological practice, as the inclusion of incomplete records could introduce misclassification and bias into the estimation of injury rates and group comparisons (Rothman et al., 2008).

Finally, after joining the aggregated datasets of the numerator and denominator, a final exclusion of individuals employed in certain sectors was made: O (public administration), U (extraterritorial organizations), T (household employees), and A (agriculture). These steps produced a robust, sector-consistent sample of salaried private-sector workers suitable for comparative analysis of occupational injury risk using the two types of denominators: one in person-years and the other in person-hours.

2.3 Analytical variables

2.3.1   Outcome variable

An accidental injury can take many different forms. Traditionally, the following three forms are distinguished in France (DARES, 2009):

1. Minor accidental injury: As it does not result in work stoppage or medical care, this type of accident does not entail any compensation.

2. Accidental injury with medical care but without work stoppage: More serious than minor accidental injury, this type of accident required medical care, in the form of only in-kind benefits (coverage, without advance payment, of medical and pharmacy expenses), but does not result in any work stoppage or payment of cash benefits.

3. "Work-related accidents with time off": This refers to accidents that were compensated by the payment of a cash benefit (daily allowance, lump-sum compensation, annuity for the victim or any beneficiaries). It primarily includes accidents that resulted in at least one day of sick leave (in addition to the day the accident occurred), but also fatal accidents. This type of accident has significant consequences, as it leads to work stoppage and in the worst cases, permanent impairment of the victim's physical integrity.

The primary outcome was the annual count of severe lost-time work accidents, defined as incidents resulting in financial compensation and at least 10 days of work absence (DARES, 2009). This threshold is widely used in official statistics and research to distinguish more serious injuries, particularly in reports from the French National Health Insurance (CNAM) and the Ministry of Labour.

2.3.2   Exposure, adjustment and stratification variables

The main exposure was managerial status, classified as a binary variable (Manager vs. Non-manager), included standardizing occupational codes and resolving discrepancies in sector classifications, using ISCO88 occupational codes , supplemented by professional qualification codes where necessary (Table 7 in the Appendix). Data from each year were filtered, harmonized and merged by managerial status, sex, age group (15 to 60+ years old), industry sector (NAF 2008) and year. Categorical variables were factorized, and the data were aggregated to yield total accidents and person-time for each stratum defined by managerial status, sex, age group, industry sector, and year.

2.4 Statistical analysis

To address confounding by age and sex, we applied inverse probability weighting (IPW)(Cole and Hernan, 2008). A logistic regression model was fit to estimate the probability of being a manager as a function of age group, sex, and their interaction, weighted by person-years, as well as person-hours. The predicted probabilities from this model were used to construct stabilized weights: non-managers were assigned a weight of *1*, while managers received a weight equal to the odds of being a non-manager given their covariates, as the equation below show.

$$Weights= \frac{P(X=1|S)}{1- P(X=1|S)}$$

P(X = 1 | S) represents the conditional probability of being a non-manager, as a function of the covariates *S.*

This approach creates a pseudo-population in which the distribution of measured confounders is independent of exposure, thereby facilitating causal interpretation of the calculated ratios. Weighted Poisson regression models were further used to estimate standardized injury ratios and quantify disparities in occupational health outcomes while accounting for overdispersion in count data (Richardson et al., 2023). For each analysis stratum, the number of work accidents was modeled as a function of managerial status, with the log of person-time included as an offset. The regression was weighted by the stabilized IPW weights described above. The exponentiated coefficient for managerial status yields the SIRx, interpreted as the rate ratio of work accidents in non-managers relative to managers, standardized for age and sex.

$$SIRX = \frac{[Y|X = 1]}{[Y|X = 0] \times \left(\frac{P(X = 1|S)}{1 - P(X = 1|S)} \right)} = e^{\beta}$$

[Y│X=1] is the number of work accidents in the non-managers group.
[Y│X= 0] is the number of work accidents in the managers group.

To quantify the uncertainty of the SIRx estimates, we employed non-parametric bootstrap resampling (Mokhtar et al., 2023). For each stratum, the entire estimation procedure, including IPW calculation and Poisson regression, was repeated on 500 bootstrap samples, generating an empirical distribution of SIRx estimates. Confidence intervals were calculated on the log scale and exponentiated to ensure valid, positive bounds. Percentile-based intervals were also computed as a robustness check.

SIRxs were estimated overall and stratified by industry sector (NAF code), sex, age group, and year. For each stratum, the same IPW and Poisson regression approach was applied, looping over all combinations as appropriate.

2.5 Software and reproducibility

Analyses were conducted mainly in R and partly in SAS, utilizing packages including *dplyr, DataExplorer, boot, purrr, and tidyr.* A part of the code, which is a function used to calculate the weights and SIRxs is shown in Figure 7 in the Appendix.

2.6 Ethical Considerations

This project was approved by the Comité du secret statistique under the number scss-4228-3. All analyses were conducted using de-identified data accessed via the Secure Data Access Center (CASD). Data use agreements were strictly adhered to, and no individual-level identifiers were accessible to the research team, in accordance with European and French data protection regulations (RGPD).

3 Results

3.1 Results from the analysis of French administrative data

This section presents the findings from the analysis of work-related injury inequalities between managers and non-managers in France, based on large-scale administrative data sources. The results are organized into three parts: first, we report on the output of the selection process using flow charts; second, we show descriptive statistics of the distribution of employment and injury burden by sex, age, and occupational status; and third, we provide analytical results using the Standardized Injury Rate Ratio (SIRx), calculated both with person-years and person-hours as denominators. Particular attention is paid to how inequalities vary by sector, sex, and age, and how these patterns are affected by different measures of time at risk.

3.1.1 Study population and data completeness

While the original dataset covers the years 2003 to 2018, our analysis is deliberately limited to the period from 2009 to 2012. This decision stems from both data quality concerns and methodological constraints.

One key reason is the lack of granularity of an important variable in the denominator before 2009, which contained information on the work contract of employees. This variable would be necessary to perform a robust filtering process in line with DARES. Furthermore, starting from 2013, around 20% of injury records in the CNAM-TS database lacked the variable needed to determine managerial status. This missing information would have introduced substantial bias into our analysis, particularly by underestimating the number of cases in certain sectors.

As such, we chose to focus exclusively on 2009–2012. It is important to point out that in this time-series, around 70% of the total observations are included. To define our analytic sample, we applied a filtering process aimed at retaining only private-sector employees affiliated with the general social security regime. Although the original dataset includes all salaried workers reported in the DADS-INSEE files, it also contains individuals from special regimes (e.g.,

agricultural workers, civil servants, employees of large public enterprises), who were excluded, following conventions developed by DARES (DARES, 2009) (Figure 3 , Figure 4). In practice this meant excluding workers affiliated with agricultural regime (MSA), civil servants without a formal employment contract, and employees of large enterprises with special social protection (e.g., EDF, SNCF, La Poste), unless they were clearly identified as contract-based staff. Additional restrictions were applied to certain sectors such as maritime transport, mining, and energy, where only workers with valid contract and collective agreement fields were retained. This approach helped isolate the target population, employees covered by the general insurance scheme and ensured consistency in exposure and reporting standards across the study sample. Moreover, we focused on severe work injuries, which make up around 63.3% of all reported work injuries during our study period. Finally, after applying all the necessary filters, the missing data in key interest variables was minimal, ranging from 0.01% to 0.73% (Figure 5 and Figure 6 in the Appendix), so we decided to exclude the missing data.

80 19 351
Total of work accidents and occupational diseases 2009–2012.

2 662 352
Keep only work accidents.

1 694 450
Keep only severe accidents.

1 694 187
Filter: Exclude less than 15 years old.

1 665 545
Filter: Exclude interns and individuals whose specified professions are not in the general social security scheme data.

1 636 890
Filter: Unknown sectors, ISCO88 and PCS.

1 635 643
Exclude missing data.

1 603 478
Exclude sectors O, U, T, A
(after joining with the DADS)

Figure 3 Sample selection: Work injuries 2009-2012

126 405 786
Total of person-years 2009–2012.

126 231 732
Individuals aged more than 15 years old (the official age of starting an apprenticeship)

124 661 855
Filter: Remove interns and people with employment benefits

90 233 709
Filter: Exclude individuals whose specified professions are not in the general social security scheme data

89 733 076
Remove missing data

84 690 398
Exclude sectors O, U, T, A
(after joining with the CNAM-TS)

Figure 4 Sample sepection: Person-years 2009-2012

3.1.2 Employment structure and crude injury disparities

The descriptive statistics confirm a pronounced occupational hierarchy in the dataset. As shown in Table 1 among men, non-managers represent almost 81.0% of total person-years, while managers account for just 19.0%. For women, the divide is even more striking: non-managers account for 86.4% of person-years, and managers only 13.6%. The same divide is seen when we are looking to person-hours in Table 2, with small variations. These differences in occupational distribution are foundational to understanding later disparities in injury burden. Age composition varies substantially between occupational categories. Non-managerial roles are strongly skewed toward younger workers. Among male non-managers, 32.2% are aged 15–29 years old, compared to just 14.7% of male managers. A similar pattern holds for women, where 31.2% of female non-managers fall into the 15–29 years old age group, whereas this figure drops to 19.3% among female managers. Conversely, older workers (50–59 years old and 60+ years old) are overrepresented in managerial positions, suggesting longer career trajectories, possible role accommodation, or survival effects in less physically demanding jobs.

The injury data mirrors these structural imbalances. Male non-managers sustained 1,087,760 injuries, compared to 20,080 among male managers, translating to rates of 2,861.4 vs. 224.7 per 100,000 person-years. For women, non-managers experienced 482,100 injuries, compared to 13,538 among managers, with rates of 1,478.0 vs. 264.4 per 100,000 person-years. These differences suggest that, even before adjustment, occupational status is associated with an order-of-magnitude difference in injury risk. Combining sexes, non-managers accounted for 97.9% of all injuries, reaffirming their disproportionately high burden. Even after weighting, the disparities persist, as men who are non-managers have a much higher injury rate compared to their managerial counterparts, with rates of 2,861.4 and 213.2, respectively. Among women, as well, the weighted injury rate for non-managers is 1,478.0, while for managers it is 269.0 per 100,000 person years.

The same patterns manifest when calculating inury rates per 1,000,000 person-hours.

Table 1 Distribution of population at risk and severe work injuries by managerial status, sex, and age in France 2009-2012 (person-years).

	Original data			Weighted data		
	Managers	**Non-Managers**	**Total**	**Managers**	**Non-Managers**	**Total**
Men	8 936 096	38 014 950	46 951 046	38 014 949	38 014 950	76 029 899
15-29 years old	14.70%	32.18%	28.85%	32.18%	32.18%	32.18%
30-39 years old	28.55%	25.18%	25.82%	25.18%	25.18%	25.18%
40-49 years old	29.16%	23.30%	24.42%	23.30%	23.30%	23.30%
50-59 years old	22.21%	16.85%	17.87%	16.85%	16.85%	16.85%
60 + years old	5.38%	2.49%	3.04%	2.49%	2.49%	2.49%
	19.03%	80.97%				
Total no of accidents	20 080	1 087 760	1 107 840	81 056	1 087 760	1 168 816
Rate per 100.000 person years	224.71	2 861.40	2 359.56	213.22	2 861.40	1 537.31
Women	**Managers**	**Non-Managers**	**Total**	**Managers**	**Non-Managers**	**Total**
	5 119 816	32 619 536	37 739 352	32 619 534	32 619 536	65 239 070
15-29 years old	19.29%	31.16%	29.55%	31.16%	31.16%	31.16%
30-39 years old	32.09%	23.51%	24.67%	23.51%	23.51%	23.51%
40-49 years old	26.90%	24.21%	24.57%	24.21%	24.21%	24.21%
50-59 years old	18.05%	18.13%	18.12%	18.13%	18.13%	18.13%
60 + years old	3.67%	2.99%	3.08%	2.99%	2.99%	2.99%
	13.57%	86.43%				
Total no of accidents	13 538	482 100	495 638	8 7685	482 100	569 785
Rate per 100.000 person years	264.42	1 477.95	1 313.32	269.00	1 477.95	873.38
Total no of accidents both sexes	33 618	1 569 860	1 603 478	168 740	1 569 860	1 738 600
Rate per 100.000 person years	239.17	2 222.51	1 893.34	238.89	2 222.51	1 230.70

Table 2 Distribution of population at risk and severe work injuries by managerial status, sex, and age in France 2009-2012 (person-hours).

	Original data			Weighted data		
	Managers	**Non-Managers**	**Total**	**Managers**	**Non-Managers**	**Total**
Men	14 308 456 773	54 413 933 716	68 722 390 489	54 413 936 578	54 413 933 716	108 827 870 294
15-29 years old	12.49%	26.55%	23.63%	26.55%	26.55%	26.55%
30-39 years old	28.74%	26.65%	27.09%	26.65%	26.65%	26.65%
40-49 years old	30.76%	26.01%	27.00%	26.01%	26.01%	26.01%
50-59 years old	23.32%	18.75%	19.70%	18.75%	18.75%	18.75%
60 + years old	4.68%	2.03%	2.58%	2.03%	2.03%	2.03%
	20.82%	79.18%		50.00%	50.00%	
Total no of accidents	20 080	1 087 760	11 07 840	74 768	1 087 760	1 162 528
Rate per 1.000.000 person hours	1.40	19.99	16.12	1.37	19.99	10.68
	Managers	**Non-Managers**	**Total**	**Managers**	**Non-Managers**	**Total**
Women	7 525 221 418	41 029 655 780	48 554 877 198	32 619 534	32 619 536	65 239 070
15-29 years old	16.71%	25.44%	24.09%	25.44%	25.44%	25.44%
30-39 years old	32.24%	24.90%	26.04%	24.90%	24.90%	24.90%
40-49 years old	28.55%	26.88%	27.14%	26.88%	26.88%	26.88%
50-59 years old	19.24%	20.14%	20.00%	20.14%	20.14%	20.14%
60 + years old	3.26%	2.63%	2.73%	2.63%	2.63%	2.63%
	15.50%	84.50%		50.00%	50.00%	
Total no of accidents	13 538	482 100	495 638	75 959	482 100	558 059
Rate per 1.000.000 person hours	1.80	11.75	10.21	1.85	11.75	6.80
Total no of accidents both sexes	33 618	1 569 860	1 603 478	150 727	1 569 860	1 720 587
Rate per 1.000.000 person hours	1.54	16.45	13.67	1.58	16.45	9.01

3.1.3 SIRx by years

To account for differences in age distribution and time exposure, SIRx values were calculated using both person-years and person-hours as denominators. This comparison reveals the extent to which inequalities are shaped not only by job type but by actual work time.

Over the study period from 2009 to 2012, SIRx estimates based on person-years showed moderate variation but no statistically significant trend, ranging from 8.42 (95% CI: 5.97–10.76) in 2009 to 9.49 (95% CI: 6.86–12.09) in 2012 (Table 3). The overlapping confidence intervals indicate that the relative injury risk between managers and non-managers remained consistent and statistically stable. SIRx values based on person-hours revealed slightly larger disparities, as expected.

Table 3 Standardized injury ratios of severe work injuries between managers and non-managers across years 2009-2012 (by person-years and person-hours).

	Person-years		Person-hours	
Year	SIRx	CI 95%	SIRx	CI 95%
2009	8.42	5.97,10.76	9.48	6.90,11.82
2010	9.40	6.80,11.90	10.48	7.84,12.87
2011	9.97	7.07,12.82	11.14	8.27,13.81
2012	9.49	6.86,12.09	10.66	7.92,13.20

3.1.4 SIRx by sector and sex

Consistently with yearly estimates, the results found over the 2009-2012 period suggest that non-managers were over-exposed to the risk of severe work injuries by a factor of 9.2 compared to managers, with an interval going from 6.59 to 11.87 (Table 4). When using person-hours, the factor went up to 10.3 (CI95%: 7.8 –12.7) (Table 5).

Among men, the SIRx was 13.42 (95% CI: 12.05–14.80) using person-years. When calculated with person-hours, the SIRx increased to 14.55 (95% CI: 13.40 - 15.67). For women, the person-year-based SIRx was 5.50 (95% CI: 5.06 - 5.96), and increased to 6.35 (95% CI: 5.81 - 6.90) when using person-hours. These results show that male non-managers experience over almost 14 times the injury risk of managers, and female non-managers between five to over six times, depending on the measure of exposure.

Sectoral analysis provides further resolution of where and for whom disparities are most acute. Among men, SIRx values were highest in sectors traditionally associated with physical labor. In manufacturing, the SIRx based on person-years was 17.09 (95% CI: 15.20-19.12), and increased slightly using person-hours. In construction, the SIRx increased from 8.97 (95% CI:

7.16 - 11.19) to 9.52 (95% CI: 7.81-11.55) when switching from person-years to hours. Similarly, transport and logistics, electricity and gas, water supply, and education suggest high SIRx in men.

Among women, the highest SIRx values were also found in manufacturing, where the estimate increased from 8.99 (95% CI: 7.37-11.09) to 10 (95% CI: 8.22-12.21) with person-hours. Other sectors manifesting high SIRx include transportation, water supply and waste management, and education.

Across most of the sectors, there is a noticeable difference between managers and non-managers across women and men holding manager and non-manager positions, with SIRx around double for men. Sectors that do not show many significant changes between managerial roles are accommodation and food service activities, human health and social work activities, as well as the sector of arts, entertainment and recreation.

Across nearly all sectors, the use of person-hours as a denominator consistently elevated SIRx estimates, particularly in male-dominated, high-risk sectors. To give a more thorough view of the severe work injuries as well as rates across industry sectors, we included *Table 8* and *Table 9* in the Appendix.

Table 4 Standardized injury ratios of severe work injuries between managers and non-managers across industry sectors and sex using person-years 2009-2012

NAF sector	SIRx	CI 95%	SIRx Men	CI 95%	SIRx Women	CI 95%
All sectors	9.22	6.59 , 11.87	13.42	12.05 , 14.80	5.50	5.06 , 5.96
Mining and quarrying	7.05	4.26 , 11.63	7.76	4.66 , 14,39	1.28	0.44 , 4.45
Manufacturing	14.63	11.00 , 18.87	17.09	15.20 , 19.12	8.99	7.37 , 11.09
Electricity, gas, air conditioning supply	14.11	10.21 , 19.13	15.76	10.32 , 24.99	4.66	2.13 , 11.56
Water supply, waste management	12.38	7.60 , 19.57	12.47	10.40 , 14.87	11.08	6.67 , 19.41
Construction	8.81	4.90 , 15.62	8.97	7.16 , 11.19	3.66	2.49 , 5.33
Wholesale and retail trade, repair of vehicles	5.55	4.33 , 7.04	7.87	7.33 , 8.46	3.66	3.38 , 3.98
Transportation and storage	10.85	8.65 , 13.41	10.83	9.57 , 12.16	10.95	9.56 , 12.52
Accommodation and food service activities	1.12	0.77 , 1.90	1.27	1.07 , 1.58	0.99	0.73 , 1.47
Information and communication	7.18	4.93 , 10.20	9.05	6.97 , 11.75	5.08	4.18 , 6.24
Financial and insurance activities	5.61	4.09 , 7.98	9.98	7.84 , 12.70	4.38	3.63 , 5.27
Real estate activities	4.05	2.96 , 5.56	5.12	4.26 , 6.10	3.22	2.61 , 4.02
Professional, scientific and technical activities	6.19	4.57 , 8.42	9.47	8.03 , 11.22	4.03	3.43 , 4.78
Administrative and support service activities	8.34	6.09 , 11.23	9.48	7.99 , 11.27	6.43	5.11 , 8.13
Education	13.6	8.13 , 22.57	19.06	13.54 , 26.00	10.14	8.21 , 12.55
Human health and social work activities	1.53	1.26 , 1.94	2.75	2.33 , 3.23	1.39	1.23 , 1.56
Arts, entertainment and recreation	2.84	1.19 , 7.61	4.02	2.53 , 6.56	1.33	0.91 , 2.09
Other service activities	3.10	2.37 , 4.40	4.82	4.20 , 5.57	2.57	2.16 , 3.12

Table 5 Standardized injury of severe work injuries ratios between managers and non-managers across industry sectors and sex using person-hours 2009-2012

NAF sector	SIRx	CI 95%	SIRx Men	CI 95%	SIRx Women	CI 95%
All sectors	10.28	7.82 , 12.73	14.55	13.40 , 15.67	6.35	5.81, 6.90
Mining and quarrying	6.94	4.38 , 11.07	7.59	4.54 , 14.15	1.34	0.45 , 4.72
Manufacturing	15.45	11.93 , 19.42	17.66	15.72 , 19.73	10.00	8.22 , 12.21
Electricity, gas, air conditioning supply	14.21	10.55 , 18.84	15.79	10.36 , 24.94	4.84	2.21 , 12.04
Water supply, waste management	12.30	7.62 , 19.25	12.34	10.17 , 14.84	11.62	7.01 , 20.26
Construction	9.37	5.51 , 15.82	9.52	7.81 , 11.55	4.16	2.83 , 6.07
Wholesale and retail trade, repair	6.32	5.04 , 7.8	8.51	7.75 , 9.31	4.35	3.95 , 4.78
Transportation and storage	11.15	9.26 , 13.41	10.95	9.66 , 12.31	12.26	10.59 , 14.15
Accommodation and food service activities	1.60	1.09 , 2.62	1.67	1.36 , 2.09	1.53	1.1 , 2.25
Information and communication	7.64	5.58 , 10.23	9.67	7.63 , 12.21	5.38	4.55 , 6.41
Financial and insurance activities	6.06	4.61 , 8.31	10.75	8.73 , 13.2	4.73	4.02 , 5.55
Real estate activities	4.49	3.42 , 5.89	5.68	4.84 , 6.59	3.56	2.96 , 4.32
Professional, scientific and technical activities	6.67	5.11 , 8.7	10.01	8.89 , 11.32	4.39	3.77 , 5.17
Administrative and support service activities	11.45	8.89 , 14.5	12.91	11.46 , 14.48	8.95	7.41 , 10.83
Education	16.62	10.29 , 26.65	22.17	16.09 , 29.58	12.78	10.73 , 15.23
Human health and social work activities	1.53	1.18 , 2.04	2.62	2.29 , 2.99	1.39	1.2 , 1.61
Arts, entertainment and recreation	2.28	0.87 , 6.75	3.06	1.77 , 5.43	1.15	0.74 , 1.94
Other service activities	3.44	2.56 , 5.03	5.41	4.61 , 6.38	2.85	2.34 , 3.53

3.1.5 SIRx by age group

The age-stratified SIRx analysis (Table 6) reveals further insights. Using person-years, the highest SIRx was observed among 30–39-year-olds, at 11.14 (95% CI: 6.89–15.22), followed by 40–49-year-olds at 9.74 (95% CI: 7.49 –11.93). The youngest group (15–29) had a lower but still substantial SIRx of 9.51 (95% CI: 5.67 –13.47). Inequality declined with age: SIRx dropped to 7.20 (95% CI: 5.32 – 9.12) in the 50–59 group and fell further to 5.97 (95% CI: 3.39 –8.64) in the 60+ category. When calculated using person-hours, the SIRx values in each age group increased modestly.

Table 6 Standardized injury ratios severe work injuries between managers and non-managers across age groups 2009-2012 (by person-years and person-hours).

Age group	Person-years		Person-hours	
	SIRx	CI 95%	SIRx	CI 95%
15-29 years old	9.51	5.67 , 13.47	11.41	6.68 , 16.31
30-39 years old	11.14	6.89 , 15.22	12.08	7.88 , 16.00
40-49 years old	9.74	7.49 , 11.93	10.57	8.31 , 12.72
50-59 years old	7.20	5.32 , 9.12	7.83	5.74 , 10.07
60 + years old	5.97	3.39 , 8.64	7.08	3.69 , 10.53

3.2 Literature summary on administrative data use in research

In the following section, we report on the literature highlighting key advantages and pitfalls identified while using administrative data for research. Meanwhile, we also reflect on our research strategy, hence somehow anticipating on the discussion of the strengths and limitations of our own work.

Administrative data have emerged as a critical resource in public health research, particularly in contexts where population-wide coverage and longitudinal follow-up are essential. However, these datasets, while comprehensive in scale, were not designed for epidemiological research and come with well-documented biases and limitations.

A key advantage of administrative data is their scale and representativeness. As noted by Hashimoto et al., (2014), Katz and Freier, (2022), and Shaw et al., (2022) administrative datasets often cover entire national populations, enabling research on rare outcomes and small subgroups that would be impractical to study using survey data alone. For instance, (Fulton-Kehoe et al., 2007) have shown that administrative measures of work disability, such as compensated time off, are strongly correlated with self-reported pain and functional limitations, suggesting that administrative sources can serve as robust proxies when survey data are limited by low response rates or recall bias. This reinforces the rationale for using administrative data in occupational injury research, particularly when comparing risk across narrowly defined strata.

This aligns with the recommendations of Jorm, (2015), who emphasizes the potential of routinely collected data to support system-level health evaluations when paired with appropriate statistical approaches such as the use of propensity scores, weighting, time-series analysis, or multilevel modeling.

Nonetheless, the benefits of administrative data must be weighed against their potential for introducing bias. In their study, Katz and Freier, (2022) discuss the concept of access bias in administrative health records, noting that disadvantaged populations often have reduced interactions with healthcare systems, either due to financial barriers, lack of sick leave, or cultural mistrust. In occupational settings, this can mean that injuries occurring in lower-status workers may never enter administrative systems if those individuals do not seek formal care or file claims. As such, our decision to focus on accidents resulting in ≥10 compensated days off work helps to reduce the influence of this bias by prioritizing more severe, and thus more likely to be reported, cases.

Another layer of complexity arises from the original purpose of administrative data. As Van Walraven and Austin (2012) caution, these datasets are typically created for operational or billing purposes, not research. This can result in coding inconsistencies, omitted variables, and ambiguous classifications that affect both exposure and outcome definitions. For example, occupational categories may not adequately reflect actual managerial responsibilities or

exposure conditions, leading to potential misclassification in manager vs. non-manager comparisons in our study.

Gavrielov-Yusim and Friger, (2014) highlight the information bias that arises when key covariates are missing or poorly recorded. In the context of our study, the absence of data on employment contracts, or even the main exposure variable (managerial status) can introduce bias in the findings. The limitation of time series included is a practical solution but makes the results more limited.

From a methodological perspective, multiple authors stress the importance of transparent reporting and bias mitigation strategies. The RECORD statement developed by Benchimol et al. (2015) proposes specific guidelines for reporting observational studies using routinely collected health data. These include clear definitions of variables, discussion of potential biases, and explanation of data cleaning and linkage procedures. Although our study is not directly governed by RECORD (Figure 8 in the Appendix), its emphasis on transparency and stratification aligns with these best practices, especially in the use of standardized injury rates and inverse probability weighting (IPW) to control for differences in demographic structure.

The methodological implications are further discussed in the paper by Togo and Yonemoto, (2022), who explore how real-world data (RWD), including administrative datasets, require rigorous adjustment techniques due to the absence of randomization. They advocate for techniques such as regression modelling, confounder adjustment, and bias correction, all of which are employed in this thesis through weighted Poisson models and bootstrapped confidence intervals. Their findings underscore the value of standardization methods, as used in this project to assess of occupational injury inequalities robustly.

Administrative data's utility in identifying population-level patterns is also emphasized by Doidge and Harron (2019), who warn, however, against assuming that large data size eliminates bias. On the contrary, they argue that selection bias, misclassification, and exposure misalignment can persist, and even be amplified, without deliberate methodological countermeasures. Their recommendation to triangulate administrative findings with qualitative knowledge or survey-based benchmarks is particularly relevant in interpreting differences in injury reporting behavior by job status.

Moreover, many authors (Penner and Dodge, 2019; Van Der Wel et al., 2019; Taylor et al., 2021) discuss the burden of ethical and operational dimensions of administrative data use with experiences from USA, United Kingdom, and Nordic countries. They note that even when data are comprehensive, issues such as informed consent, legal access, and governance remain. These concerns are evident in the limitations of this thesis, which focuses exclusively on the private sector due to the unavailability to researchers of comparable administrative data for public-sector or informal workers. This exclusion is both a practical constraint and a reminder that administrative data, while vast, do not fully capture the working population.

Finally, several studies echo the need for critical appraisal of research based on administrative sources. Hashimoto et al. (2014) refer to such databases as a potential "goldmine," but warn that without proper scrutiny, they may lead researchers into "goose chases." Misinterpretation of causality, inadequate control for confounding, or failure to recognize missingness patterns can distort conclusions. This reinforces the importance of methodological transparency in our study, especially in defining the limitations of SIRx measures and the assumptions behind exposure metrics.

In sum, the literature demonstrates that administrative data are an essential tool for studying health inequalities but must be used critically. Their large size, longitudinal structure, and sectoral granularity make them well-suited to analyzing occupational disparities. However, their original design, reporting gaps, and structural exclusions introduce non-trivial biases, particularly in relation to under-reporting among disadvantaged workers. Our study adopts several mitigation strategies inspired by this literature. These approaches aim to extract meaningful insights while acknowledging the inherent limitations of administrative data in occupational health research.

4 Discussion

4.1 Summary of key results

This study, based on comprehensive French administrative data, provides a detailed and stratified analysis of occupational injury risk disparities between managers and non-managers across economic sectors, and by sex and age. Using the Standardized Injury Rate Ratio (SIRx) as a comparative indicator, the results highlight persistent and substantial inequalities in work-related injury risks, with patterns that vary meaningfully depending on sectoral context, gender, and stage in the working life course.

4.2 Interpretation

Across all industry sectors, non-managers consistently face significantly higher rates of reported work injuries than managers, even after adjustment for age. When exposure was measured using person-years, the overall SIRx was 13.4 for men (95% CI: 12.1–14.8) and 5.5 for women (95% CI: 5.1–6.0). However, when using person-hours, a more refined measure of actual time at risk, these disparities were even greater, with SIRx rising to 14.6 for men (95% CI: 13.4–15.7) and 6.4 for women (95% CI: 5.8–6.9). These differences reflect the fact that

non-managers not only face more hazardous job conditions but also tend to work longer hours in high-risk settings, especially in male-dominated sectors (Richardson et al., 2023; Biswas et al., 2022; Niedhammer et al., 2018; Piha et al., 2013). For instance, in manufacturing, male non-managers had an SIRx of 17.1 (95% CI: 15.2–19.1) when using person-years, which increased slightly to 17.7 using hours. Among women, the SIRx in manufacturing was 9 (95% CI: 7.4–11.1), increasing to 10 based on person-hours. In education, male SIRx was 19.1 (95% CI: 13.5–26) compared to 10.1 (95% CI: 8.2–12.6) among women. These patterns point to important differences in both job assignments and actual work exposure. Male non-managers are more likely to be employed in support, maintenance, or logistics roles within these sectors, roles with substantial physical demands and less control over working conditions, while managers often occupy supervisory or office-based positions with considerably lower injury risk (Richardson et al., 2023; Niedhammer et al., 2018).

In contrast, sectors such as human health and social work, accommodation and food services, and art industry showed smaller SIRx values. For example, in health and social work, the SIRx was 1.5, with minimal change when switching from person-years to hours. These narrower gaps may reflect more homogeneous job tasks between managers and non-managers in these sectors, especially among women, or broader cultural factors around injury reporting and role structure.

Age-stratified results provide further insights. The disparity between managers and non-managers was highest among workers aged 30–39 years old, with an SIRx of 11.1 (95% CI: 6.9–15.2),

The age-based trends held across both sexes and most sectors. When recalculated using person-hours, SIRx values increased across all age groups, most notably in the youngest cohort, where SIRx rose from 9.5 to 11.4. This suggests that younger non-managers not only experience higher injury rates but also likely accumulate more time in high-risk roles, either due to job design, shift patterns, or lack of autonomy to refuse unsafe tasks. These differences reflect a well-documented progression through the labor market, where managerial status tends to increase with age and experience (Elvira et al., 2023; OECD, 2020).

The observed age gradient supports the idea that younger non-managerial workers face elevated injury risks due to their limited experience, increased physical demands, or assignment to more hazardous tasks (Smith and Mustard, 2007). The narrowing of SIRx at older ages could reflect a "healthy worker effect," whereby only those with better health or safer roles remain employed, particularly among non-managers (Wilcosky and Wing, 1987). Older managers may also benefit from more stable and protected job environments, while non-managers who experience injuries are more likely to exit the labor force prematurely or transition to less physically demanding roles (Taylor, 2017). These findings reinforce the

importance of integrating a life-course perspective into analyses of occupational health inequalities.

The consistent increase in SIRx when using person-hours rather than person-years further strengthens the argument that time-based exposure is critical to understanding inequality. Relying solely on person-years can slightly understate the disparity, especially in sectors with extended working hours or where overtime is common among non-managerial staff. The difference is most notable among men and in manual sectors, suggesting that both structural exposure and time burden compound risk (Biswas et al., 2022).

4.3 Research and Public Health implications

From a policy perspective, our findings have several implications. First, the clear disparities by sector, sex, and age underscore the need for targeted, context-specific prevention strategies. For example, injury prevention in sectors like manufacturing and education should focus on non-managers, especially men, in physically intensive or support roles. This might include investment in ergonomic improvements, better safety training, and structural oversight in high-risk environments. Among younger workers, tailored interventions such as onboarding programs, mentorship, and job hazard training could help mitigate early-career vulnerability.

Second, the study highlights the value of using person-hours when estimating time at risk. Policies, occupational health monitoring, and injury surveillance systems should aim to incorporate time-based denominators to better capture the realities of work exposure. This is particularly important in settings with high variability in contract type, shift length, or seasonal labor, where person-years fail to adequately reflect the heterogeneity of time worked.

This study also demonstrates the strength of administrative data for public health research. The working dataset had a very limited amount of missing information on key variables, as such it enabled a robust, population-level analysis with granular stratification by sex, age, and economic sector. As highlighted in recent methodological literature (Connelly et al., 2016; Shaw et al., 2022; Jorm, 2015), administrative data provide coverage that is unmatched by surveys and can detect inequality patterns invisible in smaller or self-reported samples. The capacity to compare outcomes using both person-years and person-hours is an additional advantage, revealing hidden layers of occupational risk.

4.4 Study limitations

An important limitation of this study is that we could only report occupational injuries claimed by salaried workers in the private sector who represent around 71% of the actual workforce in

France during the years included in our study. While this ensures greater consistency and reliability in our findings, it also means that any developments or shifts occurring after 2012 are not captured in the scope of our study. Public sector employees as well as the self-employed and agricultural workers were excluded because their insurance scheme for compensating work injuries and occupational disease is managed under a different scheme for which data are not accessible to researchers. Yet they are likely to carry a large burden of injuries as well that remains invisible in our work. Another limitation in terms of population coverage, as indeed in administrative data more broadly, is the systematic exclusion of informal or unregistered workers. The dataset captures only formally declared, salaried employment relationships within the private sector. As such, it omits individuals in precarious, undocumented, or informal labor arrangements, who may be disproportionately represented in certain industries such as construction, domestic services, or seasonal agricultural work. These workers often operate outside the protective reach of labor law and are likely to experience higher injury risks, lower reporting rates, and limited access to compensation systems (Kreshpaj et al., 2022). Their absence from the data may result in an underestimation of inequality, particularly in sectors with high informal employment. Future efforts to link administrative data with survey or qualitative sources could help to make these invisible workers statistically visible.

The focus on private sectors employees also prevents from drawing conclusions about activities that are largely dominated by the public sector in France, such as education and healthcare.

Additionally, the classification of individuals as "managers" versus "non-managers" deserves critical attention. While the distinction is based on standardized occupational codes, it may not consistently reflect actual work content or hierarchical status across sectors. For instance, some so-called managers may have limited supervisory authority, while some non-managers may perform quasi-managerial functions without the corresponding job title. Furthermore, classification systems may differ across firms or evolve over time, introducing potential misclassification bias. This is particularly relevant in sectors like health, education, or services, where occupational boundaries are fluid and roles may overlap. Despite using a binary distinction that is analytically useful, it is important to recognize its limitations in capturing the complexity of work organization.

Finally, and perhaps most importantly, in addition to the lack of data on informal workers and workers out the private salaried world, and potential misclassification of job roles, it is important to stress that administrative records are collected primarily for regulatory and operational purposes, not for research. They often lack contextual information about specific job tasks, work intensity, and psychosocial factors. Furthermore, differential reporting of injuries by occupational status (Kreshpaj et al., 2022) may bias comparisons if, for example, managers

are less likely to report minor injuries. Under-reporting has been documented and one way to overcome this limitation may be to consider only ever more serious, disabling injuries that are less likely to be under-reported (Kreshpaj et al., 2022).

5 Conclusion

This study investigated social inequalities in severe work-related injuries in France, focusing on differences between managers and non-managers. Using rich administrative data from 2009 to 2012 and statistical methods like inverese probability weighting and Weighted Poisson regression to determine Standardized Injury Rate Ratio (SIRx), the study found that non-managers consistently experience substantially higher injury risks than managers, underscoring the enduring role of occupational status in shaping health outcomes at work. However, by providing results by industry, it made clear that the occupational division of risk may be even more unequal in specific sectors, while in others there seem to be a (relatively) more horizontal distribution.

A major contribution of this research is its use of administrative data, which enables large-scale, detailed analyses not possible with traditional surveys. Importantly, they also make it possible to compare different exposure metrics, such as person-years and person-hours, revealing that the magnitude of observed inequalities can depend on how time at risk is measured.

However, administrative data have limitations. They rarely cover the entire population of interest (here, any person involved in paid-work in France). They often lack contextual information about workplace practices, psychosocial conditions, and informal employment, factors that may further deepen disparities but remain invisible in register-based analysis. As such, the actual scale of injury inequality in the broader labor market may be even greater than reported here. Since this study focused on private sector employees, disparities within the public sector, agriculture and among self-employed workers remain undocumented.

Moreover, the binary classification of "manager" versus "non-manager," although analytically practical, may also obscure differences within occupational groups.

Despite these limitations, this study offers a compelling demonstration of how administrative data can be used not only to quantify occupational health inequalities. The comparison between person-years and person- hours strengthens the case for adopting time-sensitive exposure measures in occupational health research.

These findings have practical implications for occupational health surveillance and policy. They point to the need for targeted prevention strategies focusing on non-managerial roles in high-risk sectors, particularly among young and mid-career workers. Injury prevention efforts should include ergonomic improvements, safety oversight, and better onboarding and training in

hazardous environments. The results also support the case for further investment in administrative data infrastructure, including efforts to capture more nuanced information on job content, informal labor, and organizational context. A further step could be to contrast the risk between workers in standard (permanent) with those in non-standard (e.g. non-permanent) employment relations. And in the near future, linked data from the EDP-Santé could allow individual-level and longitudinal analysis of the mutual influence of career paths and workplace injuries in a lifecourse perspective.

References

AT-MP-CNAM : Workplace accidents - Occupational Injuries [WWW Document], n.d. . CASD. URL https://www.casd.eu/en/source/workplace-accidents-occupational-injuries/

Bena, A., Giraudo, M., Leombruni, R., Costa, G., 2013. Job tenure and work injuries: a multivariate analysis of the relation with previous experience and differences by age. BMC Public Health 13, 869. https://doi.org/10.1186/1471-2458-13-869

Benchimol, E.I., Smeeth, L., Guttmann, A., Harron, K., Moher, D., Petersen, I., Sørensen, H.T., Elm, E. von, Langan, S.M., Committee, R.W., 2015. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLOS Medicine 12, e1001885. https://doi.org/10.1371/journal.pmed.1001885

Biswas, A., Harbin, S., Irvin, E., Johnston, H., Begum, M., Tiong, M., Apedaile, D., Koehoorn, M., Smith, P., 2022. Differences between men and women in their risk of work injury and disability: A systematic review. American Journal of Industrial Medicine 65, 576–588. https://doi.org/10.1002/ajim.23364

Brookhart, M.A., Wyss, R., Layton, J.B., Stürmer, T., 2013. Propensity Score Methods for Confounding Control in Non-Experimental Research. Circ Cardiovasc Qual Outcomes 6, 604–611. https://doi.org/10.1161/CIRCOUTCOMES.113.000359

BTS-Salariés : Base Tous Salariés : fichiers Salariés [WWW Document], n.d. . CASD. URL https://www.casd.eu/source/base-tous-salaries-fichiers-salaries/

CASD - Secure Data Access Centre [WWW Document], n.d. URL https://www.casd.eu/en/.

Cole, S.R., Hernan, M.A., 2008. Constructing Inverse Probability Weights for Marginal Structural Models. American Journal of Epidemiology 168, 656–664. https://doi.org/10.1093/aje/kwn164

Comité du Secret Statistique [WWW Document], n.d. . Comité du Secret Statistique. URL https://www.comite-du-secret.fr/.

Connelly, R., Playford, C.J., Gayle, V., Dibben, C., 2016. The role of administrative data in the big data revolution in social science research. Social Science Research 59, 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015

DARES, 2009. Les indicateurs accidents du travail de la Dares - conception, champ et interprétation. [WWW Document]. URL https://dares.travail-emploi.gouv.fr/publications/les-indicateurs-accidents-du-travail-de-la-dares-conception-champ-et

Daubas-Letourneux, V., 2009. Accidents du travail : des blessés et des morts invisibles. Mouvements 58, 29–37. https://doi.org/10.3917/mouv.058.0029

Doidge, J.C., Harron, K.L., 2019. Reflections on modern methods: linkage error bias. Int J Epidemiol 48, 2050–2060. https://doi.org/10.1093/ije/dyz203

Elvira, M.M., Quintana-García, C., Villamor, I., 2023. Aging the gap: the compensation among men and women executives. Journal of Management & Organization 1–18. https://doi.org/10.1017/jmo.2023.11

EUROGIP, 2023. EUROGIP-2023 Under-reporting of accidents at work in Europe [WWW Document]. URL https://eurogip.fr/wp-content/uploads/2024/08/EUROGIP-2023-Under-reporting-of-accidents-at-work-in-Europe.pdf

Eurostat, 2024. Accidents at work statistics [WWW Document]. Eurostat - Statistics Explained. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics.

Eurostat [WWW Document], 2023. . Share of women working part-time higher than men. URL https://ec.europa.eu/eurostat/web/products-eurostat-news/w/edn-20230303-1.

Fulton-Kehoe, D., Gluck, J., Wu, R., Mootz, R., Wickizer, T.M., Franklin, G.M., 2007. Measuring Work Disability: What Can Administrative Data Tell Us About Patient Outcomes? [WWW Document]. Journal of Occupational and Environmental Medicine. URL https://www.jstor.org/stable/44997175

Gavrielov-Yusim, N., Friger, M., 2014. Use of administrative medical databases in population-based research. J Epidemiol Community Health 68, 283–287. https://doi.org/10.1136/jech-2013-202744

Gjertsen, F., Lund, J., Wergeland, E., 2022. Accuracy of fatal occupational injury registration in a high-income country: A comparison of two-source capture-recapture estimates with the number of cases identified in four register systems in Norway, 2000-2003. Glob Epidemiol 4, 100072. https://doi.org/10.1016/j.gloepi.2022.100072

Hashimoto, R.E., Brodt, E.D., Skelly, A.C., Dettori, J.R., 2014. Administrative database studies: goldmine or goose chase? Evid Based Spine Care J 5, 74–76. https://doi.org/10.1055/s-0034-1390027

Havet, N., Penot, A., 2022. Trends in exposures to physically demanding working conditions in France in 2003, 2010 and 2017. European Journal of Public Health 32, 73–79. https://doi.org/10.1093/eurpub/ckab195

Inan, C., 2022. Quels sont les salariés les plus touchés par les accidents du travail en 2019 ? | DARES [WWW Document]. URL https://dares.travail-emploi.gouv.fr/publication/quels-sont-les-salaries-les-plus-touches-par-les-accidents-du-travail-en-2019.

International Labour Organization, 2023 [WWW Document], 2023. URL
https://www.ilo.org/resource/news/nearly-3-million-people-die-work-related-accidents-and-diseases.

Jorm, L., 2015. Routinely collected data as a strategic resource for research: priorities for methods and workforce. Public Health Res Pract 25, e2541540. https://doi.org/10.17061/phrp2541540

Katz, A., Freier, A., 2022. Recognition of inherent biases in administrative data. Can Fam Physician 68, 933–934. https://doi.org/10.46747/cfp.6812933

Kreshpaj, B., Bodin, T., Wegman, D.H., Matilla-Santander, N., Burstrom, B., Kjellberg, K., Davis, L., Hemmingsson, T., Jonsson, J., Håkansta, C., Orellana, C., 2022. Under-reporting of non-fatal occupational injuries among precarious and non-precarious workers in Sweden. Occup Environ Med 79, 3–9. https://doi.org/10.1136/oemed-2021-107856

Mokhtar, S., Yusof, Z., Sapiri, H., 2023. Confidence Intervals by Bootstrapping Approach: A Significance Review. Malaysian Journal of Fundamental and Applied Sciences 19, 30–42. https://doi.org/10.11113/mjfas.v19n1.2660

NAF 2008 [WWW Document], n.d. URL https://www.insee.fr/fr/information/2120875

Niedhammer, I., Lesuffleur, T., Labarthe, G., Chastang, J.-F., 2018. Role of working conditions in the explanation of occupational inequalities in work injury: Findings from the national French SUMER survey. BMC Public Health 18. https://doi.org/10.1186/s12889-018-5254-7

OECD, 2020. Promoting an Age-Inclusive Workforce: Living, Learning and Earning Longer. OECD. https://doi.org/10.1787/59752153-en

Penner, Dodge, 2019. Using Administrative Data for Social Science and Policy. RSF: The Russell Sage Foundation Journal of the Social Sciences 5, 1. https://doi.org/10.7758/rsf.2019.5.3.01

Piha, K., Laaksonen, M., Martikainen, P., Rahkonen, O., Lahelma, E., 2013. Socio-economic and occupational determinants of work injury absence. The European Journal of Public Health 23, 693–698. https://doi.org/10.1093/eurpub/cks162

Richardson, D.B., Cole, S.R., Martin, A.T., McClure, E.S., Nocera, M., Cantrell, J., Ranapurwala, S.I., Marshall, S.W., 2023. Disparities in Fatal Occupational Injury Rates in North Carolina, 1978–2017: Comparing Nonmanagerial Employees to Managers. Epidemiology 34, 741–746. https://doi.org/10.1097/EDE.0000000000001632

Rothman, K.J., Greenland, S., Lash, T.L., 2008. Modern epidemiology, 3rd ed. ed. Wolters Kluwer health - Lippincott Williams & Wilkins, Philadelphia.

Sato, T., Matsuyama, Y., 2003. Marginal Structural Models as a Tool for Standardization [WWW Document]. Epidemiology. URL https://www.jstor.org/stable/3703427.

Shaw, R.J., Harron, K.L., Pescarini, J.M., Pinto Junior, E.P., Allik, M., Siroky, A.N., Campbell, D., Dundas, R., Ichihara, M.Y., Leyland, A.H., Barreto, M.L., Katikireddi, S.V., 2022. Biases arising from linked administrative data for epidemiological research: a conceptual framework from registration to analyses. Eur J Epidemiol 37, 1215–1224. https://doi.org/10.1007/s10654-022-00934-w

Smith, P.M., Mustard, C.A., 2007. How many employees receive safety training during their first year of a new job? Inj Prev 13, 37–41. https://doi.org/10.1136/ip.2006.013839

Takala, J., Hämäläinen, P., Sauni, R., Nygård, C.-H., Gagliardi, D., Neupane, S., 2024. Global-, regional- and country-level estimates of the work-related burden of diseases and accidents in 2019. Scand J Work Environ Health 50, 73–82. https://doi.org/10.5271/sjweh.4132

Taylor, J.A., Crowe, S., Espuny Pujol, F., Franklin, R.C., Feltbower, R.G., Norman, L.J., Doidge, J., Gould, D.W., Pagel, C., 2021. The road to hell is paved with good intentions: the experience of applying for national data for linkage and suggestions for improvement. BMJ Open 11, e047575. https://doi.org/10.1136/bmjopen-2020-047575

Taylor, R.N., 2017. Age and Experience as Determinants of Managerial Information Processing and Decision Making Performance. Academy of Management Journal. https://doi.org/10.5465/255626

Togo, K., Yonemoto, N., 2022. Real world data and data science in medical research: present and future. Jpn J Stat Data Sci 5, 769–781. https://doi.org/10.1007/s42081-022-00156-0

Traore, N., 2024. Inégalités socio-démographiques et professionnelles vis-à-vis des accidents du travail en France.

Van Der Wel, K.A., Östergren, O., Lundberg, O., Korhonen, K., Martikainen, P., Andersen, A.-M.N., Urhoj, S.K., 2019. A gold mine, but still no Klondike: Nordic register data in health inequalities research. Scand J Public Health 47, 618–630. https://doi.org/10.1177/1403494819858046

Van Walraven, C., Austin, P., 2012. Administrative database research has unique characteristics that can risk biased results. Journal of Clinical Epidemiology 65, 126–131. https://doi.org/10.1016/j.jclinepi.2011.08.002

Wagstaff, A.S., Sigstad Lie, J.-A., 2011. Shift and night work and long working hours--a systematic review of safety implications. Scand J Work Environ Health 37, 173–185. https://doi.org/10.5271/sjweh.3146

Wilcosky, T., Wing, S., 1987. The healthy worker effect. Selection of workers and work forces. Scand J Work Environ Health 13, 70–72. https://doi.org/10.5271/sjweh.2078

Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering, in: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. Presented at the EASE '14: 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, London England United Kingdom, pp. 1–10. https://doi.org/10.1145/2601248.2601268

Appendixes

Table 7 Correspondence table PCS-ISCO88, Source: DARES 2009

| PCS ESE Classification Level 1	ISCO-88 Classification
2 and 3 - Business managers, executives, and higher intellectual professions[4]	11 - Members of the executive and legislative bodies and senior public administration managers. 12 - Company directors. 13 - Executives and managers. 21 - Specialists in physical, mathematical, and technical sciences. 22 - Specialists in life and health sciences. 23 - Teaching specialists. 24 - Other specialists in intellectual and scientific professions.
4 - Associate professionals	31 - Associate professionals in physical and technical sciences. 32 - Associate professionals in life and health sciences. 33 - Associate teaching professions. 34 - Other associate professionals
5 - Employees	41 - Office workers. 42 - Receptionists, cashiers, counter clerks, and similar workers. 51 - Personnel in direct personal services and protection and security services. 52 - Models, salespeople, and demonstrators. 91 - Unskilled service and sales employees (if the qlf_cod variable takes the categories 2 - Manager, technician, supervisor, or 3 - Employee).
6 - Workers	61 - Skilled farmers and workers in market agriculture and fishing. 62 - Subsistence farmers and workers in agriculture and fishing. 71 - Craftsmen and workers in mining and construction trades. 72 - Craftsmen and workers in metallurgy, mechanical construction, and similar trades. 73 - Craftsmen and workers in precision mechanics, arts and crafts, printing, and similar trades. 74 - Other craftspeople and workers in artisanal trades. 81 - Operators of fixed and related installations and equipment. 82 - Machine operators and assembly workers. 83 - Drivers of vehicles and heavy lifting and handling equipment. 91 - Unskilled service and sales employees (if the qlf_cod variable does not take the categories 2 - Manager, technician, supervisor, or 3 - Employee). 92 - Laborers in agriculture, fishing, and related occupations. 93 - Laborers in mining, construction and public works, manufacturing industries, and transportation.

---

[4] PCS-ESE group 1 (Agricltural workers) was excluded from the analysis.

Figure 5 Missing data plots from the DADS, 2009-2012, exported from R, constructed with "plot_missing", from DataExplorer package
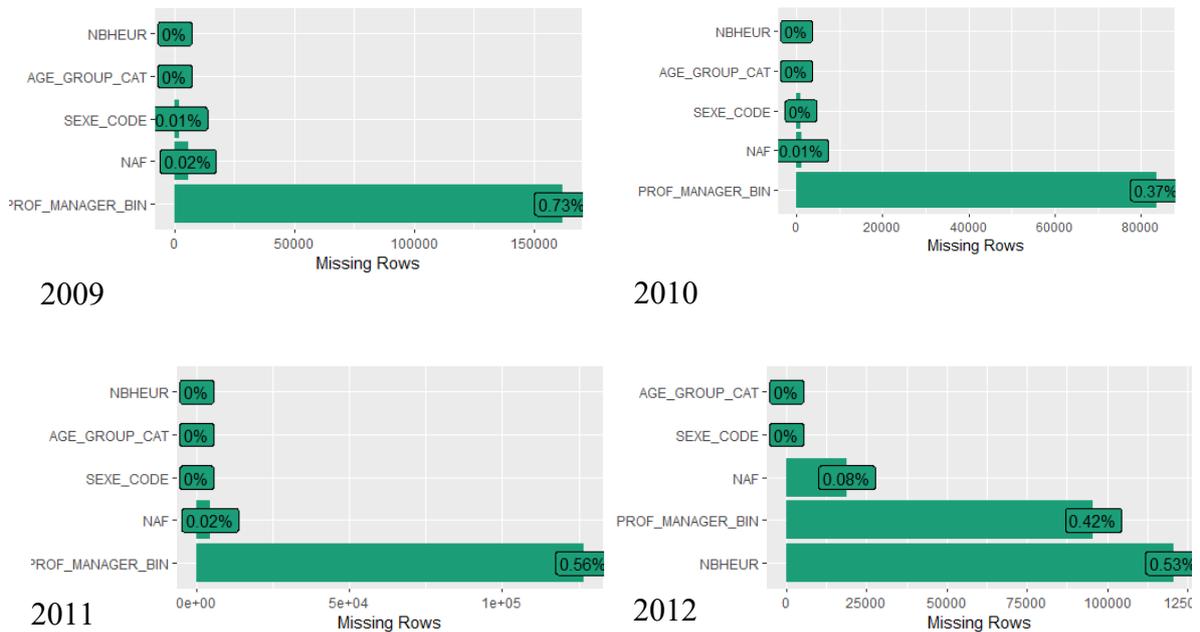
NBHEUR: Number of hours worked.
AGE_GROUP_CAT: Categorical variable of age
(1: 15-29 years old, 2: 30-39 years old, 3: 40-49 years old, 4: 50-59 years old, 5: 60 plus years old)
SEXE_CODE: Binary variable of sex (1: Male, 2: Female)
PROF_MANAGER_BIN: Binary variable of managerial status (0: Manager, 1: Non-manager)

Figure 6 Missing data plots from the CNAM-TS, 2009-2012, exported from R, constructed with "plot_missing", from DataExplorer package
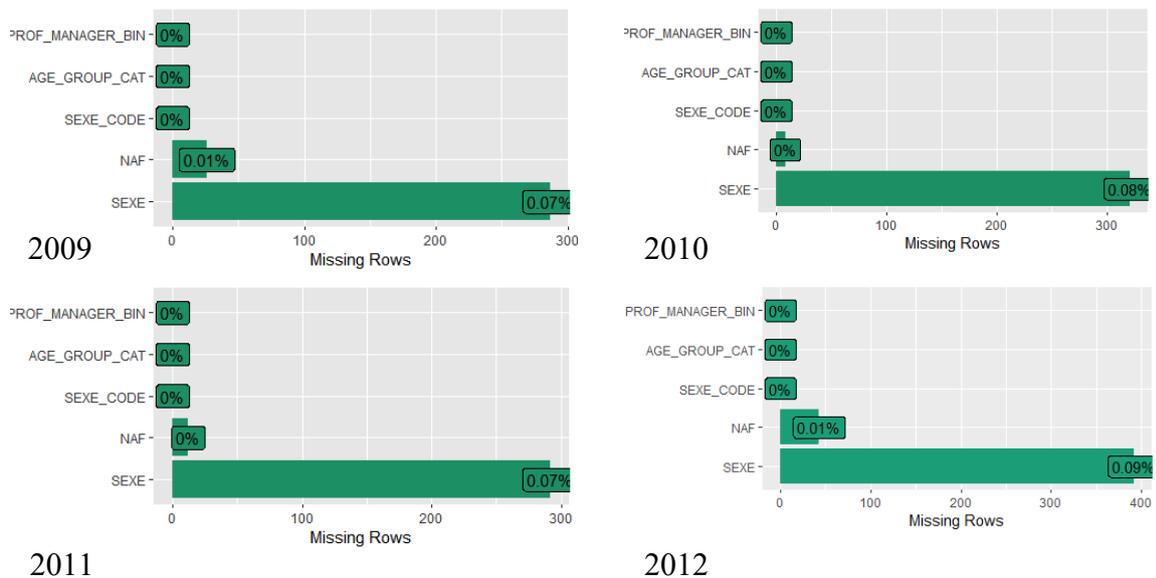
PROF_MANAGER_BIN: Binary variable of managerial status (0: Manager, 1: Non-manager)
AGE_GROUP_CAT: Categorical variable of age
(1: 15-29 years old, 2: 30-39 years old, 3: 40-49 years old, 4: 50-59 years old, 5: 60 plus years old)
NAF: Industry sectors
SEXE_CODE: Binary variable of sex (1: Male, 2: Female)
SEXE: Male, Female

Figure 7 R function estimating the SIRx comparing managers and non-managers, example adjusted for age.

```r
calcul_sirx <- function(data, indices) {
  d <- data[indices,]

  logistic_model <- glm(PROF_MANAGER_BIN ~ AGE_GROUP_CAT,
                        family = binomial (link = "logit"),
                        data = d,
                        weights = NBR_SALARIES)
  d$probability <- predict(logistic_model, type = "response")

  d <- d %>%
    mutate(logpy = log(NBR_SALARIES),
           w1 = ifelse(PROF_MANAGER_BIN == 1, 1, weights / (1 - weights)))

  poisson_model <- glm(NB_ACCIDENTS ~ PROF_MANAGER_BIN + offset(logpy),
                       family = poisson (link = "log"),
                       data = d,
                       weights = w1)
  sirx_estimate <- exp(coef(poisson_model))[2]
  return(sirx_estimate)
}
```
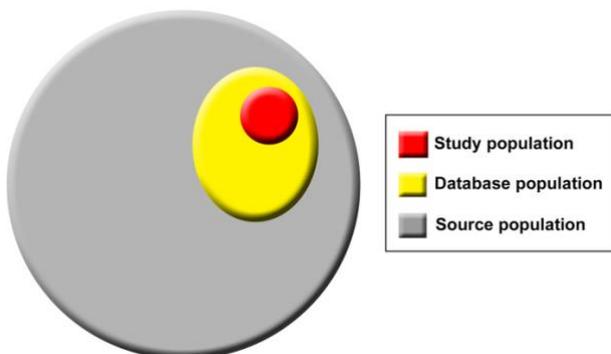
Figure 8 Guidelines for reporting observational studies based on routinely collected health data (RECORD*)*

| Population hierarchy in studies using routinely collected data sources	Checklist adapted from STROBE insists on:
Study population (red), Database population (yellow), Source population (grey)	• Selection of study population • Linkage process (with flow diagram) and assessment of linkage quality • Classification of exposures, outcomes, confounders • Data cleaning methods • Additional filters on data quality • Misclassification bias, unmeasured confounding, missing data, changing eligibility over time • Processes to access • Programming codes

Table 8 Standardized rates (per 100 000 p-y) and Ratio (SIRx) of severe work injuries among private sector employees in France, by industry, 2009-2012. (Weighted data)

NAF sector	Severe work injuries		Rate per 100,000 person years		SIRx	CI 95%
	Managers	Non-managers	Managers	Non-managers		
All sectors	167 840	1 569 860	237.62	2 222.51	9.22	6.59 , 11.87
Mining and quarrying	158	2 352	327.28	2 442.85	7.05	4.26 , 11.63
Manufacturing	14 289	250 387	149.53	2 256.47	14.63	11.00 , 18.87
Electricity, gas, air conditioning supply	137	1 365	75.66	1 126.26	14.11	10.21 , 19.13
Water supply, sewerage, waste management	897	21966	261.22	3 705.60	12.38	7.60 , 19.57
Construction	11 821	272 754	462.87	4 317.10	8.81	4.90 , 15.62
Wholesale and retail trade, repair of vehicles	33 164	253 695	351.82	2 007.55	5.55	4.33 , 7.04
Transportation and storage	5 017	168 422	278.66	3 748.69	10.85	8.65 , 13.41
Accommodation and food service activities	13 036	100 310	1 355.52	1 952.87	1.12	0.77 , 1.90
Information and communication	6 489	6 212	66.12	498.45	7.18	4.93 , 10.20
Financial and insurance activities	4 480	9 033	76.72	447.50	5.61	4.09 , 7.98
Real estate activities	1 848	12 746	340.72	1 350.77	4.05	2.96 , 5.56
Professional, scientific and technical activities	12 598	24 079	108.73	698.54	6.19	4.57 , 8.42
Administrative and support service activities	7 001	188 076	264.69	2 210.68	8.34	6.09 , 11.23
Education	6 177	14 236	63.63	989.93	13.6	8.13 , 22.57
Human health and social work activities	35 872	193 735	1 254.11	2 050.27	1.53	1.26 , 1.94
Arts, entertainment and recreation	10 372	27 462	747.89	2 774.39	2.84	1.19 , 7.61
Other service activities	4 484	23 030	324.53	1 095.01	3.10	2.37 , 4.40

Table 9 Standardized rates (per 1 000 000 p-h) and Ratio (SIRX) of severe work injuries among private sector employees in France, by industry, 2009-2012. (Weighted data)

NAF sector	Severe work injuries		Rate per 1,000,000 person hours		SIRx	CI 95%
	Managers	Non-managers	Managers	Non-managers		
All sectors	150 727	1 569 860	1,58	16,45	10.28	7.82 , 12.73
Mining and quarrying	145	2 352	1,95	14,18	6.94	4.38 , 11.07
Manufacturing	13 023	250 387	0,89	14,08	15.45	11.93 , 19.42
Electricity, gas, air conditioning supply	125	1 365	0,47	6,85	14.21	10.55 , 18.84
Water supply, sewerage, waste management	818	21 966	1,73	23,84	12.30	7.62 , 19.25
Construction	10 936	272 754	3,02	29,61	9.37	5.51 , 15.82
Wholesale and retail trade, repair of vehicles	29 742	253 695	2,20	14,53	6.32	5.04 , 7.8
Transportation and storage	5 407	168 422	1,99	23,38	11.15	9.26 , 13.41
Accommodation and food service activities	11 602	100 310	8,99	18,30	1.60	1.09 , 2.62
Information and communication	5 846	6 212	0,44	3,53	7.64	5.58 , 10.23
Financial and insurance activities	4 028	9 033	0,46	2,99	6.06	4.61 , 8.31
Real estate activities	1 666	12 746	2,13	9,44	4.49	3.42 , 5.89
Professional, scientific and technical activities	11190	24 079	0,70	4,82	6.67	5.11 , 8.7
Administrative and support service activities	6 285	18 8076	1,91	22,11	11.45	8.89 , 14.5
Education	5 508	14 236	0,52	9,03	16.62	10.29 , 26.65
Human health and social work activities	31 401	193 735	9,73	15,92	1.53	1.18 , 2.04
Arts, entertainment and recreation	9 033	27 462	7,98	26,18	2.28	0.87 , 6.75
Other service activities	3 972	23 030	2,26	8,89	3.44	2.56 , 5.03

Figure 9 Network of key articles on administrative data in health and occupational research.
Source : ResearchRabbit
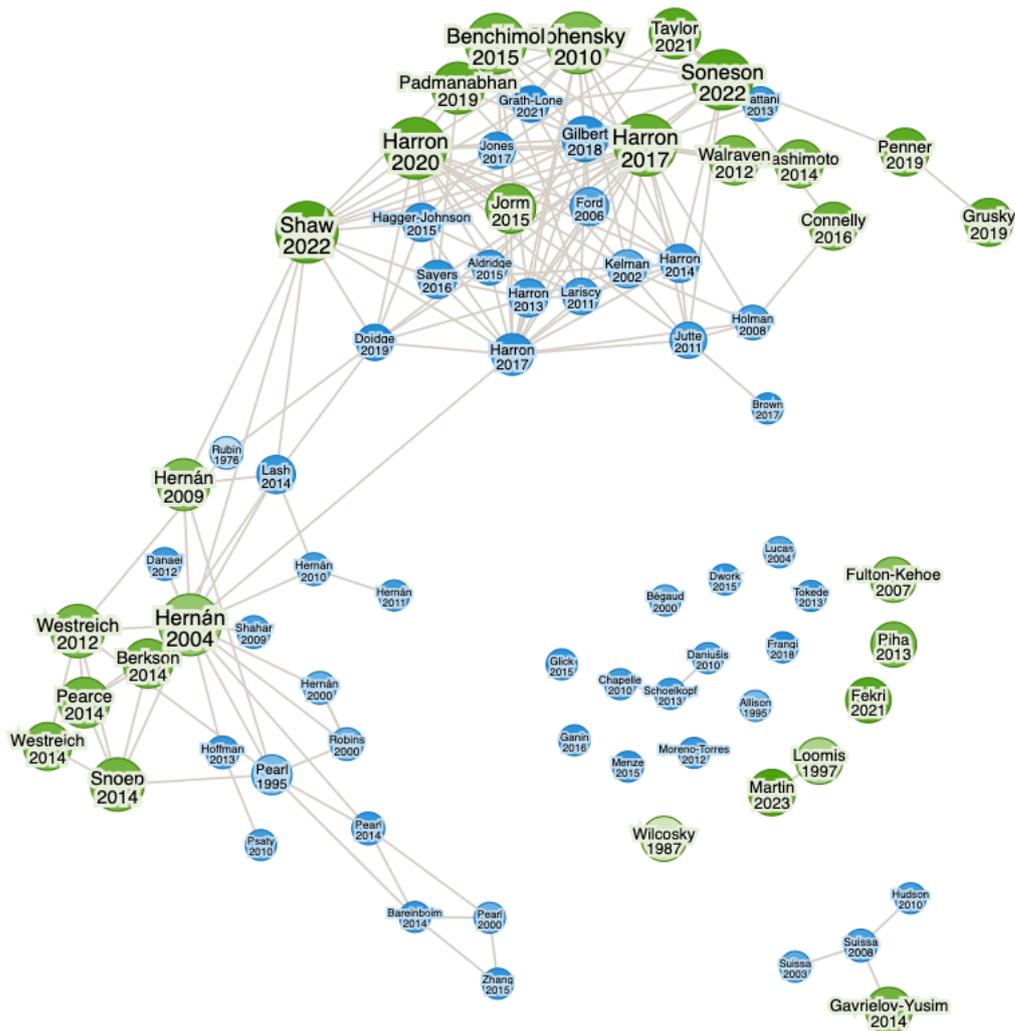
This graph visualizes the citation relationships among studies read during the literature search process, highlighting key clusters of foundational and recent work. Nodes represent individual studies (first author and year), and edges indicate citation links. Prominent clusters include methodological developments around bias and data linkage (center), theoretical frameworks on selection and measurement bias (bottom left), and studies focused on occupational health applications (right). The position and density of nodes reflect the degree of influence and connectivity within the field.