

Trends in Work-related Injury Frequency Rate in Canada*

Maritime operations have experienced a notable decline, contributing to an overall downward trend

Missy Zhang

5 April 2023

This paper aims to examine the work-related injury frequency rate in Canada over the past 12 years. To explore whether and how injury rate has declined in Canada, this paper aims to investigate the relationship between time, industry, and work-related injury frequency rate in Canada. Using Statistics Canada data, it was found that there was an overall downward trend in the frequency of injuries over time, with the decline most pronounced in maritime operations. Particularly, it is closely related to the proportion of outdoor workers in the industry. In addition, this paper shares insights into injury prevention strategies for Canadian workplaces, and highlights the industries and worker populations that may be most vulnerable to workplace injuries.

1 Introduction

In recent years, work-related injuries have emerged as a critical issue in Canada, affecting workers across various industries. These injuries can have devastating physical and psychological impacts on workers and can also result in a significant economic burden on society. Despite a decrease in the number of injuries in recent years, there were still 37,024 work-related injuries reported in 2020 alone (Canada Employment and Social Development 2022). To address this issue and improve workplace safety, it is crucial to understand injury trends over time and across different sectors. This report aims to analyze work-related injury data and explore factors that contribute to work-related injuries. By identifying areas that require improvement and highlighting successful initiatives, this report aims to contribute to the ongoing efforts to create safer workplaces and reduce the incidence of work-related injuries.

*Code and data are available at: https://github.com/MissyZhang/occupational_injury.

This report utilizes data of injury statistics in Canada by industry sector from 2008 to 2020 obtained from the Open Government Data of Statistics Canada (Statistics Canada 2023) and used programming language R (R Core Team 2020) to build models and analyze trends. The estimand of this report is work-related injury frequency rate, which is measured by injuries (disabling, fatal and minor) per one million hours worked. This paper delves into the overall trend of injury frequency rate in Canada by industry using multiple linear regression model. It was found that there was an overall downward trend in the frequency of injuries over time, with the decline most pronounced in maritime operations. Particularly, it is closely related to the proportion of outdoor workers in the industry.

The remainder of the paper is split into four sections. In Section 2, the cleaned dataset is explored to identify key variables of interest, namely the year, industry, and injury frequency rate. Visualizations in the form of tables and figures are presented to illustrate potential correlations between these variables. Before constructing the multiple linear regression model in Section 3, it is observed through data visualization that the injury frequency rate in Canada has generally decreased over time, albeit with variations across different industries. The final model is interpreted and findings on the evolution of injury frequency rate are presented in Section 4. In Section 5, the implications of these findings on time and industry are discussed, along with potential federal and provincial-level solutions and limitations of this study.

2 Data

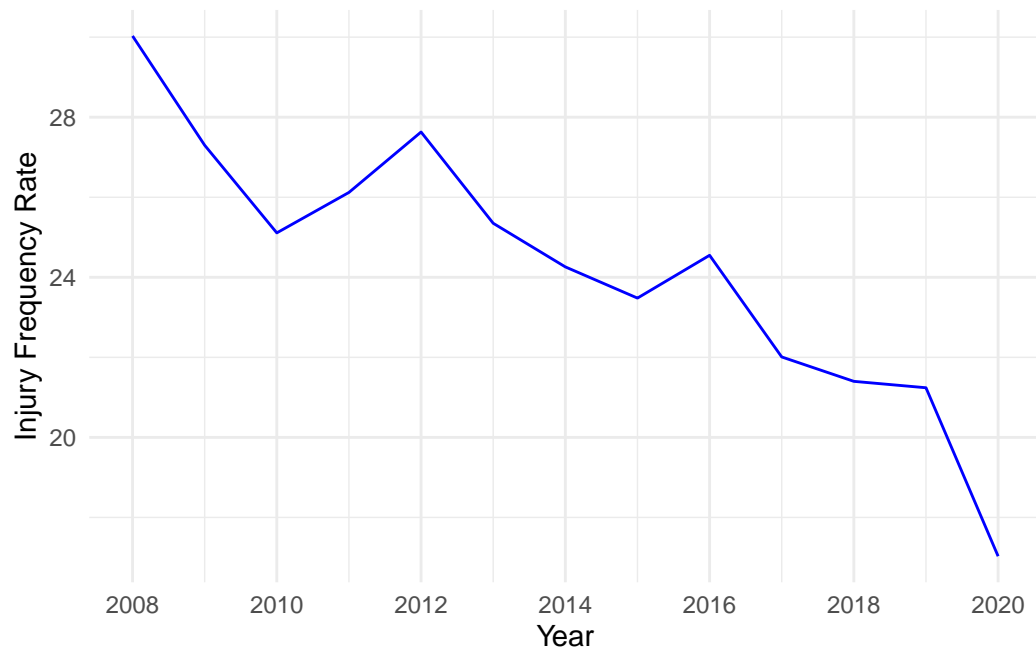


Figure 1: Work-related injury frequency rate by year in Canada from 2008 to 2020

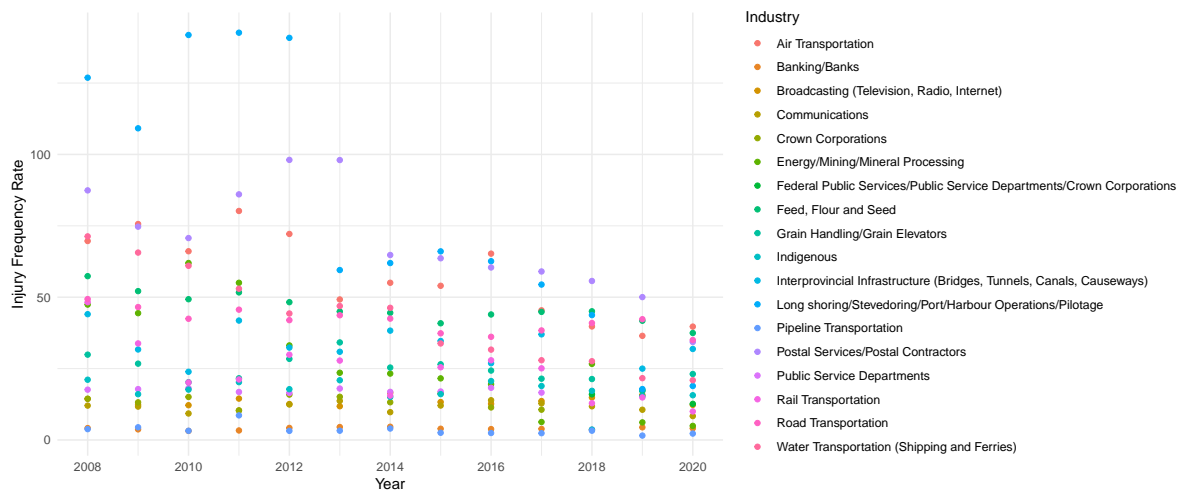


Figure 2: Work-related injury frequency rate by year and industry in Canada from 2008 to 2020

3 Model

$$Y_{ij} = \beta_0 + \beta_1 Year_i + \beta_2 Industry_j + \beta_3 Year_i Industry_j \quad (1)$$

In Model Equation 1:

- Y_{ij} is the injury frequency rate in i^{th} year and industry j .
- β_0 is the coefficient for intercept.
- β_1 is the coefficient for the continuous year variable.
- β_2 is the coefficient corresponding to industry j .
- β_3 is the coefficient for the interaction term between i^{th} year and industry j .
- The baseline of this model is year 0 and Air Transportation industry.

4 Results

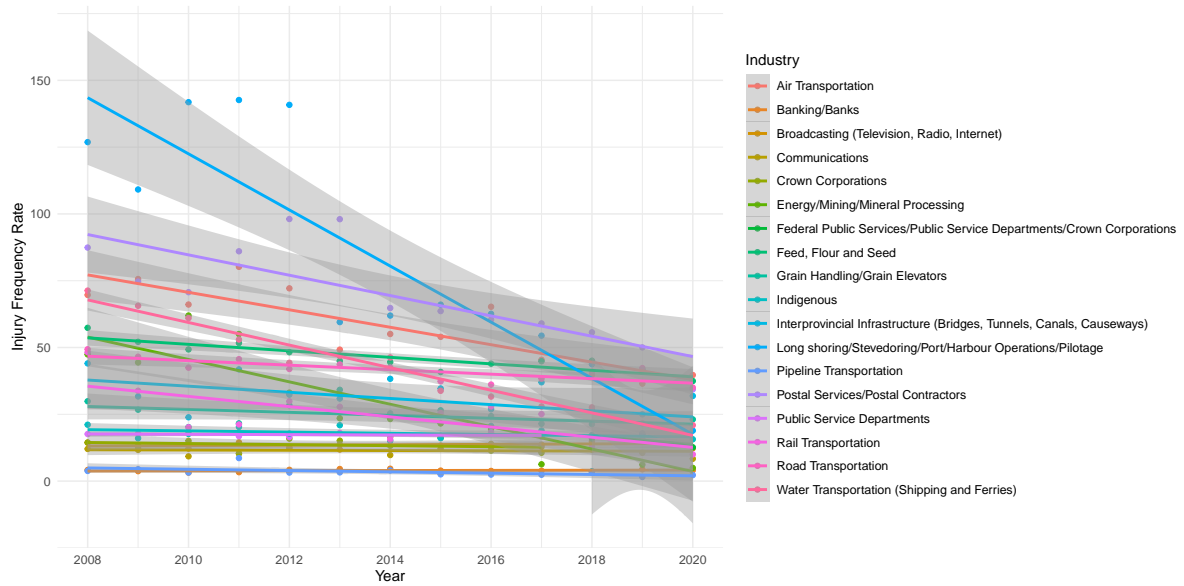


Figure 3: More bills of penguins

Table 1: Model Coefficients and 95 percent Confidence Interval

	Coefficients	Confidence Interval Lower Bound	Confidence Interval Upper Bound
Intercept	6858.4	4365.0	9351.8
Banking/Banks	-6899.6	-10592.4	-3206.7
Broadcasting (Television, Radio, Internet)	-6868.3	-10464.3	-3272.2
Communications	-6545.3	-10288.0	-2802.6
Crown Corporations	-6444.2	-10560.6	-2327.7
Energy/Mining/Mineral Processing	1498.1	-1956.8	4953.0
Federal Public Services/Public Service Departments/Crown Corporations	-3481.4	-24732.5	17769.7
Feed, Flour and Seed	-4299.8	-7708.2	-891.3
Grain Handling/Grain Elevators	-5831.6	-9165.8	-2497.4
Indigenous	-6384.7	-9793.8	-2975.6
Interprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways)	-6479.7	-10127.4	-2831.9
Long shoring/Stevedoring/Port/Harbour Operations/Pilotage	12705.9	9084.2	16327.5
Pipeline Transportation	-6303.4	-9723.9	-2882.9
Postal Services/Postal Contractors	-512.9	-4054.7	3028.9
Public Service Departments	-6583.9	-12730.4	-437.4
Rail Transportation	-3174.3	-6579.9	231.3
Road Transportation	-4923.1	-8358.9	-1487.3
Water Transportation (Shipping and Ferries)	1559.5	-2044.5	5163.6
Year	-3.4	-4.6	-2.1
Year: Banking/Banks	3.4	1.6	5.2
Year: Broadcasting (Television, Radio, Internet)	3.4	1.6	5.2
Year: Communications	3.2	1.4	5.1
Year: Crown Corporations	3.2	1.1	5.2
Year: Energy/Mining/Mineral Processing	-0.8	-2.5	1.0
Year: Federal Public Services/Public Service Departments/Crown Corporations	1.7	-8.8	12.2
Year: Feed, Flour and Seed	2.1	0.4	3.8
Year: Grain Handling/Grain Elevators	2.9	1.2	4.5
Year: Indigenous	3.1	1.5	4.8
Year: Interprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways)	3.2	1.4	5.0
Year: Long shoring/Stevedoring/Port/Harbour Operations/Pilotage	-6.3	-8.1	-4.5
Year: Pipeline Transportation	3.1	1.4	4.8
Year: Postal Services/Postal Contractors	0.3	-1.5	2.0
Year: Public Service Departments	3.2	0.2	6.3
Year: Rail Transportation	1.6	-0.1	3.3
Year: Road Transportation	2.4	0.7	4.1
Year: Water Transportation (Shipping and Ferries)	-0.8	-2.6	1.0

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

A Appendix

B Model Testing

R package “modelssummary”(Arel-Bundock 2022) helps to display the coefficients of the two models as well as the result of a series of tests to compare which model performs better.

R^2 measures how well the model explains its response variable’s variation. If R^2 is low, it indicates that the model doesn’t fit the data well. By (table-test?), in Model 1, 84.5% of the variability is explained, and in Model 2, 87.4% of the variability is explained. Both models demonstrate a pretty high R^2 value.

AIC and BIC measure the prediction ability of the multiple linear regression model. AIC (Akaike’s Information Criteria) focuses on how well the model fits unknown data, while BIC (Bayesian Information Criteria) focuses on the true model and favours simpler models (Kellen 2010). Lower AIC and BIC both indicate that the model has better prediction power. By (table-test?), Model 2 has slightly lower AIC and BIC than Model 1, implying that Model 2 has more prediction power.

Table 2: Comparing Model 1 and Model 2’s Statistics

	Model 1	Model 2
(Intercept)	3771.03 (487.75)	6858.39 (1261.02)
IndustryBanking/Banks	-54.07 (5.34)	-6899.55 (1867.60)
IndustryBroadcasting (Television, Radio, Internet)	-44.29 (5.20)	-6868.27 (1818.67)
IndustryCommunications	-49.59 (5.72)	-6545.29 (1892.81)
IndustryCrown Corporations	-48.75 (5.19)	-6444.17 (2081.85)
IndustryEnergy/Mining/Mineral Processing	-29.67 (4.97)	1498.10 (1747.28)
IndustryFederal Public Services/Public Service Departments/Crown Corporations	-34.97 (7.75)	-3481.40 (10747.51)
IndustryFeed, Flour and Seed	-12.49 (5.07)	-4299.77 (1723.77)
IndustryGrain Handling/Grain Elevators	-35.55 (4.97)	-5831.56 (1686.24)
IndustryIndigenous	-41.71	-6384.71

	Model 1	Model 2
IndustryInterprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways)	(5.19) -26.92	(1724.11) -6479.69
IndustryLong shoring/Stevedoring/Port/Harbour Operations/Pilotage	(5.19) 23.57	(1844.81) 12705.87
IndustryPipeline Transportation	(5.19) -56.38	(1831.60) -6303.43
IndustryPostal Services/Postal Contractors	(5.07) 10.57	(1729.89) -512.93
IndustryPublic Service Departments	(5.33) -47.68	(1791.22) -6583.88
IndustryRail Transportation	(5.75) -35.16	(3108.52) -3174.31
IndustryRoad Transportation	(4.97) -16.92	(1722.35) -4923.10
IndustryWater Transportation (Shipping and Ferries)	(5.08) -17.59	(1737.60) 1559.53
Year	(5.19) -1.84 (0.24)	(1822.71) -3.38 (0.63)
IndustryBanking/Banks \times Year		3.40 (0.93)
IndustryBroadcasting (Television, Radio, Internet) \times Year		3.39 (0.90)
IndustryCommunications \times Year		3.23 (0.94)
IndustryCrown Corporations \times Year		3.18 (1.03)
IndustryEnergy/Mining/Mineral Processing \times Year		-0.76 (0.87)
IndustryFederal Public Services/Public Service Departments/Crown Corporations \times Year		1.71 (5.32)
IndustryFeed, Flour and Seed \times Year		2.13 (0.86)
IndustryGrain Handling/Grain Elevators \times Year		2.88 (0.84)
IndustryIndigenous \times Year		3.15 (0.86)

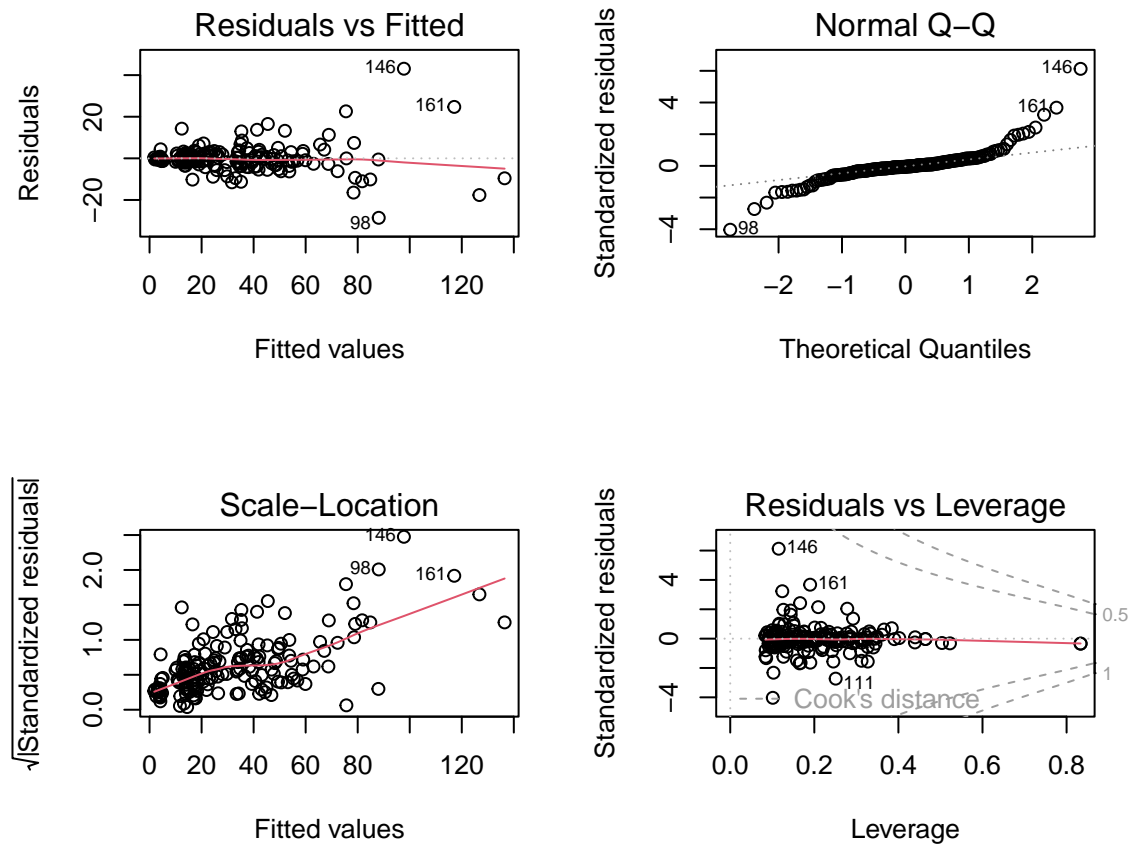
	Model 1	Model 2
IndustryInterprovincial Infrastructure (Bridges, Tunnels, Canals, Causeways) \times Year		3.21 (0.92)
IndustryLong shoring/Stevedoring/Port/Harbour Operations/Pilotage \times Year		-6.30 (0.91)
IndustryPipeline Transportation \times Year		3.10 (0.86)
IndustryPostal Services/Postal Contractors \times Year		0.26 (0.89)
IndustryPublic Service Departments \times Year		3.25 (1.55)
IndustryRail Transportation \times Year		1.56 (0.86)
IndustryRoad Transportation \times Year		2.44 (0.86)
IndustryWater Transportation (Shipping and Ferries) \times Year		-0.78 (0.91)
Num.Obs.	174	174
R2	0.817	0.932
R2 Adj.	0.795	0.915
AIC	1366.6	1227.5
BIC	1429.8	1344.4
Log.Lik.	-	-576.767
	663.324	
RMSE	10.95	6.66

Table 3: Comparing RMSE between two models

Dataset	Model 1	Model 2
train	10.9	6.7
test	15.9	9.3

The RMSE measures how far the predicted values of the multiple linear regression model are from their actual values on average. The lower the RMSE, the better the model performs regarding prediction. Based on (**table3?**), the RMSE for both two models are very similar for training and testing dataset, indicating that the dataset is unbiased and the two models are performing as expected. Moreover, Model 2 has lower RMSE than Model 1 for both training and testing dataset, indicating that Model 2 predicts data more accurately.

C Model Assumption Check



Check for assumptions: Assumptions to the multiple linear regression model are checked to ensure that the model is valid for this dataset. (figure3?) shows the plots used to check for the assumption.

The Residuals vs Fitted plot checks for the linear relationship assumption. Since the red line is almost horizontal and there isn't any pattern, the model satisfies the linearity assumption. The Normal QQ plot checks for the residual normality assumption. Since almost all the dots are on the dashed line, the residuals follow a normal distribution. The Scale-Location plot checks for the homoscedasticity assumption. The red line is horizontal, and the dots are evenly scattered, indicating that the variance of the residuals is constant (Kassambara 2018).

References

- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- Canada Employment and Social Development. 2022. *2020 Annual Report - Occupational Injuries in the Canadian Federal Jurisdiction*. <https://www.canada.ca/en/employment-social-development/services/health-safety/reports/2020-injuries.html>.
- Kassambara, Alboukadel. 2018. *Regression Model Diagnostics*. <http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/>.
- Kellen, Dave. 2010. *Is There Any Reason to Prefer the AIC or BIC over the Other?* <https://stats.stackexchange.com/questions/577/is-there-any-reason-to-prefer-the-aic-or-bic-over-the-other>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Statistics Canada. 2023. *Open Government Data of Statistics Canada*. <https://www150.statcan.gc.ca/n1/en/type/data>.