Article

HOW TO USE BIOMECHANICAL JOB EXPOSURE MATRICES WITH JOB HISTORY TO ACCESS WORK EXPOSURE FOR MUSCULOSKELETAL DISORDERS? APPLICATION OF MATHEMATICAL MODELING IN SEVERE KNEE PAIN IN THE CONSTANCES COHORT.

Guillaume Deltreil [1,2,†], Patrick Tardivel [3], Piotr Graczyk [2,‡], Mikael Escobar-Bach [2,‡], and Alexis Descatha [1,4,5‡]

[1] Université d'Angers, CHU Angers, Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMRS 1085, SFR ICAT, Angers, France; guillaume.deltreil@univ-angers.fr alexis.descatha@inserm.fr
[2] Université d'Angers, CNRS, LAREMA, SFR MATHSTIC, Angers, France;
[3] Université de Bourgogne Franche-Comté, UMR 5584 CNRS, Dijon France;
[4] Epidemiology and Prevention, Donald and Barbara Zucker School of Medicine, Hofstra Northwell, USA ;
[5] CHU Angers, Centre Antipoison-Centre de Données Cliniques, Angers, France ;
* Correspondence: guillaume.deltreil@univ-angers.fr
† Current address: Affiliation 1.
‡ These authors contributed equally to this work.

Abstract: Introduction. Musculoskeletal disorders related to work might be caused by the cumulative effect of occupational exposures during working life. We aimed to develop a new model which allows to compare the accuracy of duration of work and intensity/frequency associations in application to severe knee pain. Methods. From the CONSTANCES cohort, 66,553 subjects who were working at inclusion and coded were included in the study. The biomechanical job exposure matrix "JEM Constances" was used to assess the intensity/frequency of heavy lifting and kneeling/squatting at work together with work history to characterize the association between occupational exposure and severe knee pain. An innovative model G was developed and evaluated, allowing to compare the accuracy of duration of work and intensity/frequency associations. Results. The mean age was 49 years at inception with 46 percent of women. The G model developed was slightly better than regular models. Among the men subgroup, odds ratios of the highest quartile for the duration and low intensity were not significant for both exposures, whereas intensity/duration were for every duration. Results in women were less interpretable. Conclusion. Though higher duration increased strength of association with severe knee pain, intensity/frequency were important predictors among men. Exposure estimation along working history should have emphasis on such parameters, though other outcomes should be studied and have a focus on women.

Keywords: Occupational; Musculskeletal; pain; lifecourse; mathematical modeling.

0. Introduction

Musculoskeletal disorders (MSDs) related to repetitive and physically demanding working conditions continue to represent one of the largest occupational disease in industrialized countries[1,2]. MSD related to work are caused by non-traumatic injuries, with a possible cumulative effect of occupational exposures during working life, mainly for degenerative disorders like osteoarthritis, and severe knee pain[3,4].

The evaluation of biomechanical exposures can be done in different ways and expressed around two main dimensions, intensity/frequency per day and duration over the year[5]. The exposure assessment can be obtained by estimations based on subjective judgments (self-reports, expert judgments), systematic observations (observations at the

 doi:10.20944/preprints202209.0489.v1

workplace, video recording), and direct measurements (at the workplace or in laboratory). However, these techniques are problematic when past exposure evaluations are needed.

Job exposure matrices (JEMs) are commonly used in occupational epidemiology research for the evaluation of past exposures [6]. Indeed, JEMs allow estimating participants exposures to occupational factors based on job titles, industry sector, and population exposure data. Several biomechanical JEMs have become available recently [7–11]. A cumulative exposure index is commonly used to assess cumulative work exposure by multiplying duration and intensity/frequency. However, it is not clear how to consider the combination of intensity/frequency when assessing exposure over the years.

Thus, before optimizing models using relevant statistical methods, we first aimed to determine if low level exposure with high duration is equivalent to high level exposure with low duration in the example of severe knee pain and two occupational exposures: carrying heavy loads and kneeling/squatting.

We then aimed develop a new model and compare the accuracy of duration of work and intensity/frequency associations in application to the knee disorders using severe knee pain as an outcome in a large cohort study, by developing an innovative model.

1. Materials and Methods

1.1. Population

The CONSTANCES study is a French general population-based cohort[12]. More than 200,000 participants, aged 18-69 years, were recruited between 2012 and 2020 in 23 health screening centers across France. The recruitment was limited to people affiliated to the French National Health Insurance Fund who correspond to active or former salaried workers and their families and excludes agricultural and self-employed workers[12]. At enrolment, self-administered questionnaires were sent to participants to collect data including lifestyle, life events, health, and occupations. Variables of interest were collected from the baseline self-administered questionnaire and the medical interview.

For this work, we used French CONSTANCES clean data from 2020. Subjects from this cohort were active at their inclusion with work trajectory coded.

1.2. Variables of interest

Participants' sex, age at inception, known inflammatory disease of the joints, regular leisure activity (sports, gardening yes/no), Center for Epidemiologic Studies-Depression Scale (CES-D) into two categories (yes/no), were retrieved from the baseline questionnaire, and body mass index from the medical examination.

JEM Constances, which is based on self-reported exposure was used to evaluate occupational exposure[13]. In the JEM, occupational exposure is rated from 0 to 4 for intensity/frequency of heavy lifting ("lifting") and 1 to 4 of kneeling of squatting ("kneeling") based on reported job titles. The JEM Constances was combined with participants reported work trajectory that were coded at baseline retrospectively.

The main outcome was reporting severe knee pain, collected from the self-reported questionnaire at inception: yes if knee pain intensity >5/10 or having knee pain for more than a month per year.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2022 doi:10.20944/preprints202209.0489.v1

1.3. Mathematical modeling

In order to study the influence of duration and intensity/frequency on the onset of the disease, several logistic regression models were built for the two exposures separetely: heavy lifting and kneeling.

The method we propose in this article is based on a generalization of the logistic regression approach. Formally, we define $(Y, X_{i1}, \dots, X_{ip})_{[1,n]}$ an i.i.d sample in $\{0,1\} \times \mathbb{R}^p$ of size $n \in \mathbb{N}^*$ where Y is the response variable and corresponds to the illness status of each subject (if the subject is sick $Y = 1$, if else $Y = 0$) and $X_i = (X_{i1}, \dots, X_{ip})$ are the variables $X_{i1}, \dots, X_{ip-6}$ corresponding to a score based on occupation times and levels exposures and $X_{ip-5}, \dots, X_{ip}$ corresponding to others adjustements variables.

For a given i conditionally on $X_i := (X_{i1}, ..., X_{ip})$, Y_i follows a Bernoulli distribution such as:

$$\mathbb{P}(Y_i = 1|X_i) = \pi_{\beta^*}(X_i) := \frac{\exp(<\beta^*, X_i>)}{1 + \exp(<\beta^*, X_i>)}, \tag{1}$$

where $\beta^* = (\beta_0^*, \beta_1^*, \dots, \beta_p^*) \in \mathbb{R}^p$ is an unknown vector of parameters to be estimated. We estimate the parameter β^*, given in (1), by Maximal Likelihood Estimation, i.e. by minimizing the normalized opposite of the log likelihood $\gamma_n(\beta)$ over $\mathbb{R}^p$:

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \gamma_n(\beta).$$

1.3.1. Statistical models

For the occupational health data from the cohort CONSTANCES, several logistic models were possible, depending on the total duration value of the careers and the average exposures of the individuals. We describe them in our cohort via the variables $T_i = (T_{i1}, \dots, T_{ik_i})$, $N_i = (N_{i1}, \dots, N_{ik_i})$ and $a_{1i}, \dots, a_{6i}$ respectively for occupation times, levels of exposures for each k_i jobs held by the subject i as well as six adjustment variables (a_{1i}= sex, a_{2i}= age, a_{3i}= imc, a_{4i}= leisure, a_{5i}= arthrite, a_{6i}= depression). We must take into account that the number of jobs k_i can be very different. The models A,B and C are defined through the three following transformations:

$$X_i^A = \sum_{j=1}^{m} N_{ij} T_{ij}$$

$$X_i^B = \sum_{j=1}^{m} N_{ij}^B T_{ij}, \quad \text{with } N_{ij}^B = \begin{cases} 0, & \text{if } N_{ij} \leq 1 \\ N_{ij}, & \text{if else.} \end{cases}$$

$$X_i^C = \sum_{j=1}^{m} N_{ij}^C T_{ij}, \quad \text{with } N_{ij}^C = \begin{cases} 0, & \text{if } N_{ij} \leq 3 \\ 1, & \text{if else.} \end{cases}$$

Thus, Y_i follows a Bernoulli distribution such that $P(Y_i = 1|X_i) = \text{logit}(\beta_0^* + \beta_1^* X_i^\bullet + \sum_{j=1}^{6} \beta_{j+1}^* a_{ji})$ where $X_i^\bullet$ is to be replaced by X_i^A, X_i^B or X_i^C depending on the variable considered. We specify that the model B differs, from the model A, because in the computation of its transformation the exposure levels between 0 and 1 are confounded. The model C differs, from the model A, because in the computation of its transformation the lowest exposure levels are nullified. In order to define the model G we need to introduce two transformations ϕ_1 and ϕ_2:

- $\phi_1(T_i) = \sum_{j=1}^{k_i} T_{ij}$ corresponding to the total duration of the career range.
- $\phi_2(T_i, N_i) = \frac{\sum_{j=1}^{k_i} N_{ij} g(T_{ij})}{\sum_{j=1}^{k_i} g(T_{ij})}$, where $g(x) = 1 - \exp(-x)$ which is the time-weighted average of the exposure level.

The construction of the design matrix $(X_{ij})_{(1 \leq i \leq n, 1 \leq j \leq p)}$ of the model G is based on a semi-coding of the variables $\{\phi_1(T_i)\}_{1 \leq i \leq n}$ and $\{\phi_2(T_i, N_i)\}_{1 \leq i \leq n}$. More precisely, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2022 doi:10.20944/preprints202209.0489.v1

membership in a group is determined by the belonging of the values $\phi_1(T_i)$ and $\phi_2(T_i, N_i)$ to different given intervals. For this study, we consider the classes of intervals

$$\mathcal{G}_1 := \{[q_j, q_{j+1}[, \ j = 1, \dots, q\} \quad \text{and} \quad \mathcal{G}_{2,c} := \{\ [j-1, j[\ \}_{c \le j \le 4} \cup \{4\}, c = 1, 2$$

where $q_j \in \mathbb{N}^*$ defines fixed empirical quantiles of the sample $\{\phi_1(T_i)\}_{1 \le i \le n}$. When $c = 1$, the exposure considered is carrying a heavy load, and when $c = 2$ when the exposure considered is kneeling. The set of groups is defined as the class $\mathcal{G} := \{G_j, \ j = 1, \dots, p\} = \mathcal{G}_1 \times \mathcal{G}_2$ where $p = (6 - c)q$. An individual is then associated with the group $j \in [\![1, p]\!]$ with $X_{ij} = 1$ if and only if $(\phi_1(T_i), \phi_2(T_i, N_i)) \in G_j$.

1.3.2. Selection of the design matrix for model G

We propose to choose the design matrix X for the model G by a recent model selection procedure, introduced and described in the article "Model selection in logistic regression" by Kwemou et al. [14]. The mathematical guarantees for this model selection method are based on oracle inequalities from Birgé and Massart [15]. This model selection is performed using penalized maximum likelihood estimators which will allow us to choose the best design matrix.

Let $\mathcal{F}$ be a family of design matrices $X^{(m)}$, $m \in \{1, \dots, M\}$. For each $m \in \{1, \dots, M\}$, we define $\hat{\beta}_m$ as the estimator obtained by minimizing γ_n over $\mathbb{R}^{p_m}$ where p_m is the number of columns of $X^{(m)}$, namely:

$$\hat{\beta}_m = \arg \min_{\beta \in \mathbb{R}^{p_m}} \gamma_n(\beta)$$

We use a data driven strategy that selects the best matrix among the family $\mathcal{F}$. For this purpose we use a penalized maximum likelihood criterion for choosing the index m associated to the appropriate design. Here, we consider the Akaike information criterion where p_m corresponds to the number of parameters in the model (1), i.e. p_m is the number of columns of the matrix $X^{(m)}$ and

$$\widehat{m} = \arg \min_{m \in \{1, \dots, M\}} \left\{ \gamma_n(\widehat{\beta}_m) + p_m \right\}.$$

Hence, minimizing this criterion allows us to find the best design matrix $X^{\widehat{m}}$.
The "lifting" and "kneeling" quartiles of high exposure is considered in terms of duration and intensity/frequency:

- into 5 categories for heavy lifting $\mathcal{G}_{2,1} := \{\ [j-1, j[\ \}_{1 \le j \le 4} \cup \{4\}$,
- into 4 categories for kneeling $\mathcal{G}_{2,2} := \{\ [j-1, j[\ \}_{2 \le j \le 4} \cup \{4\}$.

1.4. α-divergence and Statistical tests

The logistic regression formula given in (1) depends on the model A, B, C and G. We wish to compare these statistical models, via a differentiation criterion, in order to identify the most appropriate model for this formula. Let r be the unknown probability that a randomly selected subject from the population is sick. It is natural to estimate the unknown parameter r by the proportion of patients observed in our dataset: $\hat{r}_1 = \frac{1}{n} \sum_{i=1}^{n} Y_i$. One can also estimate r from the logistic regression formula via the estimator $\hat{r}_2 = \frac{1}{n} \sum_{i=1}^{n} \pi_{\beta^*}(X_i)$. Note that unlike $\hat{r}_1$, the estimator $\hat{r}_2$ depends on the model. For an appropriate model it is natural to choose the model for which the estimators $\hat{r}_1$ and $\hat{r}_2$ are close.

We here chose to rely on the work of A.Basu et al. [16] who developped a robust estimator for the density function. The criterion used in this article is based on the α-divergence between two densities f and g (relative to a measure μ) defined for an $\alpha > 0$ as follows:

$$d_\alpha(g, f) = \int \left\{ f^{1+\alpha}(z) - \left(1 + \frac{1}{\alpha}\right) g(z) f^\alpha(z) + \frac{1}{\alpha} g^{1+\alpha}(z) \right\} d\mu(z)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2022 doi:10.20944/preprints202209.0489.v1

Here, we need to use the α-divergence between two Bernoulli laws with parameters $r_1 \in]0.1[$ and $r_2 \in]0.1[$ which we define by

$$d_\alpha(r_1, r_2) = \begin{cases} r_2^{1+\alpha} + (1-r_2)^{1+\alpha} - \left(1+\frac{1}{\alpha}\right)((1-r_2)^\alpha(1-r_1) + r_2^\alpha r_1) \\ +\frac{1}{\alpha}((1-r_1)^{1+\alpha} + r_1^{1+\alpha}) & \alpha > 0 \\ r_1 \ln\left(\frac{r_1}{r_2}\right) + (1-r_2)\ln\left(\frac{1-r_1}{1-r_2}\right) & \alpha = 0 \end{cases}$$

The parameter α here defines a trade-off between the estimation efficiency and the variability robustness. Note that when α is close to 0 we retrieve the Kullback–Leibler divergence and when $\alpha = 1$ we have $d_1(r_1, r_2) = 2(r_1 - r_2)^2$.
Next, we compared the p-value of the Wald's test (with Bonferroni correction) and odds ratios (OR) of the highest quartile for the duration with low intensity/frequency and the highest quartile for the intensity/frequency with low duration. Stratification on sex was performed as sensitivity analysis. For both exposures, logistic models were built adjusted on relevant variables. We compared the p-value of the Wald's test (with Bonferroni adjustments) and odds ratios of different quartile of duration and intensity/frequency and with low duration and lowest intensity as reference.

1.5. Analysis plan

After a brief description of the sample and the available adjustement variables, we assessed the performance score based on the α-divergence allowing the comparison of different models with α varying with each relevant occupational exposure. For a small value of parameter α, the decision of the model choice can be questioned but when α increases the choice of model becomes easier and the model G is selected.

Then, we were able to compare Odds ratios with the Wald's test. Sex stratification has been performed as primary analysis (whereas 45 year of age stratification has been considered in secondary analysis).

Analyses were performed on Python 3.8.8 (statsmodels,statistics,pandas,numpy).

2. Results

The sample included 66,553 subjects (Table 1), with 30,765 women (46.2%) and with a median of 49 years old (21 years of employment). The subjects reported regular primary leisure activities in 43.1% of cases, with half of sample who were overweight/ obese, and an important part (21.2%). who had a positive CES-D, suggesting a depression.

2.1. Figures, Tables and Schemes

Table 1. Description of the sample.

Variables		**N(%)**	**Mean (SD)**
Sex	Men	35788 (53.7)	
	Women	30765 (46.2)	
Age (years)			48.5 (13.1)
Body Mass Index (kg/m²)			25.07 (24.0)
Leisure activites	Yes	28693 (43.1)	
	No	37860 (56.9)	
Inflammatory Ostearthritis	Yes	950 (1.4)	
	No	65603 (98.6)	
Depression	Yes	14095 (21.2)	
	No	52458 (78.8)	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2022 doi:10.20944/preprints202209.0489.v1

Table 2. Comparison of performance score for severe knee pain with lifting.

α	**A**	**B**	**C**	**G**
0	0.305	0.329	0.326	**0.192**
0.25	0.162	0.228	0.218	**0.019**
0.5	0.155	0.221	0.211	**0.017**
0.75	0.146	0.210	0.200	**0.015**
1	0.136	0.198	0.189	**0.013**

Table 3. Comparison of performance score for severe knee pain with kneeling.

α	**A**	**B**	**C**	**G**
0	0.282	0.304	0.297	**0.194**
0.25	0.101	0.160	0.145	**0.021**
0.5	0.096	0.153	0.138	**0.018**
0.75	0.092	0.143	0.13	**0.016**
1	0.088	0.134	0.121	**0.014**

In order to study intensity and duration of exposure, we first selected the best model between A to G. We compared this familly of models by increasing α. For any value of parameter alpha, the model G is selected (Tables 2 and 3).

For heavy lifting as well as kneeling, intensity/ frequency was the most important predictor. Duration increased risks only for men with a dose response relationship. For women, only intensity and frequency in low duration for both exposures seemed associated with the knee pain (Tables 4 and 5). Although the statistical power was lower, the pattern of association was similar for participants aged less and more than 45 years (supplemental Tables 6 and 7).

Table 4. Results of adjusted logistic regression of severe knee pain with lifting, for men and women separately.

Variable		**Men**			**Women**		
Duration	Intensity/Frequency	OR	IC 95%	p-value	OR	IC 95%	p-value
low	[1,2[	1.83	[1.53 , 2.19]	$< 10^{-4}$	1.46	[1.29 , 1.66]	$< 10^{-4}$
low	[2,3[	2.08	[1.71 , 2.52]	$< 10^{-4}$	1.47	[1.20 , 1.81]	$< 10^{-4}$
low	[3,4[	1.96	[1.42 , 2.70]	$< 10^{-4}$	1.38	[0.91 , 2.09]	0.13
low	{4}	2.49	[1.86 , 3.33]	$< 10^{-4}$	2.81	[1.60 , 4.92]	$< 10^{-4}$
medium	[0,1[	1.07	[0.91 , 1.27]	0.41	0.86	[0.77 , 0.96]	0.01
medium	[1,2[	1.63	[1.35 , 1.97]	$< 10^{-4}$	1.11	[0.97 , 1.28]	0.13
medium	[2,3[	2.39	[1.97 , 2.90]	$< 10^{-4}$	1.28	[1.02 , 1.61]	0.04
medium	[3,4[	2.45	[1.78 , 3.39]	$< 10^{-4}$	1.08	[0.62 , 1.86]	0.79
medium	{4}	2.71	[1.84 , 3.98]	$< 10^{-4}$	1.09	[0.41 , 2.93]	0.86
high	[0,1[	1.09	[0.92 , 1.29]	0.33	0.87	[0.77 , 0.98]	0.03
high	[1,2[	1.77	[1.49 , 2.11]	$< 10^{-4}$	1.07	[0.93 , 1.23]	0.36
high	[2,3[	2.12	[1.77 , 2.53]	$< 10^{-4}$	1.09	[0.86 , 1.37]	0.49
high	[3,4[	2.41	[1.87 , 3.11]	$< 10^{-4}$	1.37	[0.99 , 1.91]	0.06
high	{4}	3.56	[2.68 , 4.72]	$< 10^{-4}$	1.67	[0.68 , 4.12]	0.27

Adjusted on body mass index, leisure activity, inflammatory osteoarthritis, depression, and age/sex when not stratified ; Reference duration low, Intensity/Frequency [0,1[.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2022 doi:10.20944/preprints202209.0489.v1

Table 5. Results of adjusted logistic regression of severe knee pain with kneeling, men and women separately.

Variable		**Men**			**Women**		
Duration	Intensity/Frequency	OR	IC 95%	p-value	OR	IC 95%	p-value
low	[2,3[	1.97	[1.67 , 2.31]	$< 10^{-4}$	1.34	[1.16 , 1.53]	$< 10^{-4}$
low	[3,4[	1.75	[1.40 , 2.20]	$< 10^{-4}$	1.44	[1.22 , 1.71]	$<10^{-4}$
low	{4}	1.98	[1.51 , 2.60]	$< 10^{-4}$	1.35	[1.01 , 1.82]	0.04
medium	[1,2[	1.06	[0.91 , 1.23]	0.48	0.83	[0.74 , 0.93]	$<10^{-4}$
medium	[2,3[	1.86	[1.57 , 2.20]	$< 10^{-4}$	1.02	[0.88 , 1.17]	0.83
medium	[3,4[	2.19	[1.72 , 2.79]	$< 10^{-4}$	1.23	[1.03 , 1.48]	0.02
medium	{4}	2.66	[1.91 , 3.71]	$< 10^{-4}$	0.91	[0.59 , 1.39]	0.66
high	[1,2[	1.10	[0.94 , 1.28]	0.26	0.85	[0.76 , 0.96]	0.01
high	[2,3[	1.80	[1.54 , 2.11]	$< 10^{-4}$	1.03	[0.89 , 1.19]	0.67
high	[3,4[	2.16	[1.78 , 2.62]	$< 10^{-4}$	0.88	[0.74 , 1.04]	0.13
high	{4}	2.55	[1.99 , 3.28]	$< 10^{-4}$	1.38	[1.03 , 1.84]	0.03

Adjusted on body mass index, leisure activity, inflammatory osteoarthritis, depression, and age/sex when not stratified ; Reference duration low, Intensity/Frequency [1,2[.

3. Discussion

This is the first study that used a developed mathematical model to compare the effect of duration and intensity during working life, on the association with severe knee as a proxy of degenerative musculoskeletal disorders. The new model G was found to be better than the usual ones, though the difference was minor. For men, we found that the OR of the highest quartile for the duration and low intensity is not significant for both exposures, whereas intensity/duration is significant for every duration, with a dose response relationship. Results for women were limited.

As expected, there was as an important effect of the intensity of heavy lifting and kneeling on severe knee pain. Both exposures are known to be associated with knee disorders [17–19]. The dose response relationship has been described previously. Jensen calculated an equivalent of our model A using an individual exposure from the number of knee-straining activities and the number of years in the trade within a collective of floor layers, carpenters and compositors. The ORs for knee complaints and radiographically determined knee osteoarthritis were 3.0 (95% CI, 0.5 to 17.2) in the low-exposure group, 4.2 (95% CI, 0.6 to 27.6) in the medium-exposure group, and 4.9 (95% CI, 1.1 to 21.9) in the high-exposure group compared with the zero-exposure group [20]. There is an important difference in the strength of the associations compared to our work but it should be explained by the large population design with JEM exposure methods. Indeed, high exposure is considered using the proxy of job title with a large variation inside job categories.

Also expected, in a previous review on occupational exposure and knee osteoarthritis, [21] lifting and carrying of loads was significantly associated with severe knee pain. Knee osteoarthritis was also found associated with lifting and carrying of loads with a dose-response relationship : OR of 2.0 (95% CI, 1.1 to 3.6) in the exposure group 630 to <5,120 kg-hours over life, up to an OR of 2.6 (95% CI, 1.1 to 6.1) in the highest exposure group (>37,000 kg-hours over life) in men [20].

The lack of clear association for women was also found by D'Souza et al. who reported on an analysis of the US national survey, where they describe relationships between work activities and symptomatic knee osteoarthritis [22]. A significant exposure-response relationship was only found between symptomatic knee osteoarthritis and kneeling in men but not women. Different explanations might be suggested: since our model included adjusting factors like BMI and depression, there might be a more complex causal pathways than in men such as considered in back pain [23]. JEM Constances is not gender stratified and applying a specific JEM for sex could be a lead for another study. Furthermore, selection effect similar to healthy workers effect is also possible. The main strength of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 September 2022 doi:10.20944/preprints202209.0489.v1

our study was the possibility to use JEM with working life course on a large cohort study. Limitations might also be raised by using a large but not representative population of French workers, and specific jobs in agriculture (not included by design) and mining (almost disappeared in France) should be considered since they are known factors related to lower limb MSD [19]. Second, the outcome was focused on severe knee pain. Even, it is self-declared and might correspond to heterogenous disorders, work-related or not. It was used as a good example of a proxy of degenerative musculoskeletal disorders, and the use of pain intensity and severity is recommended [24]. We have previously shown that working in a kneeling or squatting position was significantly associated with severe knee pain [18]. More recently, similar trends of associations between severe knee pain and knee arthroplasty groups were showed in the same cohort [25]. This result is also found elsewhere, with non-managerial jobs associated with higher prevalence of knee osteoarthritis and knee symptoms [26]. Third, as we already mentioned, the use of JEMs might also be questioned since it is a global average evaluation that does not consider the differences inside job [27]. However, assessing exposure during long periods of time and for a big number of subjects is challenging and JEMs are appropriate tool to consider. Furthermore, assessement of carrying heavy loads exposure using JEMs was found to be valid compared to a self-administrated questionnaire [28].

4. Conclusions

This innovative approach using mathematic modeling of working history and a JEM, shows that duration in years has a smaller impact than frequency/intensity and should be considered at least among men. Our new model G seems to be an interesting approach though the improvement is light. Further study, should be done on other outcomes, and have a focus on women.

Table 6. Results of adjusted logistic regression of severe knee pain with lifting, for <45 years old and 45 or more participants.

Variable		**<45 years**			**≥45 years**		
Duration	Intensity/Frequency	OR	IC 95%	p-value	OR	IC 95%	p-value
low	[1,2[	1.59	[1.35 , 1.87]	$< 10^{-4}$	1.51	[1.26 , 1.81]	$< 10^{-4}$
low	[2,3[	1.71	[1.37 , 2.14]	$< 10^{-4}$	1.62	[1.31 , 2.00]	$< 10^{-4}$
low	[3,4[	1.45	[0.95 , 2.19]	0.08	1.42	[0.93 , 2.16]	0.107
low	{4}	2.63	[1.83 , 3.79]	$< 10^{-4}$	1.31	[0.86 , 2.02]	0.212
medium	[0,1[	0.87	[0.75 , 1.02]	0.09	0.92	[0.79 , 1.06]	0.234
medium	[1,2[	1.33	[1.10 , 1.59]	0.002	1.24	[1.05 , 1.46]	0.01
medium	[2,3[	1.64	[1.28 , 2.10]	$< 10^{-4}$	1.57	[1.28 , 1.93]	$< 10^{-4}$
medium	[3,4[	1.79	[1.19, 2.70]	0.005	1.95	[1.34 , 2.84]	0.001
medium	{4}	3.30	[2.04, 5.34]	$< 10^{-4}$	2.25	[1.42, 3.58]	0.001
high	[0,1[	0.93	[0.77, 1.12]	0.442	0.88	[0.78, 1.01]	0.073
high	[1,2[	1.31	[1.07, 1.62]	0.01	1.21	[1.05, 1.39]	0.007
high	[2,3[	1.93	[1.51, 2.47]	$< 10^{-4}$	1.46	[1.25, 1.71]	$< 10^{-4}$
high	[3,4[	1.36	[0.85, 2.19]	0.204	1.69	[1.36, 2.09]	$< 10^{-4}$
high	{4}	1.63	[0.89, 2.98]	0.114	2.58	[1.97, 3.37]	$< 10^{-4}$

Adjusted on body mass index, leisure activity, inflammatory osteoarthritis, depression, and age/sex when not stratified ; Reference duration low, Intensity/Frequency [0,1[.

Table 7. Results of adjusted logistic regression of severe knee pain with kneeling, for <45 years old and 45 or more participants.

Variable		**<45 years**			**≥45 years**		
Duration	Intensity/Frequency	OR	IC 95%	p-value	OR	IC 95%	p-value
low	[2,3[	1.50	[1.28 , 1.78]	$< 10^{-4}$	1.49	[1.24 , 1.79]	$< 10^{-4}$
low	[3,4[	1.62	[1.31 , 2.01]	$< 10^{-4}$	1.45	[1.16 , 1.80]	0.001
low	{4}	1.80	[1.33 , 2.44]	$< 10^{-4}$	1.40	[1.04, 1.88]	0.025
medium	[1,2[	0.86	[0.74 , 1.00]	0.06	0.91	[0.79 , 1.05]	0.19
medium	[2,3[	1.45	[1.22 , 1.73]	$< 10^{-4}$	1.20	[1.02 , 1.41]	0.028
medium	[3,4[	1.34	[1.04 , 1.73]	0.026	1.47	[1.20 , 2.47]	$< 10^{-4}$
medium	{4}	1.24	[0.78, 1.97]	0.361	1.75	[1.24, 2.47]	0.001
high	[1,2[	0.91	[0.76, 1.09]	0.300	0.88	[0.78, 1.00]	0.050
high	[2,3[	1.38	[1.14, 1.69]	0.001	1.23	[1.08, 1.41]	0.002
high	[3,4[	1.59	[1.22, 2.09]	0.001	1.22	[1.05, 1.42]	0.011
high	{4}	1.42	[0.89, 2.26]	0.139	1.73	[1.41, 2.12]	$< 10^{-4}$

Adjusted on body mass index, leisure activity, inflammatory osteoarthritis, depression, and age/sex when not stratified ; Reference duration low, Intensity/Frequency [1,2[.

Author Contributions: Conceptualization, A.D. ; methodology P.G., M.E. and P.T.; software, G.D.; validation, PG,ME and PT; formal analysis, G.D.; investigation,G.D., P.G., M.E. and A.D.; writing—original draft preparation, A.D. and G.D.; writing—review and editing, P.G., M.E. and P.T.; All authors have read and agreed to the published version of the manuscript.

Funding: The study was founded by TEC-TOP project (Pays de la Loire region, Angers Loire Métropole, University of Angers, CHU of Angers). None of the institution had had any role in the collection of data, analyses and the writing of the manuscript. The CONSTANCES Cohort Study was supported and funded by the Caisse nationale d'assurance maladie (CNAM); it is an "Infrastructure nationale en Biologie et Santé" and benefits from ANR (ANR-11-INBS-0002) grant funding. CONSTANCES is also partly funded by MSD, AstraZeneca and Lundbeck.

Institutional Review Board Statement: "The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of CCTIRS (protocol in 2015, number 15-636) in addition of the Inserm ethic commitee for Constances cohort (in 2011 and updated in 2021, 01-011, 21-842).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study

Data Availability Statement: Personal health data underlying the findings of our study are not publicly available due to legal reasons related to data privacy protection. However, the data are available upon request to all interested researchers after authorization of the French "Commission nationale de l'informatique et des libertés". The CONSTANCES email address is contact@constances.fr.

Acknowledgments: We thank Sabrina Pitet for her help in improving manuscript tables and reference list. The authors also thank the team of the "Population-based Epidemiologic Cohorts Unit" (Cohortes en population) that designed and manages the CONSTANCES Cohort Study, as well as the volunteers.

Conflicts of Interest: The authors declare no conflict of interest (paid by their affiliations. Authors are paid by their institution, and A.D. is also paid as editor of the Archives des Maladies professionnelles et de l'Environnement (Elsevier).The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results".

Abbreviations

The following abbreviations are used in this manuscript:

OR odds ratio

References

1. Hulshof, C.T.; Pega, F.; Neupane, S.; van der Molen, H.F.; Colosio, C.; Daams, J.G.; Descatha, A.; Kc, P.; Kuijer, P.P.; Mandic-Rajcevic, S.; et al. The prevalence of occupational exposure to ergonomic risk factors: A systematic review and meta-analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. *Environment international* **2021**, *146*, 106157.
2. Yassi, A. Repetitive strain injuries. *The Lancet* **1997**, *349*, 943–947.
3. Descatha, A.; Evanoff, B.A.; Leclerc, A.; Roquelaure, Y. Occupational determinants of musculoskeletal disorders. *Handbook of disability, work and health* **2020**, pp. 169–188.
4. Verbeek, J.; Mischke, C.; Robinson, R.; Ijaz, S.; Kuijer, P.; Kievit, A.; Ojajärvi, A.; Neuvonen, K. Occupational exposure to knee loading and the risk of osteoarthritis of the knee: a systematic review and a dose-response meta-analysis. *Safety and health at work* **2017**, *8*, 130–142.
5. Van Der Beek, A.J.; Frings-Dresen, M. Assessment of mechanical exposure in ergonomic epidemiology. *Occupational and environmental medicine* **1998**, *55*, 291–299.
6. Siemiatycki, J.; Lavoué, J. Availability of a new job-exposure matrix (CANJEM) for epidemiologic and occupational medicine purposes. *Journal of occupational and environmental medicine* **2018**, *60*, e324–e328.
7. Vad, M.; Frost, P.; Svendsen, S. Occupational mechanical exposures and reoperation after first-time inguinal hernia repair: a prognosis study in a male cohort. *Hernia* **2015**, *19*, 893–900.
8. Rubak, T.S.; Svendsen, S.W.; Andersen, J.H.; Haahr, J.P.L.; Kryger, A.; Jensen, L.D.; Frost, P. An expert-based job exposure matrix for large scale epidemiologic studies of primary hip and knee osteoarthritis: the Lower Body JEM. *BMC Musculoskeletal Disorders* **2014**, *15*, 1–8.
9. Dalbøge, A.; Hansson, G.Å.; Frost, P.; Andersen, J.H.; Heilskov-Hansen, T.; Svendsen, S.W. Upper arm elevation and repetitive shoulder movements: a general population job exposure matrix based on expert ratings and technical measurements. *Occupational and Environmental Medicine* **2016**, *73*, 553–560.
10. Evanoff, B.; Zeringue, A.; Franzblau, A.; Dale, A.M. Using job-title-based physical exposures from O* NET in an epidemiological study of carpal tunnel syndrome. *Human factors* **2014**, *56*, 166–177.
11. Dale, A.M.; Zeringue, A.; Harris-Adamson, C.; Rempel, D.; Bao, S.; Thiese, M.S.; Merlino, L.; Burt, S.; Kapellusch, J.; Garg, A.; et al. General population job exposure matrix applied to a pooled study of prevalent carpal tunnel syndrome. *American journal of epidemiology* **2015**, *181*, 431–439.
12. Goldberg, M.; Carton, M.; Descatha, A.; Leclerc, A.; Roquelaure, Y.; Santin, G.; Zins, M.; et al. CONSTANCES: a general prospective population-based cohort for occupational and environmental epidemiology: cohort profile. *Occupational and Environmental Medicine* **2017**, *74*, 66–71.
13. Evanoff, B.A.; Yung, M.; Buckner-Petty, S.; Andersen, J.H.; Roquelaure, Y.; Descatha, A.; Dale, A.M. The CONSTANCES job exposure matrix based on self-reported exposure to physical risk factors: development and evaluation. *Occupational and environmental medicine* **2019**, *76*, 398–406.
14. Kwemou, M.; Taupin, M.L.; Tocquet, A.S. Model selection in logistic regression. *arXiv preprint arXiv:1508.07537* **2015**.
15. Birgé, L.; Massart, P. Gaussian model selection. *Journal of the European Mathematical Society* **2001**, *3*, 203–268.
16. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M. Robust and efficient estimation by minimising a density power divergence. *Biometrika* **1998**, *85*, 549–559.
17. Klussmann, A.; Gebhardt, H.; Nübling, M.; Liebers, F.; Quirós Perea, E.; Cordier, W.; von Engelhardt, L.V.; Schubert, M.; Dávid, A.; Bouillon, B.; et al. Individual and occupational risk factors for knee osteoarthritis: results of a case-control study in Germany. *Arthritis research & therapy* **2010**, *12*, 1–15.
18. Descatha, A.; Cyr, D.; Imbernon, E.; Chastang, J.F.; Plénet, A.; Bonenfant, S.; Zins, M.; Goldberg, M.; Roquelaure, Y.; Leclerc, A. Long-term effects of biomechanical exposure on severe knee pain in the Gazel cohort. *Scandinavian journal of work, environment & health* **2011**, *37*, 37.
19. Fransen, M.; Agaliotis, M.; Bridgett, L.; Mackey, M.G. Hip and knee pain: Role of occupational factors. *Best Practice & Research Clinical Rheumatology* **2011**, *25*, 81–101.
20. Jensen, L.K. Knee-straining work activities, self-reported knee disorders and radiographically determined knee osteoarthritis. *Scandinavian journal of work, environment & health* **2005**, pp. 68–74.
21. Jensen, L.K. Knee osteoarthritis: influence of work involving heavy lifting, kneeling, climbing stairs or ladders, or kneeling/squatting combined with heavy lifting. *Occupational and environmental medicine* **2008**, *65*, 72–89.
22. D'Souza, J.C.; Werner, R.A.; Keyserling, W.M.; Gillespie, B.; Rabourn, R.; Ulin, S.; Franzblau, A. Analysis of the Third National Health and Nutrition Examination Survey (NHANES III) using expert ratings of job categories. *American journal of industrial medicine* **2008**, *51*, 37–46.

23. Leclerc, A.; Gourmelen, J.; Chastang, J.F.; Plouvier, S.; Niedhammer, I.; Lanoë, J.L. Level of education and back pain in France: the role of demographic, lifestyle and physical work factors. *International archives of occupational and environmental health* **2009**, *82*, 643–652.
24. Hagberg, M.; Violante, F.S.; Bonfiglioli, R.; Descatha, A.; Gold, J.; Evanoff, B.; Sluiter, J.K. Prevention of musculoskeletal disorders in workers: classification and health surveillance–statements of the Scientific Committee on Musculoskeletal Disorders of the International Commission on Occupational Health. *BMC musculoskeletal disorders* **2012**, *13*, 1–6.
25. Valter, R.; Godeau, D.; Leclerc, A.; Descatha, A.; Fadel, M. Influence of severe knee pain, meniscus surgery and knee arthroplasty on physical ability: an observational study of 114 949 adults in the CONSTANCES cohort. *BMJ open* **2019**, *9*, e031549.
26. Lee, J.Y.; Han, K.; Park, Y.G.; Park, S.H. Effects of education, income, and occupation on prevalence and symptoms of knee osteoarthritis. *Scientific Reports* **2021**, *11*, 1–8.
27. Descatha, A.; Fadel, M.; Sembajwe, G.; Peters, S.; Evanoff, B.A. Job-Exposure Matrix: A Useful Tool for Incorporating Workplace Exposure Data Into Population Health Research and Practice.
28. Ngabirano, L.; Fadel, M.; Leclerc, A.; Evanoff, B.A.; Dale, A.M.; Roquelaure, Y.; Descatha, A. Comparison between a job-exposure matrix (JEM) score and self-reported exposures for carrying heavy loads over the working lifetime in the Constances cohort. *Annals of Work Exposures and Health* **2020**, *64*, 455–460.