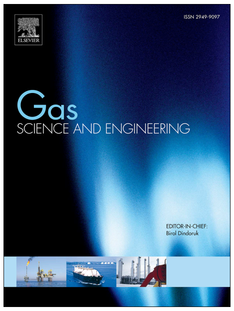


A statistical perspective on natural gas distribution pipeline incidents in the United States

Rui Xiao, Tarek Zayed, Mohamed A. Meguid, Laxmi Sushama

PII: S2949-9089(24)00320-0

DOI: https://doi.org/10.1016/j.jgsce.2024.205524

Reference: JGSE 205524

To appear in: *Gas Science and Engineering*

Received Date: 15 May 2024

Revised Date: 26 September 2024

Accepted Date: 10 December 2024

Please cite this article as: Xiao, R., Zayed, T., Meguid, M.A., Sushama, L., A statistical perspective on natural gas distribution pipeline incidents in the United States, *Gas Science and Engineering*, https://doi.org/10.1016/j.jgsce.2024.205524.

A statistical perspective on natural gas distribution pipeline incidents in the United States

Rui Xiao[a,b,*], Tarek Zayed[a], Mohamed A.Meguid[b], Laxmi Sushama[c]

[a] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[b] Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada

[c] Department of Civil Engineering, Trottier Institute for Sustainability in Engineering and Design, McGill University, Montreal, QC, Canada

Abstract

Natural gas distribution pipelines are essential for transporting natural gas from larger transmission pipelines to end users, including residential, commercial, and industrial facilities. The frequency of pipeline incidents and the potential for significant resulting losses have garnered considerable attention from stakeholders. This study conducts a statistical analysis of the incident characteristics of natural gas distribution systems in the U.S. Over the past thirty years, a general decline has been observed in both the number of incidents and the incident rate, reflecting similar downward trends in annual injuries and fatalities. However, the annual costs associated with incidents demonstrate minimal correlation with the total number of incidents, suggesting that multiple minor incidents may contribute substantially to overall losses. Pipelines constructed from different materials exhibit distinct failure patterns, particularly concerning the age of the pipelines. Statistical analysis reveals significant differences in the consequences of incidents, including injuries, fatalities, and total costs, across various pipeline attributes, such as pipe material, system type, incident location, cause, and type of incident. These findings highlight the necessity of incorporating these factors when modeling the consequences of natural gas distribution pipeline incidents. A detailed comparison of incident characteristics for distribution mains and service lines is also presented, utilizing a Zero-Inflated Poisson model to assess casualties associated with each pipeline incident. The results of this study offer valuable insights into the incident characteristics of natural gas distribution pipelines and can enhance safety and integrity management for these critical infrastructure systems.

Keywords

Gas distribution pipelines; Pipeline incidents; Incident consequences; Statistical tests.

---

[*] Corresponding author
Email address: rui.xiao@mcgill.ca

1. Introduction

Gas distribution systems are a vital component of the energy infrastructure in modern cities (Ma et al., 2013). Unlike gas transmission pipelines, which transport natural gas over long distances often spanning multiple states (Jo and Ahn, 2005), gas distribution pipelines are responsible for delivering natural gas from the transmission pipelines to end-users for various purposes, such as heating, power generation, and industrial processes, thereby ensuring the availability of a reliable and versatile energy source within specific urban regions (Cimellaro et al., 2015; Herrán-González et al., 2009). Gas distribution pipelines consist of distribution main lines and service lines. Distribution main lines are generally installed in underground utility easements alongside streets and highways, while distribution service lines run from the main lines into homes or businesses. In the U.S., natural gas distribution systems operate 2.2 million miles of mains and 1.6 million miles of service lines, serving over 72 million customers as of 2023 (PHMSA, 2023a). Given that gas distribution systems must be concentrated where residents live and work, ensuring their safety presents significant challenges for stakeholders due to the severe consequences that may arise from pipeline failures. According to incident reports, gas distribution systems are involved in significantly higher rates of fatalities and injuries compared to other types of pipelines, such as gas transmission, hazardous liquid, and liquefied natural gas pipelines, often resulting in explosions and evacuations, with 613 incidents, 23 fatalities, and 39 injuries reported in 2023 (PHMSA, 2023b). Therefore, it is essential to examine the detailed information on gas distribution pipeline incidents provided by operators to gain a better understanding of the hazards, their causes, and the circumstances surrounding them, and to develop more targeted strategies for managing and controlling the associated risks.

On March 12, 2014, an explosion occurred in Manhattan, New York, United States, resulting in the destruction of two multi-story, mixed-use buildings and leading to eight fatalities and at least 70 reported injuries (Wikimedia, 2014). The United States National Transportation Safety Board (NTSB) determined that the explosion was primarily caused by a defective fusion joint in the service tee of two gas pipes, allowing natural gas to leak from the main and migrate into the building, where it ignited (McGeehan, 2015). Experts warn that this incident exemplifies a troubling trend affecting modern cities, exposing the aging and hazardous nature of gas distribution infrastructure and the risks of natural gas leaks. These issues pose serious threats to urban economies and quality of life. In addition to the severe Manhattan incident, nearly 600 such gas distribution pipeline incidents have been reported annually over the past two decades. This includes the September 13, 2018 overpressure incident in the low-pressure gas distribution system in Lawrence, Massachusetts, which led to 1 fatality, 25 injuries, and damage to over 100 buildings, amounting to an estimated $1.59 billion in losses (Wikipedia, 2018). The frequency of these critical pipeline incidents has heightened public awareness regarding safety concerns. Accordingly, the relevant regulatory bodies have placed greater emphasis on enhancing pipeline safety measures, with the aim of mitigating the potential social risks associated with pipeline infrastructure.

The factors associated with pipeline incidents can be categorized into physical, environmental, and operational factors (Soomro et al., 2022; Zakikhani et al., 2020). The failure behavior of a gas pipeline system, as well as the consequences of incidents, is influenced by the interplay of these factors (Hocine et al., 2024; Soomro et al., 2022). It is acknowledged that the failure of gas pipelines is a complicated process, and the detailed mechanisms underlying each incident type require further in-depth analysis, including physical modeling of the specific degradation mechanisms involved (Rajani and Kleiner, 2001), experimental investigations (Wu et al., 2022), and the application of finite element method (FEM) approaches (Silva et al., 2007). Such methods can provide detailed, case-by-case analysis of the failure processes for individual pipeline incidents. However, they cannot easily provide general findings or insights applicable on a larger scale, particularly given the general inaccessibility of detailed information for each pipeline (Zamenian et al., 2017). On the other hand, statistical methods may offer a solution to provide potential insights into the characteristics of pipeline incidents based on the analysis of larger datasets of numerous pipeline incidents.

Several studies investigated the failure analysis, consequence modeling, and safety assessment of gas distribution pipelines. Simonoff et al. developed an approach to model the consequences and costs of gas transmission and distribution pipeline incidents, highlighting differences between pipeline types using PHMSA data (Restrepo et al., 2009). Siler-Evans et al. explored trends, causes, and consequences of natural gas pipeline incidents, finding decreasing fatalities and injuries over time, with a small number of incidents accounting for significant property damage (Siler-Evans et al., 2014). Hendrick et al. investigated gas emissions from leaks in cast iron distribution mains and found a positively skewed distribution, with 7% of leaks accounting for 50% of total emissions (Hendrick et al., 2016). Bianchini et al. examined U.S. gas distribution pipeline incidents from 2004 to 2015, finding an average of 2.09×10^{-5} accidents/km, with low-pressure and small-diameter systems accounting for most injuries and fatalities (Bianchini et al., 2018). Vetter et al. presented a comprehensive analysis of significant transmission and distribution pipeline incidents, categorizing circumstances and causes based on various factors (Vetter et al., 2019). Li et al. explored spatial and temporal patterns of correlations between natural gas distribution pipeline incident severity and contributing factors using GTWOLR, identifying several factors with significant spatiotemporally varying correlations (Li et al., 2021). Rahimi et al. employed geospatial analysis to assess urban gas pipeline risks, integrating population dynamics and building vulnerabilities to create risk maps that inform decision-making for pipeline management and safety measures (Rahimi et al., 2024). Shen and Zhou analyzed onshore oil and gas pipeline statistics from Canada and the US, comparing incident rates and failure causes to identify integrity threats and enhance maintenance prioritization for pipeline safety (Shen and Zhou, 2024). While these studies have examined various aspects of gas pipeline incidents, there is a notable lack of comprehensive research that offers a holistic understanding of the underlying factors contributing to these incidents and their consequences (Ramírez-Camacho et al., 2017; Xiao et al., 2024).

The investigations into the failure behavior of gas transmission pipeline incidents have been extensively examined (Lam and Zhou, 2016; Siler-Evans et al., 2014; Wang et al., 2022; Xiao et al., 2023; Zerouali et al., 2024), but studies regarding gas distribution pipeline incidents are more limited, as evident from the previous descriptions. Given the distinct characteristics of gas transmission and gas distribution pipelines, such as pipe diameter, operating pressure, and surrounding environment, it is essential to conduct separate studies for a comprehensive analysis of incident characteristics and failure behaviors, including incident causes, failure components, failure types, and consequences. However, such detailed investigations focusing specifically on gas distribution pipeline incidents have been limited in the literature. Furthermore, there is a lack of understanding regarding the differences in typical pipeline attributes, leading to uncertainty about whether these factors should be incorporated into the modeling of pipeline incidents. Developing an initial model to estimate the potential number of casualties in gas distribution pipeline incidents would be beneficial for stakeholders, enabling them to make informed decisions regarding pipeline layout, maintenance, and integrity management (Teng et al., 2021; Xu et al., 2023). The limited factor analysis and understanding of gas distribution pipeline failures hinder the scientific validity and accuracy of such predictive modeling efforts.

This study examined gas distribution pipeline incidents in the U.S. using data from the PHMSA. The primary objectives were to analyze the characteristics of these incidents, investigate variations across different states, and assess key factors such as pipe materials, pipe types, incident causes, and incident types. These analyses provide a foundation for potential modeling of incident consequences in gas distribution systems. Statistically significant differences were identified among these attributes. The study also compared the characteristics of incidents occurring in distribution mains versus service lines. Predictive models were developed to estimate the potential number of casualties resulting from an incident. The findings offer valuable insights for stakeholders to better understand pipeline failures, informing strategies to prevent future incidents, reduce risks, and enhance the safety of gas distribution systems.

2. Methodology

2.1 Data sources and description

As defined in the U.S. Code of Federal Regulations (49 C.F.R. § 191.3) (Office of the Federal Register, 2023), PHMSA has collected gas distribution pipeline incident reports since 1970, maintaining thousands of incident records. This study will utilize this dataset to present a statistical investigation into the incident characteristics regarding various pipeline attributes. PHMSA categorizes the gas distribution pipeline incidents into four different files based on the incident year: 1970–1986, 1986–2004, 2004–2009, and 2010 to present (PHMSA, 2023c). This is due to changes in reporting regulations and incident report formats over the years (PHMSA, 2017). Since the incidents that occurred between 1970 and 1986 included less information than the more recent incident records, these incidents are excluded from the subsequent analysis in this study.

The incident files from 1986 to the present consist of a total of 4,774 incidents. In contrast to gas transmission pipeline incidents, which are primarily made from steel (Lam and Zhou, 2016; Xiao et al., 2023), gas distribution pipeline incidents generally involve plastic, steel, and iron materials. Considering the functions, geographic locations, and regulatory factors, the proportion of plastic pipelines has been increasing and has become the main material used in gas distribution pipelines (Bachir-Bey and Belhaneche-Bensemra, 2020; Khademi-Zahedi, 2019). The attributes shared across the different incident report file periods, such as pipe diameter, material, fatalities, and injuries, were selected for further investigation. After carefully reviewing the attributes of the different files manually, sixteen attributes were selected for incidents that occurred from 1986 to 2004, including incident State, year, month, pipe material, diameter, thickness, operating pressure, installed year, location class, system incident occurred (main, service line, etc.), incident area (above ground, underground, etc.), part failure occurred (pipe body, joint, etc.), incident cause, fatalities, injuries, and total cost. It is important to note that the composition of natural gas may also influence pipeline incidents. However, due to the unavailability of relevant information, this aspect is not analyzed in the current study and warrants further investigation in future research. The pipe age was further calculated by subtracting the installed year from the incident year, while the incident type of the involved pipelines was added to the dataset for incidents since 2004. The annual incident rate was also calculated to quantify the number of incidents relative to the total length of gas distribution pipelines, as given by Shan et al. (2018)

$$f = N/L \tag{1}$$

where f is the incident rate, N is the number of incidents that occurred within a given time duration, typically one year, relative to the total length L of gas distribution pipelines, which is obtained from the gas distribution annual data report provided by PHMSA (PHMSA, 2023a).

2.2 Statistical tests

In this study, the Mann-Kendall test was used to identify any monotonic trends in the annual number of reported gas distribution pipeline incidents (Mann, 1945). Additionally, the Pearson correlation coefficient was calculated to determine the correlations, e.g., between the number of gas distribution pipeline incidents and the total mileage of existing service gas distribution pipelines in each state (Schober et al., 2018). When comparing incident consequences (fatalities, injuries, and cost) among multiple groups (e.g., various pipe materials, incident cause, etc.), the non-parametric Kruskal-Wallis H test was used as an omnibus test to identify statistically significant differences among the groups (Kruskal and Wallis, 1952). If a significant difference was detected, the post-hoc Dunn's test was performed to compare multiple groups in a pairwise manner and determine if there were significant differences between each pair (Dunn, 1964).

Bonferroni's correction was used to adjust the significance level for each individual comparison (Miller, 1966). This study assumed a significance level of 0.05 for all statistical tests.

2.3 Zero-Inflated Poisson model

The Poisson regression model is widely used for analyzing count data, particularly when the data follows a Poisson distribution. It is ideal for modeling event counts, rates, and non-negative integer outcomes, as it handles scenarios where traditional linear regression fails due to violations of normality and homoscedasticity (Nelder, 1974). The model assumes that the logarithm of the expected outcome value μ, where $\mathrm{E}(Y) = \mu$, can be expressed as a linear combination of independent variables x, the intercept α, and the regression coefficients $\boldsymbol{\beta}$. Mathematically, this is represented as (Coxe et al., 2009)

$$\log(\mu) = \alpha + \boldsymbol{\beta x} \tag{2}$$

Given a Poisson regression model and an input vector $\boldsymbol{x}$, the expected value of Y is a multiplicative function of $\boldsymbol{x}$ as given by

$$\mu = e^{\alpha+\boldsymbol{\beta x}} = e^{\alpha} e^{\boldsymbol{\beta x}} \tag{3}$$

However, the Poisson model assumes that the variance equals the mean, which may not hold in cases of overdispersion, particularly when there are an excessive number of zeros in the data.

To address this issue, the Zero-Inflated Poisson model is introduced. The ZIP model accounts for overdispersion and excess zeros by assuming two underlying processes (Lambert, 1992). The first process generates zeros with probability π_i, while the second process follows a standard Poisson distribution with mean λ_i. The probability mass function of the ZIP model is defined as (Hall, 2000)

$$P(Y_i = 0) = \pi_i + (1-\pi_i)e^{-\lambda_i}, \qquad P(Y_i = k) = (1-\pi_i)\frac{\lambda_i^k e^{-\lambda_i}}{k!}, k = 1,2,\cdots \tag{4}$$

The parameters λ_i and π_i are linked to covariates through the log and logit link functions, respectively

$$log(\lambda_i) = \boldsymbol{X_i \beta}, \qquad logit(\pi_i) = \boldsymbol{Z_i \gamma} \tag{5}$$

where$\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ represent covariate matrices, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the respective parameter vectors. The ZIP model accounts for overdispersion by allowing the variance to exceed the mean, expressed as

$$E(Y_i) = (1-\pi_i)\lambda_i, \qquad Var(Y_i) = (1-\pi_i)\lambda_i(1+\pi_i\lambda_i) \tag{6}$$

Like Poisson models, ZIP models use maximum likelihood estimation (MLE) for parameter estimation, and hypothesis testing can be employed to detect zero inflation by comparing the ZIP model to simpler Poisson models.

The developed models will be compared against the observed values, and various commonly used quantitative metrics will be employed to evaluate the performance of each model, including

1) Mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{7}$$

2) Mean square error (MSE)

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{8}$$

3) Coefficient of determination *R*-squared (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(\hat{y}_i - \bar{y}_i)^2} \tag{9}$$

where $\hat{y}_i$ and y_i are the predicted and actual values, respectively, $\bar{y}_i$ is the average value of the actual values and N is the number of samples.

3. Results and discussion

3.1 Incident overview

Fig. 1 illustrates the annual incident number and corresponding incident rate of gas distribution pipelines in the U.S., starting from 1986 to the present. Both curves exhibit similar decreasing trends over time. This can be attributed to the fact that the construction of gas pipelines in the U.S. experienced significant growth primarily in the 1950s and 1960s, and has remained relatively stable over the past three decades. The Mann-Kendall test revealed a statistically significant moderate downward trend ($\tau = -0.400$, $p = 3.57\times10^{-4}$), confirming that the observed declines are unlikely to be caused by random variability. Exceptions occurred between 2004-2009 and in 2019, where the incident numbers were higher. However, in recent years, the number of incidents has fallen to a historic low of less than 80 per year. This downward trend is likely due to improved safety regulations and enforcement, pipeline replacement and upgrade programs, implementation of integrity management programs, and enhanced maintenance and inspection practices.

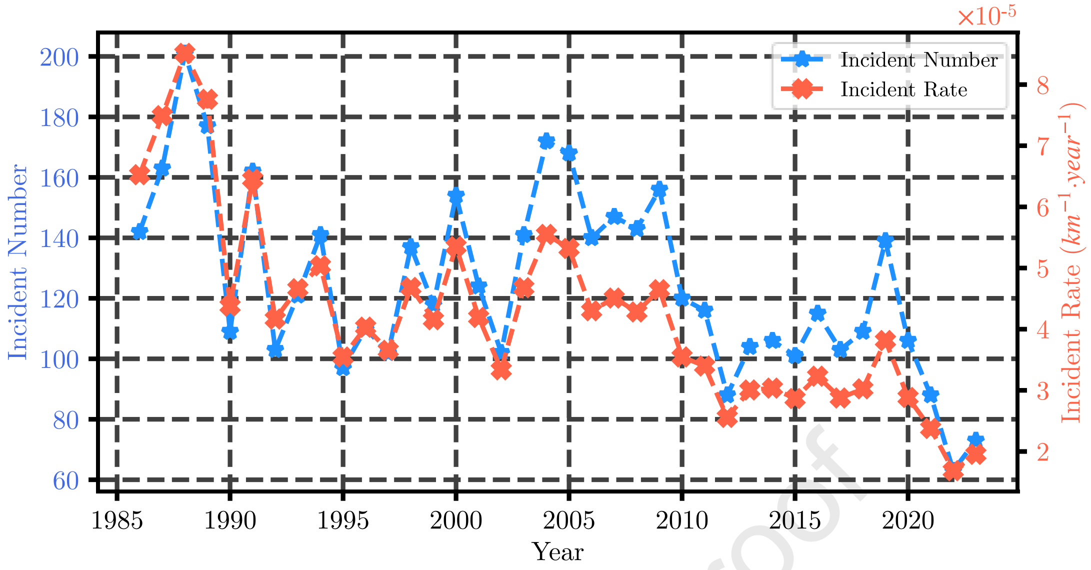


Figure 1. Annual incident number and incident rate in US gas distribution pipelines.

Fig. 2 portrays the annual injuries, fatalities, and total costs involved in gas distribution pipeline incidents. The annual injuries appear to follow a similar trend to the annual incident numbers, with notable surges in 2014, 2016, and 2018. Meanwhile, the annual fatalities exhibit a general decreasing pattern but reached their highest value in 2023 over the past two decades. The Mann-Kendall test results suggest statistically significant moderate downward trends for both injuries ($\tau = -0.588$, $p = 1.50\times10^{-7}$) and fatalities ($\tau = -0.378$, $p = 8.52\times10^{-4}$). In contrast, the total costs present a statistically significant moderate upward trend ($\tau = 0.371$, $p = 8.79\times10^{-4}$), which is a curious result given the decreasing trend of annual incident numbers. The incidents that occurred in 2005 (Nairn, LA) and 2018 (Lawrence, MA) contributed the most to the total losses, suggesting a small number of incidents can have a catastrophic impact on overall costs. Further analysis explored the correlation between annual incident numbers and consequences. The results indicate a statistically significant moderate positive correlation between the annual injuries ($\rho = 0.514$, $p = 8.13\times10^{-4}$) and fatalities ($\rho = 0.611$, $p = 3.57\times10^{-5}$) with incident numbers. However, there is nearly no statistically significant correlation between annual total costs and incident numbers ($\rho = 0.045$, $p = 0.781$), indicating a more complex relationship that warrants further investigation.

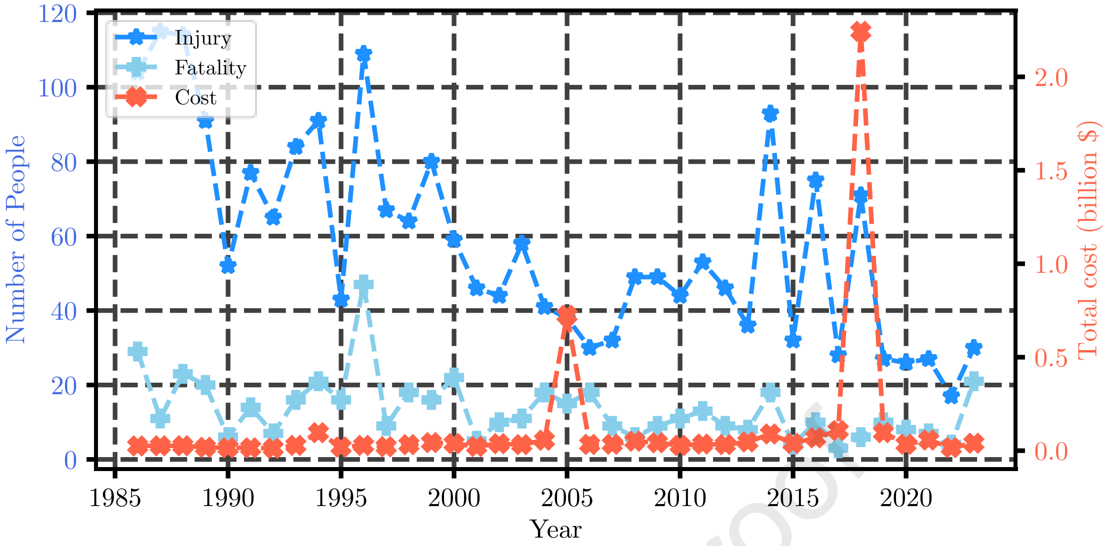


Figure 2. Annual injuries, fatalities and costs involved in incidents in US gas distribution pipelines.

As stated previously, the pipe materials used in U.S. gas distribution pipelines primarily consist of steel, plastic, iron, and other materials. **Fig. 3** further displays the annual incident numbers by pipe material. The incident numbers involving steel pipelines have consistently remained at a higher level compared to other pipe materials. Given that plastic, specifically polyethylene (PE), has become the dominant material for gas distribution pipelines (approximately 1.327 million kilometers (PHMSA, 2023a)), while steel is the second most common material (about 0.825 million kilometers (PHMSA, 2023a)), steel pipelines pose a higher risk than other pipe materials. This suggests the need for enhanced regular inspection, maintenance, replacement, and upgrade programs for steel pipelines. Plastic pipelines account for the second-highest number of incidents, with a noticeable increase in incidents between 1998 and 2007. In contrast, iron pipelines exhibit a statistically significant gradual decreasing trend in incident numbers ($\tau = 0.649$, $p = 1.39\times10^{-8}$), likely due to the reduced use of iron in gas distribution systems over time in the U.S. Incidents involving pipelines made of other materials display no discernible pattern but did reach a peak in 2010.

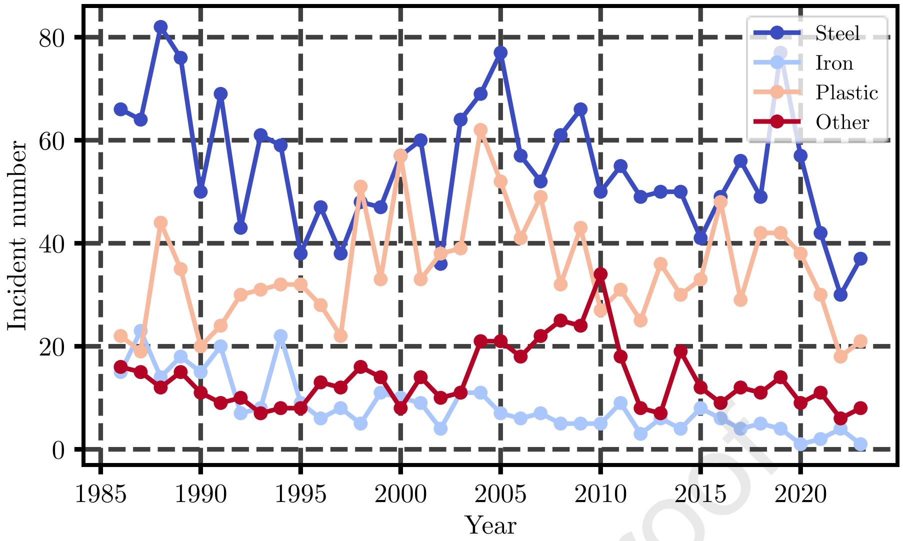


Figure 3. Annual incident number by pipe material in US gas distribution pipelines.

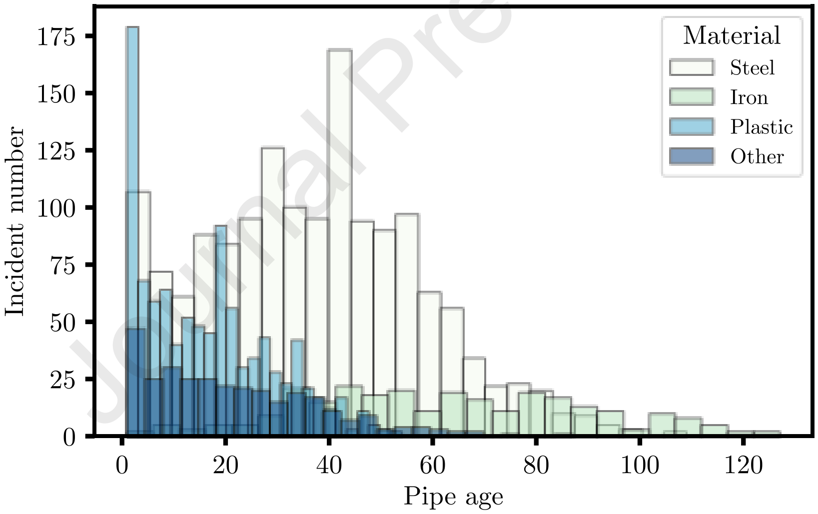


Figure 4. Age distribution of failed gas distribution pipelines by pipe material in the US.

Fig. 4 provides additional insights on the age of the failed pipelines by material. The data demonstrates that the failed steel pipelines had a longer average lifespan ($\mu_{\text{age}} \approx 36.8$ years) compared to plastic and other pipeline materials. However, a certain number of incidents also occurred early in the lifespan of steel pipelines, likely due to manufacturing defects or installation issues. The failed plastic pipelines exhibited a distinct pattern, where the service life was predominantly within 20 years ($\mu_{\text{age}} \approx 16.6$ years), with the highest incident frequency occurring within the first two years after installation. This evidence highlights the importance of improved installation practices and guidelines, as well as regular inspections, to enhance the performance, reliability, and safety of plastic gas pipelines. Interestingly, the failed iron pipelines had the longest average lifespan ($\mu_{age} \approx 62.4$ years) compared to other materials. This can be attributed to the fact that iron pipelines were commonly used when natural gas distribution systems were first established, leading to a longer service life before eventual failure.

Previous studies have demonstrated that climate can contribute to the failure of gas pipelines (Fan et al., 2022; Zakikhani et al., 2021). This study investigates the effect of temperature on gas distribution pipeline failures as a form of incident by month. **Fig. 5** presents a histogram of the incidents by month, which shows that gas distribution pipelines experienced fewer incidents from April through June, followed by a slight increase during the summer months. However, the highest frequency of gas pipeline incidents occurred during the winter season, particularly in January. This evidence suggests that both high and low temperatures may contribute to the failure of gas distribution pipelines, potentially related to the mechanical properties of the pipelines and the behavior of joints under such thermal conditions. Nonetheless, further research is needed to fully understand the impact of climate, especially in regions that endure extended periods of extreme hot or cold temperatures, such as Saudi Arabia and Canada.

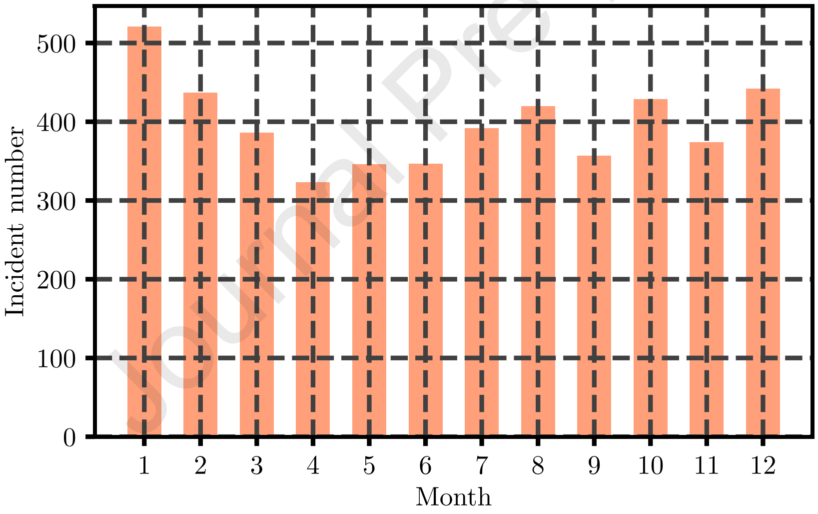
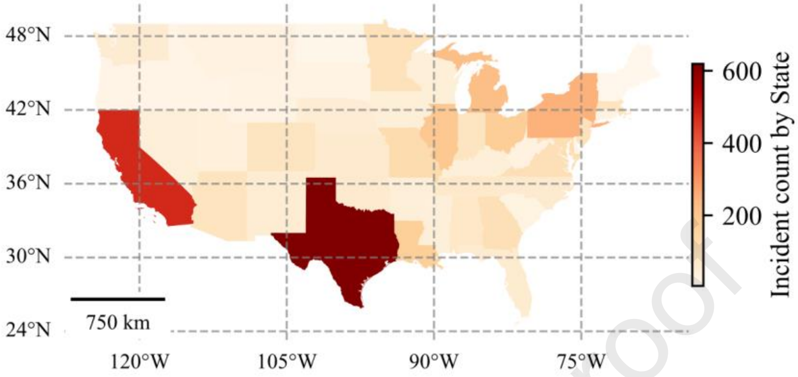
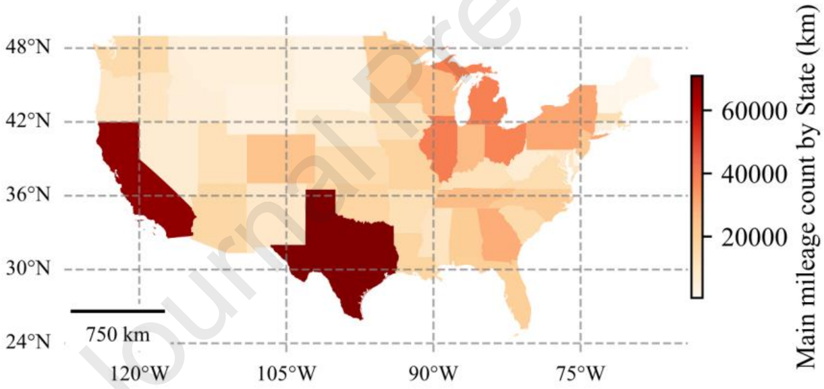


Figure 5. Histogram of gas distribution pipeline incidents by month in the US.

The subsequent analysis examines the incident characteristics at the state level. **Fig. 6** presents the incident numbers and total gas distribution pipeline mileage for each U.S. state. Texas and California experienced the highest number of pipeline incidents, representing 13.6% and 10.0% of the total incident numbers, respectively. Notably, these two states also have the longest pipeline systems, accounting for 8.42% and 8.04% of the national total, respectively. This observation suggests a potential positive correlation between incident numbers and the overall length of gas distribution pipelines in each state. This correlation is further explored in **Fig. 7**. The correlation coefficient value of 0.905 indicates a very strong positive correlation between incident numbers and total pipeline length. Additionally, the p-value of 3.46×10^{-20} suggests this correlation is statistically significant, making it extremely unlikely to have occurred by chance. These results imply that pipeline operators in states with more extensive distribution systems should be given increased attention, and more resources should be allocated to strengthening safety regulations and protocols in these areas.



(a)



(b)

Figure 6. Geographic distribution of (a) incident numbers, and (b) pipe length, by state in US gas distribution pipelines.

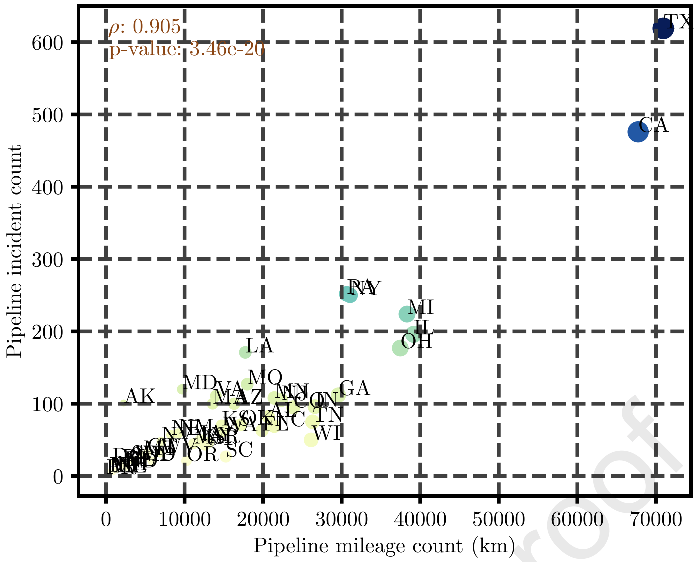
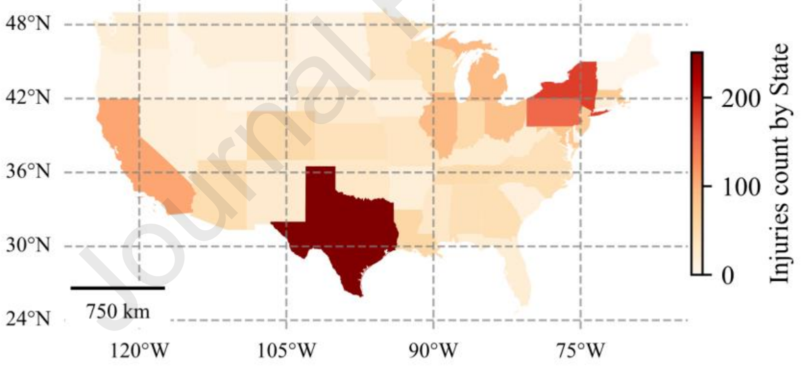
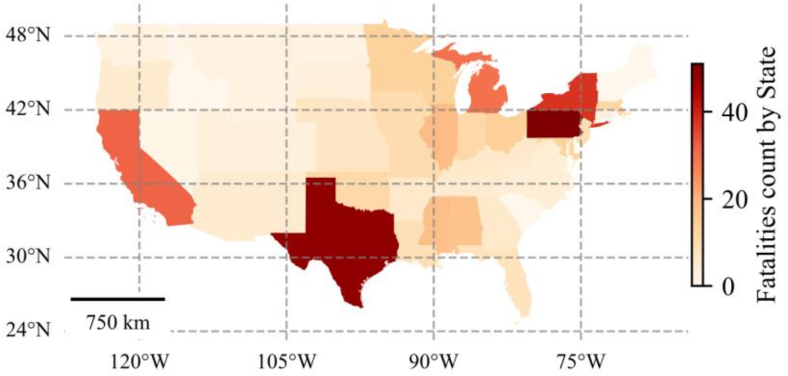


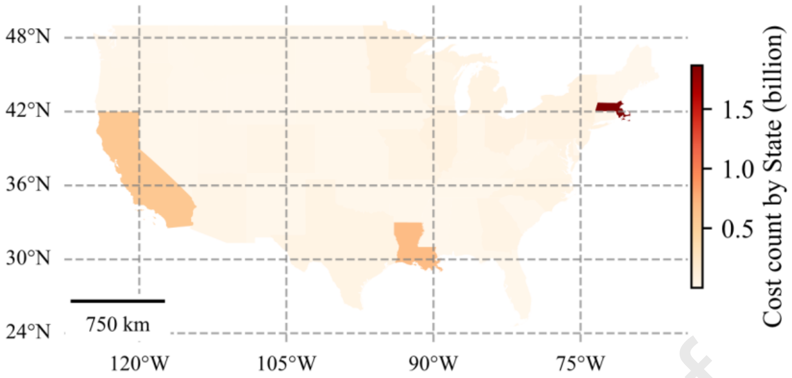
Figure 7. Correlation of incident numbers and total pipeline length by state in US gas distribution infrastructure.



(a)



(b)



(c)

Figure 8. Geographic distribution of (a) injuries, and (b) fatalities, (c) total costs, by state involved in US gas distribution pipelines.

The consequences of gas distribution pipeline incidents, including injuries, fatalities, and total costs, are further analyzed at the state level in **Fig. 8**. The findings are similar to those presented in **Fig. 2**, which demonstrated a statistically significant, strong positive correlation between incident numbers and both injuries (ρ = 0.894, p = 4.67×10^{-19}) and fatalities (ρ = 0.775, p = 1.54×10^{-11}) at the state level. Texas and New York reported the highest numbers of injuries, while Texas and Pennsylvania had the highest numbers of fatalities. In contrast, the correlation between total costs and incident numbers was weakly positive and not statistically significant (ρ = 0.229, p = 0.102), with Massachusetts incurring the most substantial costs from pipeline incidents. These findings highlight the need for each state to enhance the safety and integrity management of gas distribution pipelines to reduce the potential consequences to society.

3.2 Attributes-focused statistical analysis

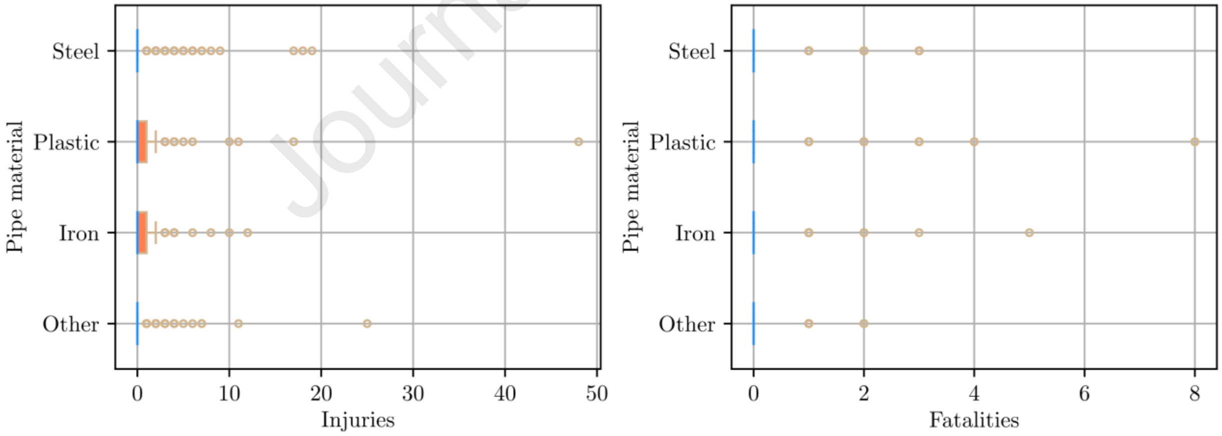
This section statistically analyzes the consequences of gas distribution pipeline incidents in relation to various pipeline attributes, including pipe material, involved system, incident part, incident cause, and incident type.

3.2.1 Pipe material

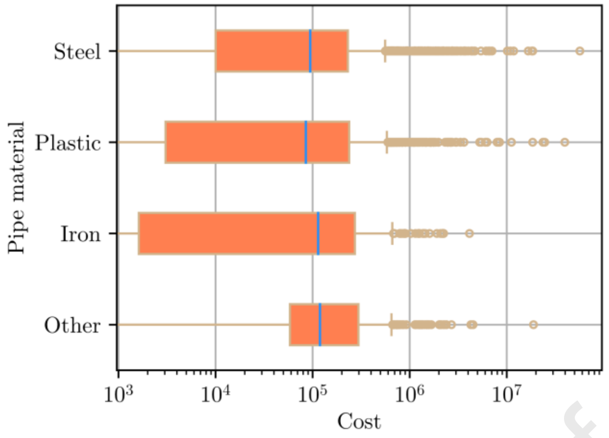
Table 1 provides a summary of the statistics for injuries, fatalities, and costs associated with gas distribution pipeline incidents, broken down by pipe material (steel, plastic, iron, and other). These findings are further visualized using box plots in **Fig. 9**. Iron pipes had the highest mean number of injuries (0.674) and fatalities (0.122) per incident, despite having fewer total incidents. Across all materials, the median number of injuries and fatalities was zero, indicating that many incidents resulted in no injuries or fatalities. However, the maximum values reveal several incidents with exceptionally high counts of injuries and fatalities as outliers. Plastic pipes had the highest mean cost per incident ($330,274), followed by steel and iron pipes. The maximum total costs were extremely high across all materials, reflecting the occurrence of a few exceptionally costly outlier incidents.

Table 1. Statistical summary of incident consequences in US gas distribution pipelines by pipe material (1986 - present).

Pipe material	count	mean	std	min	25%	50%	75%	max
					Injury			
Steel	2083	0.385	1.068	0	0	0	0	19
Plastic	1323	0.485	1.626	0	0	0	1	48
Iron	319	0.674	1.274	0	0	0	1	12
Other	519	0.405	1.467	0	0	0	0	25
					Fatality			
Steel	2083	0.078	0.335	0	0	0	0	3
Plastic	1323	0.086	0.467	0	0	0	0	8
Iron	319	0.122	0.456	0	0	0	0	5
Other	519	0.092	0.328	0	0	0	0	2
					Cost			
Steel	2083	316816	1565261	0	10035	94536	230824	56678788
Plastic	1323	330274	1660692	0	3057	85177	238136	39641121
Iron	319	245517	436921	0	1621	113983	272827	4130846
Other	519	292571	928994	0	58486	119149	296585	18867225



(a) (b)



(c)

Figure 9. Box plots illustrating (a) injuries, (b) fatalities, and (c) total costs, from incidents by pipe material in US gas distribution pipelines.

The Kruskal-Wallis H test results indicate that injuries ($H = 67.15$, $p = 1.73\times10^{-14}$), fatalities ($H = 8.60$, $p = 0.035$), and total costs ($H = 18.79$, $p = 3.00\times10^{-4}$) from gas distribution pipeline incidents differ significantly across the different pipe material groups. The specific group differences were further determined using post-hoc Dunn's tests with appropriate adjustments, as shown in **Table 2**. The results reveal significant differences in injuries between steel and iron, as well as iron and other materials. Plastic pipes also exhibited significant differences in injuries compared to steel and iron. In contrast, no significant differences were found in fatalities between steel, plastic, and iron, although iron showed marginally significant differences with steel and plastic. This divergence from the significant overall group differences detected by the Kruskal-Wallis H test. This discrepancy suggests the actual inter-group differences are relatively small, despite the Kruskal-Wallis H test's sensitivity to minor variations. The cost analysis revealed no significant differences between steel, plastic, and iron, but the other pipe material category had a significant difference compared to the other three pipe material categories. Given the findings across multiple consequences, it would be prudent to include pipe material as a factor in comprehensive modeling efforts aimed at understanding and predicting the potential consequences, especially injuries, of pipeline incidents.

Table 2. Pairwise comparisons for incident consequences in US gas distribution pipelines by pipe material (1986 - present).

Pipe material	Steel	Plastic	Iron	Other
		Injury		
Steel	1	2.05×10^{-5}	8.57×10^{-11}	**1**
Plastic	2.05×10^{-5}	1	5.92×10^{-4}	6.33×10^{-5}
Iron	8.57×10^{-11}	5.92×10^{-4}	1	2.14×10^{-10}

Other	**1**	6.33×10^{-5}	2.14×10^{-10}	1
Fatality				
Steel	1	**1**	**0.169**	**0.671**
Plastic	**1**	1	**0.0846**	**0.335**
Iron	**0.169**	**0.0846**	1	**1**
Other	**0.671**	**0.335**	**1**	1
Cost				
Steel	1	**1**	**1**	2.31×10^{-3}
Plastic	**1**	1	1	1.02×10^{-4}
Iron	**1**	**1**	**1**	**0.217**
Other	2.31×10^{-3}	1.02×10^{-4}	**0.217**	1

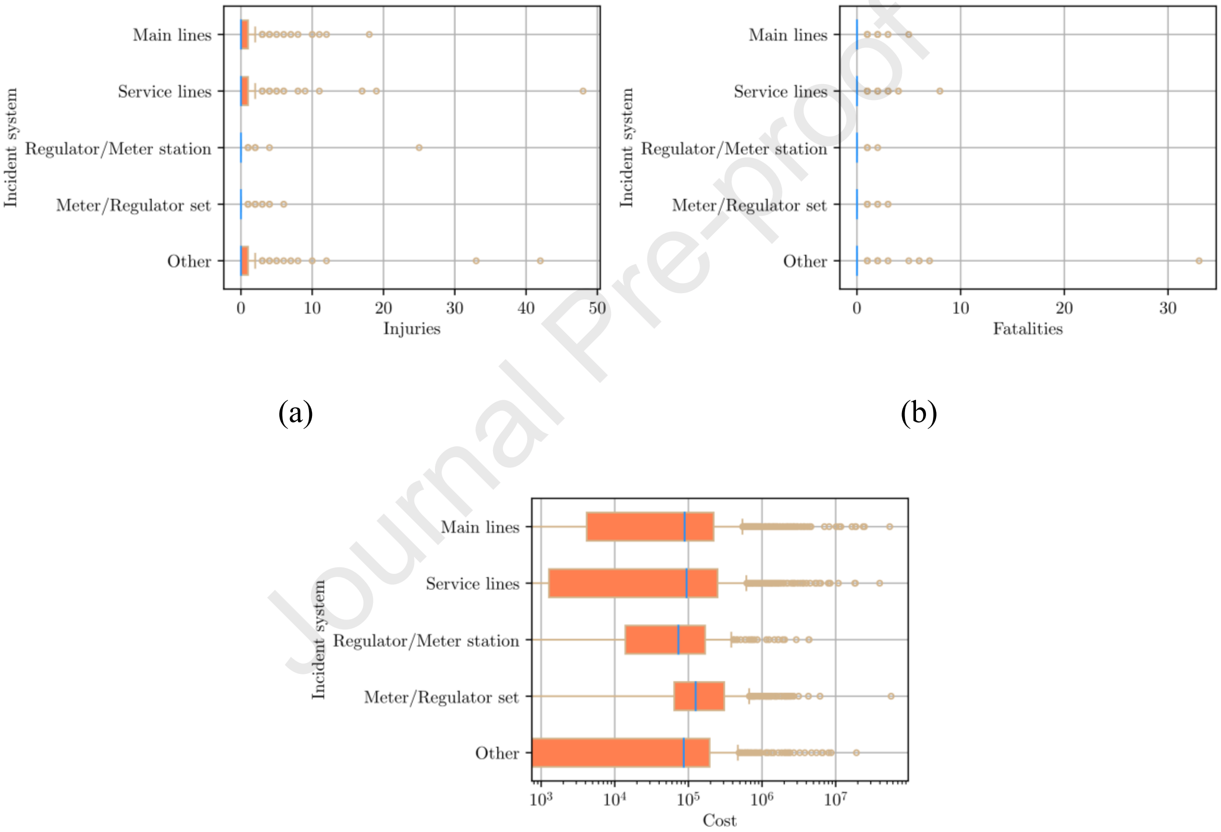
3.2.2 Involved system

The statistics for injuries, fatalities, and costs associated with gas distribution pipeline incidents, broken down by the involved system (main lines, service lines, regulator/meter station, regulator/meter set, and other), are summarized in **Table 3** and visualized in **Fig. 10**. The results show that service lines had the highest mean number of injuries (0.603) and fatalities (0.151) per incident, followed by the other category. In contrast, main lines, regulator/meter stations, and meter/regulator sets had lower mean injuries and fatalities. The median number of injuries and fatalities was 0, suggesting that many incidents resulted in no injuries or fatalities, but the maximum values showed some incidents with high injury and fatality counts. Regarding the economic impact, service lines had the highest mean cost per incident ($353,635), followed by meter/regulator sets and main lines. The median and 75th percentile costs were highest for main lines, and the maximum costs were extremely high across all system components, highlighting the substantial financial consequences of pipeline incidents.

Table 3. Statistical summary of incident consequences in US gas distribution pipelines by the involved system (1986 - present).

Involved system	count	mean	std	min	25%	50%	75%	max
Injury								
Main lines	1946	0.437	1.012	0	0	0	1	18
Service lines	1158	0.603	1.932	0	0	0	1	48
Regulator/Meter station	174	0.328	1.954	0	0	0	0	25
Meter/Regulator set	91	0.196	0.601	0	0	0	0	6
Other	605	0.709	2.454	0	0	0	1	42
Fatality								
Main lines	1946	0.049	0.267	0	0	0	0	5
Service lines	1158	0.151	0.578	0	0	0	0	8
Regulator/Meter station	174	0.069	0.276	0	0	0	0	2

Meter/Regulator set	91	0.073	0.308	0	0	0	0	3
Other	605	0269	1.530	0	0	0	0	33
					Cost			
Main lines	1946	316772	1729066	0	4161	88888	220159	5437986
Service lines	1158	353635	1578156	0	1277	94382	249068	39641121
Regulator/Meter station	174	249878	595399	0	13938	73239	168956	4361205
Meter/Regulator set	91	339523	1947484	0	64322	125709	306771	56678788
Other	605	295529	1090207	0	0	86598	194324	19149105



(a)

(b)

(c)

Figure 10. Box plots illustrating (a) injuries, (b) fatalities, and (c) total costs, from incidents by involved systems in US gas distribution pipelines.

The Kruskal-Wallis test results indicate statistically significant differences in the distributions of injuries ($H = 139.60$, $p = 3.44\times10^{-29}$), fatalities ($H = 78.72$, $p = 3.24\times10^{-16}$), and total costs ($H = 63.01$, $p = 6.74\times10^{-13}$) across the different incident system components being compared. The post-hoc Dunn's test results, listed in **Table 4**, indicate significant differences between various systems in terms of injury, fatality, and cost. For injuries, significant differences were identified between main lines and other systems, as well as between service lines and regulator/meter stations. Similarly, for fatalities, significant differences were found between main lines and service lines, as well as between service lines and meter/regulator sets. In contrast, fewer significant differences were observed for costs, with the main differences occurring between meter/regulator sets and other systems. These findings suggest that the incident system component should be incorporated as a variable in pipeline consequence modeling, to better capture the complexities of incident outcomes and associated consequences.

Table 4. Pairwise comparisons for incident consequences in US gas distribution pipelines by the involved system (1986 - present).

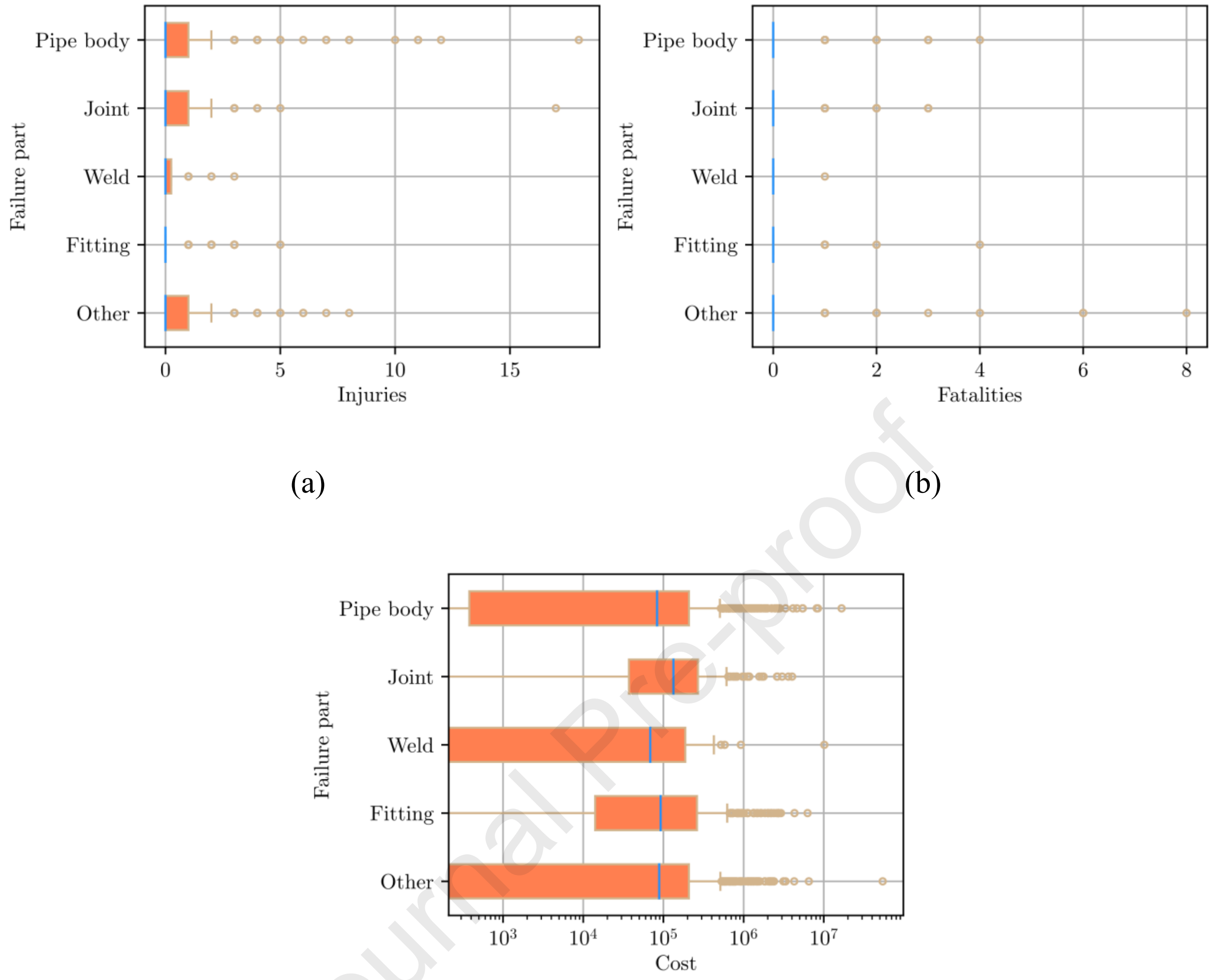
Involved system	Main lines	Service lines	Regulator/Meter station	Meter/Regulator set	Other
			Injury		
Main lines	1	0.0179	8.38×10^{-4}	9.53×10^{-16}	0.0450
Service lines	0.0179	1	1.49×10^{-6}	3.51×10^{-23}	**1**
Regulator/Meter station	8.38×10^{-4}	1.49×10^{-6}	1	**1**	2.54×10^{-6}
Meter/Regulator set	9.53×10^{-16}	3.51×10^{-23}	**1**	1	6.06×10^{-18}
Other	0.0450	**1**	2.54×10^{-6}	6.06×10^{-18}	1
			Fatality		
Main lines	1	4.83×10^{-9}	**1**	**0.436**	9.99×10^{-14}
Service lines	4.83×10^{-9}	1	**0.753**	8.02×10^{-3}	**0.0998**
Regulator/Meter station	**1**	**0.753**	1	**1**	0.0145
Meter/Regulator set	**0.436**	8.02×10^{-3}	**1**	1	1.23×10^{-6}
Other	9.99×10^{-14}	**0.0998**	0.0145	1.23×10^{-6}	1
			Cost		
	Main lines	Service lines	Regulator/Meter station	Meter/Regulator set	Other
Main lines	1	**1**	**1**	2.06×10^{-11}	**1**
Service lines	**1**	1	**1**	8.65×10^{-7}	**0.220**
Regulator/Meter station	**1**	**1**	1	3.13×10^{-3}	**1**
Meter/Regulator set	2.06×10^{-11}	8.65×10^{-7}	3.13×10^{-3}	1	1.95×10^{-10}
Other	**1**	**0.220**	**1**	1.95×10^{-10}	1

3.2.3 Incident part

The incident part refers to the specific component of the pipeline where the failure occurred, including the pipe body, joint, weld, fitting, and other components. The probability of failure may vary depending on the incident part, as failures originating from welds or fittings may have higher probabilities compared to failures in the pipe body, due to the complexity of these components and potential fabrication or installation issues. The statistics for injuries, fatalities, and costs associated with gas distribution pipeline incidents, broken down by incident part, are summarized in **Table 5** and visualized in **Fig. 11**. The analysis of pipeline incident consequences reveals distinct patterns across different incident parts. For injuries, pipe body and joint incidents have the highest mean and maximum values, while welds and fittings exhibit the lowest. Fatalities show a similar trend, with joint incidents having the highest mean. Regarding costs, weld incidents incur the highest mean, followed by fittings, joints, and other parts, while pipe body incidents have the lowest mean. The positively skewed distributions for all metrics indicate a small number of outliers with significantly higher consequences. These findings highlight the importance of considering the specific incident part when assessing and mitigating the risks associated with pipeline operations.

Table 5. Statistical summary of incident consequences in US gas distribution pipelines by incident part (1986 - present).

Incident part	count	mean	std	min	25%	50%	75%	max
Injury								
Pipe body	1287	0.477	1.142	0	0	0	1	18
Joint	306	0.477	1.143	0	0	0	1	17
Weld	32	0.375	0.751	0	0	0	0.25	3
Fitting	369	0.377	0.781	0	0	0	0	5
Other	1032	0.468	0.910	0	0	0	1	8
Fatality								
Pipe body	1287	0.085	0.364	0	0	0	0	4
Joint	306	0.163	0.512	0	0	0	0	3
Weld	32	0.031	0.177	0	0	0	0	1
Fitting	369	0.062	0.320	0	0	0	0	4
Other	1032	0.120	0.509	0	0	0	0	8
Cost								
Pipe body	1287	232982	713996	0	380	83280	208132	16656188
Joint	306	254021	467812	0	37125	133259	269431	4020405
Weld	32	454609	1780632	0	0	68544	187263	10143124
Fitting	369	287758	602496	0	14090	92366	262195	6273029
Other	1032	255882	1733698	0	0	88521	207646	54237986



(a) (b)

(c)

Figure 11. Box plots illustrating (a) injuries, (b) fatalities, and (c) total costs, from incidents by incident part in US gas distribution pipelines.

The group differences in the consequences for different incident parts were analyzed using the Kruskal-Wallis test. The results indicate no significant difference in the distributions of injuries ($H = 4.65$, $p = 0.325$) across incident parts. However, there were statistically significant differences in the distributions of fatalities ($H = 15.86$, $p = 3.21\times10^{-3}$) and total costs ($H = 24.20$, $p = 7.29\times10^{-5}$). Subsequent multiple comparisons were conducted, and the results are summarized in **Table 6**. Regarding fatalities, the analysis revealed statistically significant differences between pipe body and joint incidents, as well as between joint and fitting incidents. However, no significant differences were found between pipe body and weld, fitting, or other incident parts, nor between joint and weld or other. In terms of costs, the findings indicated a significant difference between pipe body and joint incidents, as well as between joint and other incidents, but not between pipe body and fitting incidents. No other pairwise comparisons showed significant differences in either fatality or cost outcomes. These findings suggest that the incident part is an important factor in understanding pipeline incident consequences, as different components of a gas pipeline are susceptible to different failure modes, such as corrosion, material defects, and mechanical damage.

Table 6. Pairwise comparisons for incident consequences in US gas distribution pipelines by incident part (1986 - present).

Incident part	Pipe body	Joint	Weld	Fitting	Other
			Injury		
Pipe body					
Joint					
Weld					
Fitting					
Other					
			Fatality		
Pipe body	1	0.0255	**1**	**1**	**0.418**
Joint	0.0255	1	**0.847**	0.0109	**1**
Weld	**1**	**0.847**	1	**1**	**1**
Fitting	**1**	0.0109	**1**	1	**0.163**
Other	**0.418**	**1**	**1**	**0.163**	1
			Cost		
Pipe body	1	4.35×10^{-4}	**1**	**0.0812**	**1**
Joint	4.35×10^{-4}	1	**0.441**	**1**	6.58×10^{-4}
Weld	**1**	**0.441**	1	**1**	**1**
Fitting	**0.0812**	**1**	**1**	1	**0.101**
Other	**1**	6.58×10^{-4}	**1**	**0.101**	1

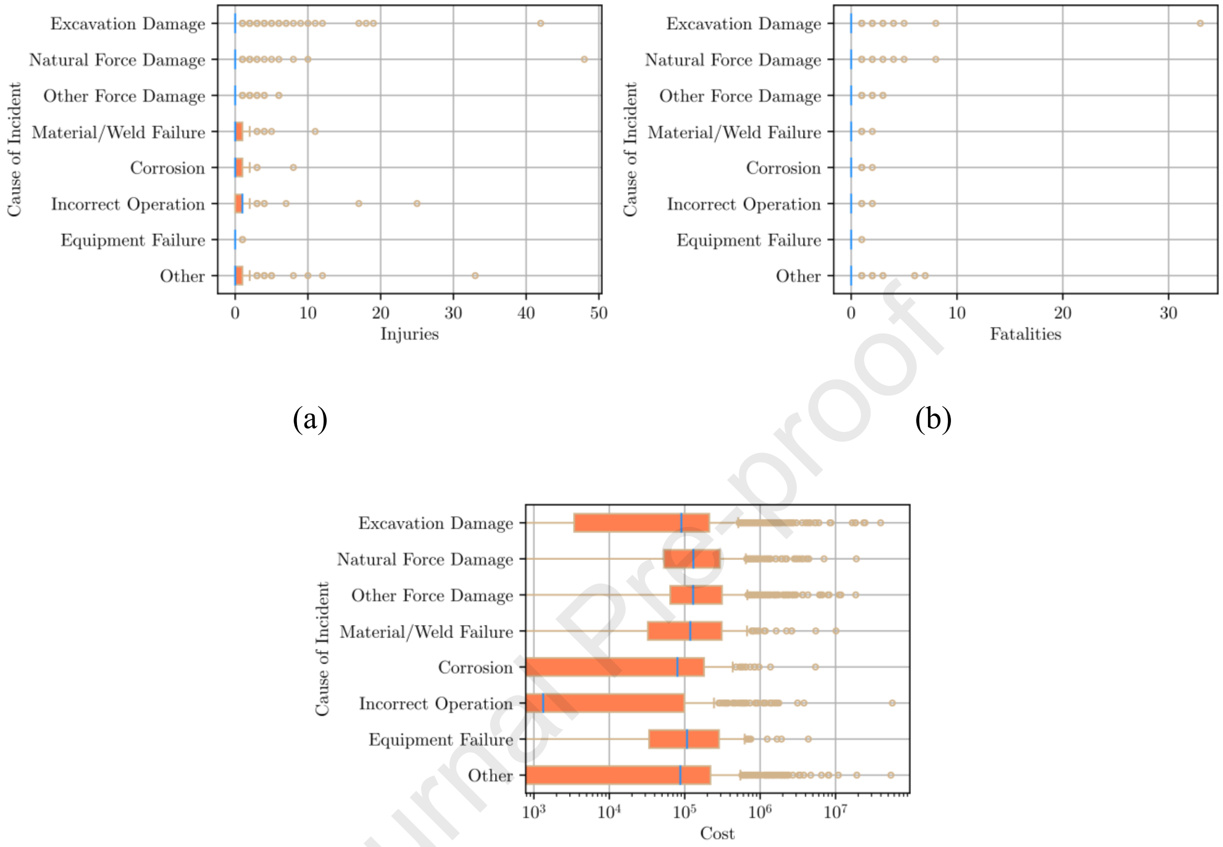
3.2.4 Incident cause

PHMSA assigns an incident cause to each reported pipeline incident to indicate the apparent cause, after a careful examination of the incident data or experimental investigation (Halim et al., 2020). These incident causes can be categorized into excavation damage, natural force damage, other outside force damage, material/weld failure, corrosion, incorrect operation, equipment failure, and other. **Table 7** and **Fig. 12** summarize the statistics of injuries, fatalities, and costs associated with gas distribution pipeline incidents by incident cause. The analysis of pipeline incident failure consequences reveals notable differences across incident causes. Regarding injuries, incidents caused by incorrect operation had the highest mean (1.017) with a wide range, likely due to such incidents involving operators in the vicinity. This was followed by incidents caused by corrosion (0.732) and material/weld failure (0.553). In terms of fatalities, incidents caused by natural force damage (0.121), corrosion (0.123), and excavation damage (0.094) had the highest mean counts, while other causes were relatively lower. Regarding incident cost, natural force damage had the highest mean ($375,907), followed by other outside force damage ($363,601) and incorrect operation ($343,165), while corrosion and other categories had the lowest means.

Table 7. Statistical summary of consequences in US gas distribution pipelines by incident cause (1986 - present).

Incident cause	count	mean	std	min	25%	50%	75%	max
Injury								
Excavation Damage	1852	0.411	1.531	0	0	0	0	42
Natural Force Damage	379	0.496	2.631	0	0	0	0	48
Other Outside Force Damage	942	0.231	0.660	0	0	0	0	6
Material/Weld Failure	226	0.553	1.123	0	0	0	1	11
Corrosion	138	0.732	1.029	0	0	0	1	8
Incorrect Operation	286	1.017	1.929	0	0	1	1	25
Equipment Failure	79	0.114	0.320	0	0	0	0	1
Other	872	0.592	1.525	0	0	0	1	33
Fatality								
Excavation Damage	1852	0.094	0.859	0	0	0	0	33
Natural Force Damage	379	0.121	0.627	0	0	0	0	8
Other Outside Force Damage	942	0.086	0.351	0	0	0	0	2
Material/Weld Failure	226	0.066	0.267	0	0	0	0	0
Corrosion	138	0.123	0.351	0	0	0	0	2
Incorrect Operation	26	0.056	0.272	0	0	0	0	2
Equipment Failure	79	0.025	0.158	0	0	0	0	1
Other	872	0.181	0.649	0	0	0	0	7
Cost								
Excavation Damage	1852	296147	1495078	0	3427	90396	211903	39641122
Natural Force Damage	379	375907	1171782	0	53127	130332	293021	18867225
Other Outside Force Damage	942	363601	10389988	0	64329	129242	312099	18531510
Material/Weld Failure	226	300181	825427	0	32429	118428	309848	10143124
Corrosion	138	189065	500501	0	0	80106	180294	5423798
Incorrect Operation	26	343165	3367189	0	0	1328	98659	56678788

Equipment Failure	79	286524	578151	0	33760	107405	285236	4361205
Other	872	189065	500501	0	0	80106	180294	5423798



(a) (b)

(c)

Figure 12. Box plots illustrating (a) injuries, (b) fatalities, and (c) total costs, from incidents by incident cause in US gas distribution pipelines.

The results of the Kruskal-Wallis statistical test indicate significant differences in the consequences of gas pipeline incidents, including injuries ($H = 356.38$, $p = 5.30\times10^{-73}$), fatalities ($H = 45.68$, $p = 1.01\times10^{-7}$), and costs ($H = 220.45$, $p = 5.28\times10^{-44}$), depending on the underlying incident causes [35]. To determine which specific incident cause pairs exhibit significantly different outcomes in terms of injuries, fatalities, and costs, post-hoc pairwise comparisons were conducted, with the results listed in **Table 8**. For injuries, significant differences were found between excavation damage and corrosion, incorrect operation, and other factors; between natural force damage and corrosion, incorrect operation, and other factors; and between other outside force damage and corrosion, incorrect operation, and other factors. However, no significant differences were observed for some factor pairs. Regarding fatalities, the analysis indicates significant differences between excavation damage, natural force damage, other outside force damage and other factors, as well as between incorrect operation and other factors. The cost analysis shows significant differences between excavation damage and natural force damage, other outside force damage, and incorrect operation; between natural force damage and corrosion, incorrect operation, and other factors; and between other outside force damage, corrosion, incorrect operation, and other factors. These findings suggest that the type of incident cause is a critical variable in determining the severity of pipeline failure consequences.

Table 8. Pairwise comparisons for incident consequences in US gas distribution pipelines by incident cause (1986 - present).

Incident cause	Excavation Damage	Natural Force Damage	Other Force Damage	Material/ Weld Failure	Corrosion	Incorrect Operation	Equipment Failure	Other
				Injury				
Excavation Damage	1	**1**	4.21×10^{-3}	3.10×10^{-3}	7.99×10^{-10}	2.89×10^{-47}	**0.626**	2.64×10^{-10}
Natural Force Damage	**1**	1	**0.118**	**0.082**	3.75×10^{-7}	9.94×10^{-30}	**0.597**	7.93×10^{-4}
Other Force Damage	4.21×10^{-3}	**0.118**	1	2.89×10^{-7}	1.47×10^{-14}	1.59×10^{-56}	**1**	1.16×10^{-18}
Material/ Weld Failure	3.10×10^{-3}	**0.082**	2.89×10^{-7}	1	**0.100**	3.52×10^{-12}	1.19×10^{-3}	**1**
Corrosion	7.99×10^{-10}	3.75×10^{-7}	1.47×10^{-14}	**0.100**	1	0.025	4.83×10^{-8}	**0.022**
Incorrect Operation	2.89×10^{-47}	9.94×10^{-30}	1.59×10^{-56}	3.52×10^{-12}	0.025	1	1.58×10^{-19}	3.04×10^{-20}
Equipment Failure	**0.626**	**0.597**	**1**	1.19×10^{-3}	4.83×10^{-8}	1.58×10^{-19}	1	1.09×10^{-4}
Other	2.64×10^{-10}	7.93×10^{-4}	1.16×10^{-18}	**1**	**0.022**	3.04×10^{-20}	1.09×10^{-4}	1
				Fatality				
Excavation Damage	1	**1**	**1**	**1**	**0.303**	**1**	**1**	1.32×10^{-7}
Natural Force Damage	**1**	1	**1**	**1**	**1**	**1**	**1**	0.0163
Other Force Damage	**1**	**1**	1	**1**	**1**	**1**	**1**	5.32×10^{-4}
Material/ Weld Failure	**1**	**1**	**1**	1	**1**	**1**	1	**0.0798**
Corrosion	**0.303**	**1**	**1**	**1**	1	**0.268**	**0.381**	**1**

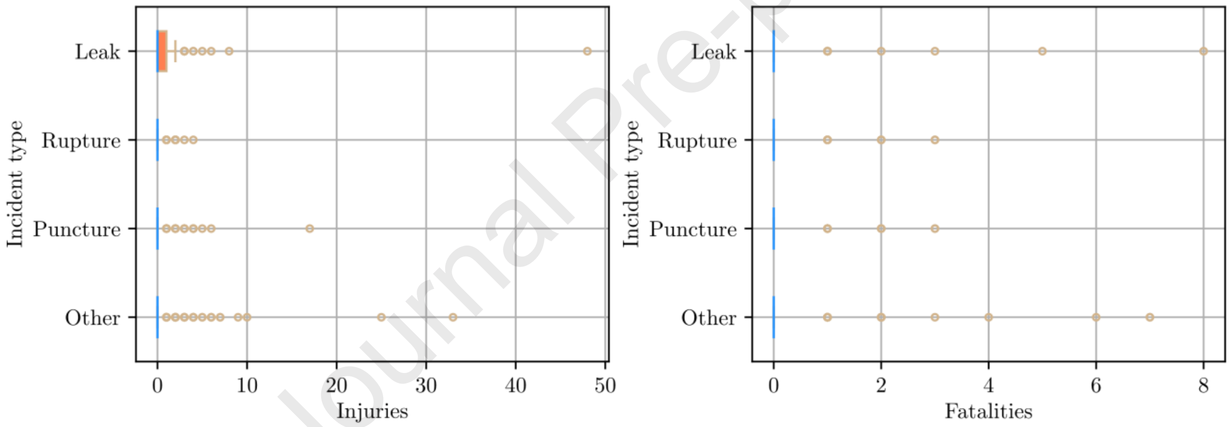
Incorrect Operation	**1**	**1**	**1**	**1**	**0.268**	1	**1**	8.51×10^{-4}
Equipment Failure	**1**	**1**	**1**	**1**	**0.381**	**1**	1	**0.0548**
Other	1.32×10^{-7}	0.0163	5.32×10^{-4}	**0.0798**	**1**	8.51×10^{-4}	**0.0548**	1
Cost								
Excavation Damage	1	9.67×10^{-6}	9.28×10^{-16}	**0.143**	**1**	8.07×10^{-16}	**0.999**	**1**
Natural Force Damage	9.67×10^{-6}	1	**1**	**1**	9.66×10^{-4}	1.88×10^{-24}	**1**	3.84×10^{-6}
Other Force Damage	9.28×10^{-16}	**1**	1	**1**	1.12×10^{-5}	6.34×10^{-37}	**1**	4.57×10^{-14}
Material/ Weld Failure	**0.143**	**1**	**1**	1	**0.0815**	4.50×10^{-15}	**1**	0.0481
Corrosion	**1**	9.66×10^{-4}	1.12×10^{-5}	**0.0815**	1	19.3×10^{-3}	**0.267**	**1**
Incorrect Operation	8.07×10^{-16}	1.88×10^{-24}	6.34×10^{-37}	4.50×10^{-15}	19.3×10^{-3}	1	2.57×10^{-8}	6.01×10^{-12}
Equipment Failure	**0.999**	**1**	**1**	**1**	**0.267**	2.57×10^{-8}	1	**0.502**
Other	**1**	3.84×10^{-6}	4.57×10^{-14}	0.0481	**1**	6.01×10^{-12}	**0.502**	1

3.2.5 Incident type

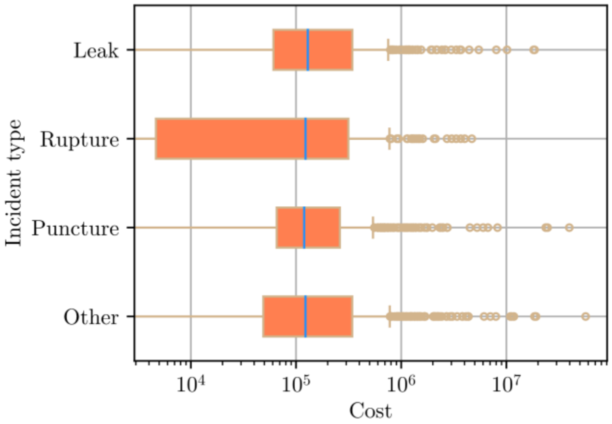
Since 2004, pipeline incident reports have been required to include information on the type of incident, such as leak, rupture, puncture, or other. Generally, the incident type reflects the extent of pipe damage, with more severe gas releases potentially leading to more severe consequences if ignited. **Table 9** and **Fig. 13** provide summary statistics and box plots of the injuries, fatalities, and costs associated with gas distribution pipeline incidents by incident type. Notable variations exist in the mean number of injuries, fatalities, and costs across different incident types. Leak incidents have the highest mean number of injuries at 0.472, while rupture incidents have the lowest at 0.239. Regarding fatalities, rupture incidents have the highest mean at 0.104, and puncture incidents have the lowest at 0.047. The other category of incident types has the highest mean cost of $491,257, whereas rupture incidents have the lowest mean cost of $308,063. The presence of outliers suggests that while the mean values offer a general summary, individual incidents may have substantially higher impacts in terms of injuries, fatalities, and costs.

Table 9. Statistical summary of consequences in US gas distribution pipelines by incident type (2004 - present).

Incident type	count	mean	std	min	25%	50%	75%	max
Injury								
Leak	502	0.472	2.283	0	0	0	1	48
Rupture	280	0.239	0.596	0	0	0	0	4
Puncture	617	0.245	0.955	0	0	0	0	17
Other	871	0.429	1.679	0	0	0	0	33
Fatality								
Leak	502	0.080	0.491	0	0	0	0	8
Rupture	280	0.104	0.397	0	0	0	0	3
Puncture	617	0.047	0.278	0	0	0	0	3
Other	871	0.118	0.535	0	0	0	0	7
Cost								
Leak	502	416045	1385558	0	60967	129375	342820	18531510
Rupture	280	308063	598748	0	4655	123232	314339	4692619
Puncture	617	428245	2185961	0	65585	119541	261061	39641121
Other	871	491257	2379491	0	48869	123059	342154	56678788



(a) (b)



(c)

Figure 13. Box plots illustrating (a) injuries, (b) fatalities, and (c) total costs, from incidents by incident type in US gas distribution pipelines.

The Kruskal-Wallis test results indicate statistically significant differences in the distribution of injuries ($H = 26.67$, $p = 6.92\times10^{-6}$) and fatalities ($H = 14.65$, $p = 2.14\times10^{-3}$) across incident types. However, no significant difference was found in the cost distribution ($H = 5.24$, $p = 0.155$). The post-hoc Dunn's test was then employed to identify the exact pairwise differences between the incident types, with the results shown in **Table 10**. The analysis revealed statistically significant differences in the distribution of injuries between leaks and ruptures, leaks and punctures, and punctures and others. However, no statistically significant differences were found in injuries between leaks and others, ruptures and punctures, or ruptures and others. Regarding fatalities, the analysis showed a statistically significant difference between punctures and others, but no significant differences were observed among the other incident types. Overall, the findings indicate that incident type is closely associated with injury outcomes, with some types being associated with more severe consequences.

Table 10. Pairwise comparisons for incident consequences in US gas distribution pipelines by incident part (2004 - present).

Incident type	Leak	Rupture	Puncture	Other
		Injury		
Leak	1	0.0272	1.42×10^{-5}	**0.794**
Rupture	0.0272	1	**1**	**0.382**
Puncture	1.42×10^{-5}	**1**	1	9.12×10^{-4}
Other	**0.794**	**0.382**	9.12×10^{-4}	1
		Fatality		
Leak	1	**0.779**	**1**	**0.278**
Rupture	**0.779**	1	**0.0584**	**1**
Puncture	**1**	**0.0584**	1	2.66×10^{-3}
Other	**0.278**	**1**	2.66×10^{-3}	1
		Cost		
Leak				
Rupture				
Puncture				
Other				

3.3 Distribution mains and service lines

As discussed in Section 2.2.2, five systems are involved in a pipeline incident, with the proportions of the different systems illustrated in **Fig. 14**. The distribution mains (40.8%) and service lines (24.3%) comprise the two largest segments. Given the differences between these two pipeline types, where distribution mains are the larger, higher-pressure pipelines that form the primary distribution network, while service lines are the smaller, lower-pressure pipelines that provide the final connection to individual customers - the incident characteristics will be analyzed in further detail in this section.

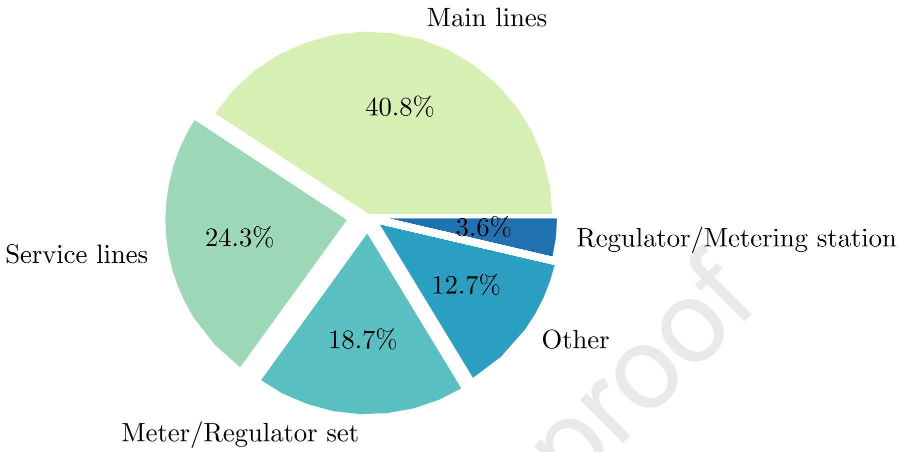
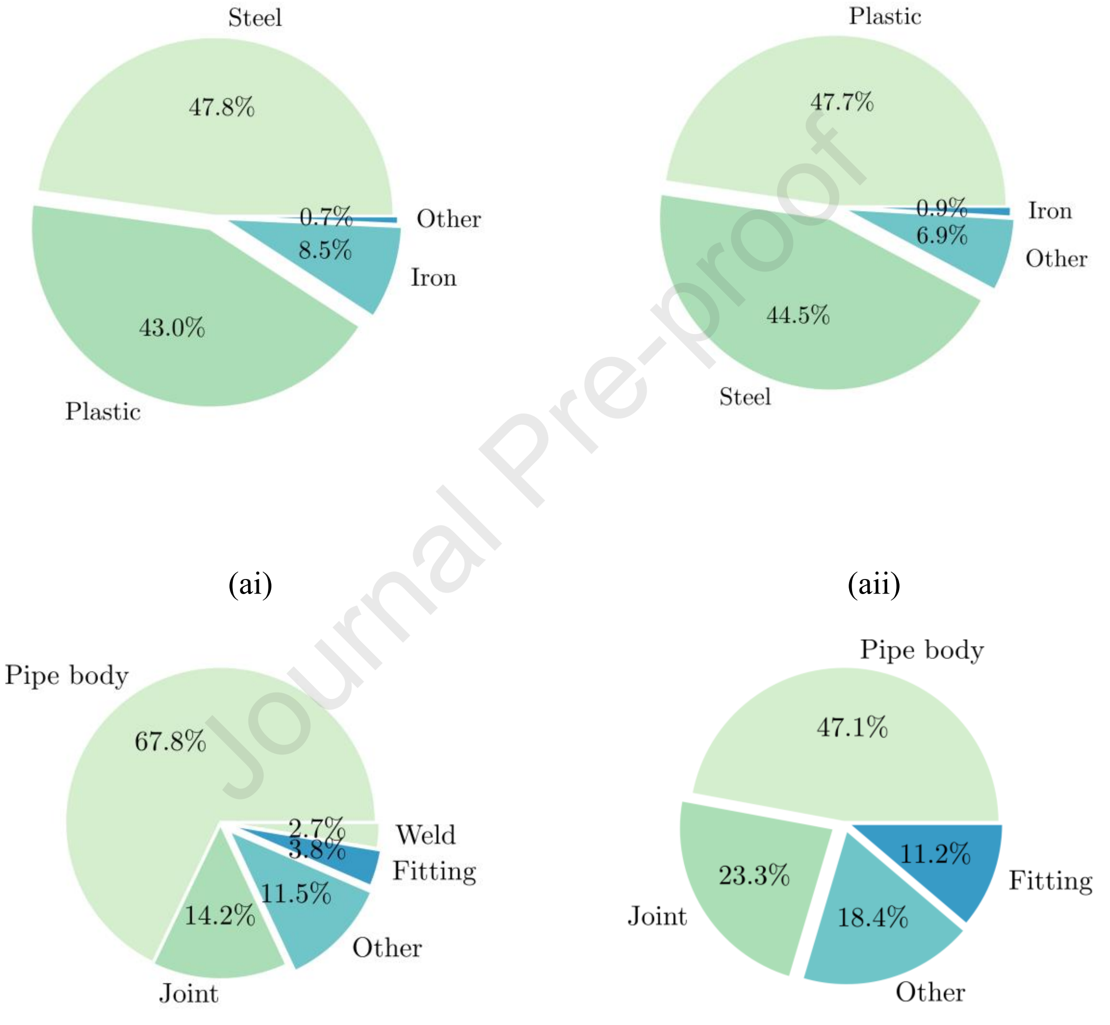


Figure 14. Distribution of different systems involved in gas distribution pipeline incidents in the US.

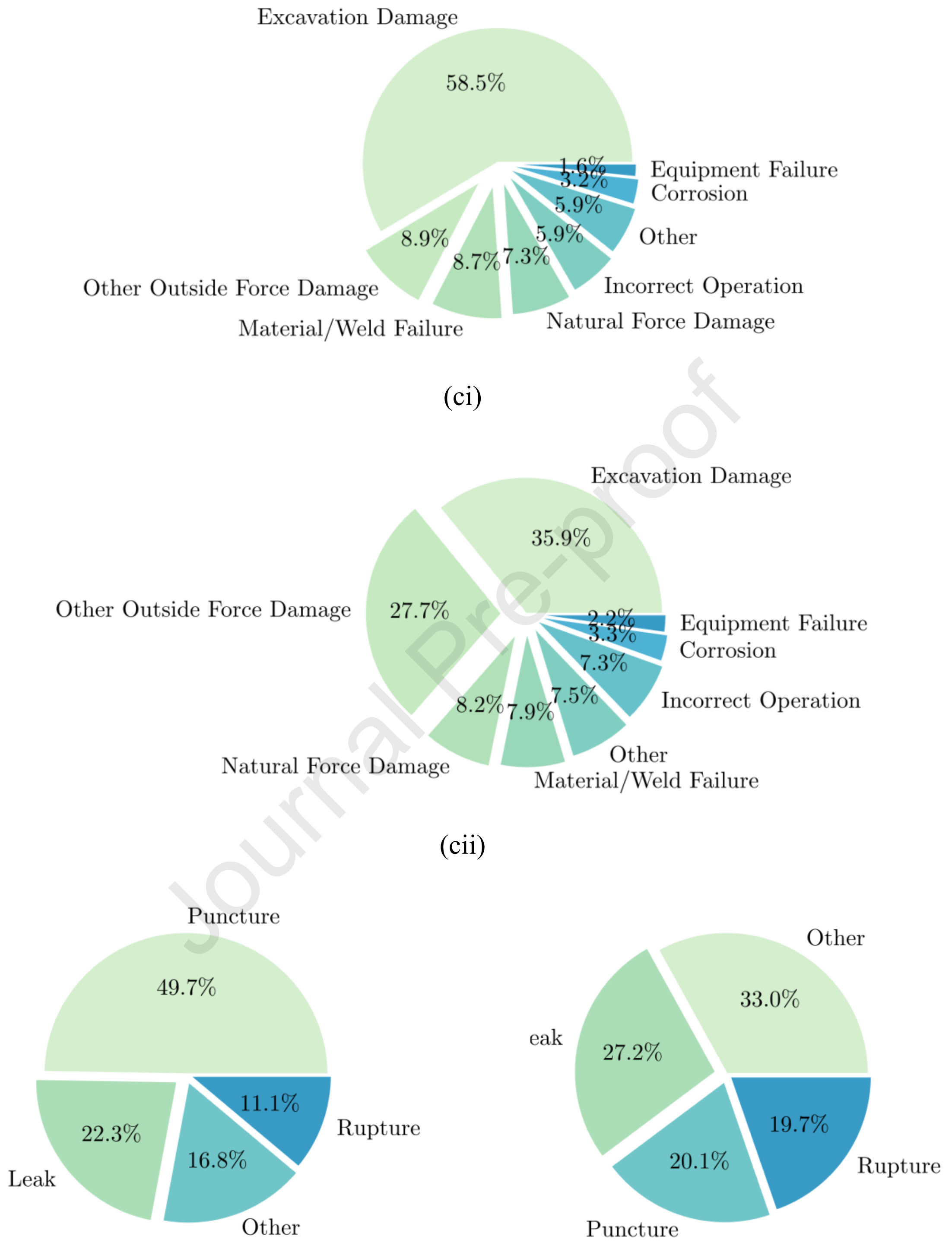
As shown in **Fig. 15(a)**, the distribution of pipe materials involved in pipeline incidents indicates that steel (47.8%) and plastic (43.0%) are the predominant materials used in main lines, while plastic (47.7%) and steel (44.5%) are more common in-service lines. Iron and other materials account for smaller proportions in both cases. Generally, the distribution of pipe materials is quite similar between the main and service lines. The distribution of pipeline incident involvement, shown in **Fig. 15(b)**, indicates that the pipe body is the primary component in main lines (67.8%) and service lines (47.1%), followed by joint failures and other components in both. Fitting and weld issues are more prevalent in service line incidents compared to main lines, likely due to design or material differences that result in varying failure modes and incident patterns. The greater prevalence of pipe body failures in main lines versus more joint, and fitting issues in service lines suggests potential differences in the dominant failure modes and points of weakness between the two systems. However, further detailed investigations are needed in future studies.

The distribution of pipeline incident causes, shown in **Fig. 15(c)**, reveals excavation damage as the predominant factor in main lines (58.5%), followed by smaller proportions of other external forces, material/weld failures, and natural causes. In contrast, service line incidents present a more balanced distribution, with excavation damage (35.9%) and other outside force damage (27.7%) as the leading contributors, alongside natural forces, operational errors, and material issues. Notably, force-related damage accounts for the vast majority (over 70%) of both distribution pipelines, highlighting the critical need to enhance excavation safety protocols, damage prevention programs, and monitoring practices as a top priority for improving the overall integrity and reliability of the gas distribution network. As shown in **Fig. 15(d)**, punctures are the dominant incident mode in main lines (49.7%), while service lines exhibit a more diverse breakdown, with a greater proportion of other incident types (33.0%), alongside punctures, ruptures, and leaks. This suggests that the failure modes and mechanisms may differ between the two pipeline systems, likely driven by factors such as material properties, operating pressures, and failure triggers.



(ai) (aii)

(bi) (bii)



(ci)

(cii)

(di)

(dii)

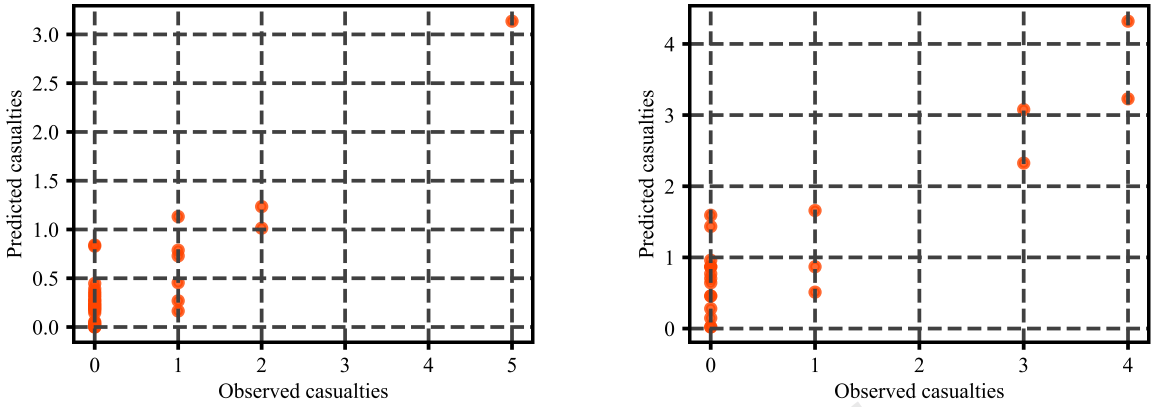
Figure 15. Distribution of (a) pipe material, (b) incident part, (c) incident cause, and (d) incident type, of gas distribution pipeline incidents in the US between (i) main lines, and (ii) service lines.

Based on the findings, Zero-Inflated Poisson regression was applied to model the number of casualties associated with each gas distribution pipeline for the main and service lines, respectively. The response variable in the analysis was the total count of injuries and fatalities per pipeline incident, while the predictor variables included various characteristics of the incidents, such as pipe age, pipe material, and incident cause. **Table 11** lists the final predictor variable sets determined for the establishment of the Zero-Inflated Poisson regression models for main and service lines. One-hot encoding was used to process categorical variables in the dataset into binary indicator (dummy) variables. The dataset was then split into a training and testing dataset with an 70:30 ratio for the incident records between main and service lines, where the training dataset was used to fit the models and the testing dataset was used to evaluate their performance. The iteratively reweighted least squares (IRLS) method was used to update the parameter estimates until convergence.

Table 11. Predictor variables and evaluation metrics of the Zero-Inflated Poisson regression models for main and service lines.

Pipe system	Predictor variables	MAE	MSE	R^2
Main line	Diameter, thickness, material, operating pressure, incident part, incident type, incident cause	0.287	0.468	0.640
Service line	Thickness, material, operating pressure, incident part, incident type, incident cause	0.609	0.738	0.697

Fig. 16 presents scatter plots illustrating the relationship between observed and predicted casualties using Zero-Inflated Poisson regression models. In **Fig. 16(a)**, the scatter of points for the main lines indicates a moderate correlation, suggesting that while the model provides some predictive power, it may not fully capture the data's complexities. Conversely, the tighter clustering of points in **Fig. 16(b)** for the service line model suggests a stronger visual correlation and a closer alignment with a linear trend. However, a closer examination of the error metrics, summarized in **Table 11**, presents a different perspective. The main line model demonstrates greater predictive accuracy, with a lower MAE of 0.287 and MSE of 0.468, compared to the service line model, which exhibits higher errors (MAE = 0.609, MSE = 0.738). Despite this, the service line model achieves a higher R^2 value of 0.697, indicating that it explains a larger proportion of the variance in the dependent variable than the main line model, which has an R^2 of 0.640. While the service line model better captures overall trends, it may still struggle with individual predictions. These results underscore the need for further refinement. Incorporating additional variables, such as environmental conditions, pipeline age and length, or operational characteristics, could enhance the models' predictive power. The current study also acknowledges that exploring alternative modeling approaches with greater capability to model pipeline incidents may improve performance. Such improvements are critical for enhancing the safety of gas distribution systems, enabling more targeted interventions and facilitating proactive risk management.



(a) (b)

Figure 16. Scatter plot of the predicted casualties and observed casualties involved in (a) distribution mains, and (b) service lines.

4. Conclusion

This study statistically analyzes pipeline incidents in the U.S. gas distribution system, focusing on incident numbers, rates, injuries, fatalities, and total costs. The geographic distribution of incidents at the state level was also examined, revealing significant variations. The analysis explored pipeline attributes, pipe material, system involved, incident part, cause, and type, to assess their impact on injuries, fatalities, and costs. Finally, the study compared distribution mains and service lines, and developed Zero-Inflated Poisson regression models to predict casualties per incident.

The findings show a moderate decline in incident numbers and rates over time, likely due to improved safety measures. However, incident costs have risen, driven by rare catastrophic events. Steel pipelines account for the highest number of incidents, highlighting the need for enhanced inspection, while plastic pipelines face higher risks during early use, indicating the need for improved installation and inspection practices. Seasonal factors, particularly temperature, were also found to influence failure rates. The state-level analysis demonstrated a strong positive correlation between pipeline length and incident numbers, with Texas and California having the highest incident counts.

Significant differences were found in the consequences of incidents based on pipe materials and systems. Iron pipelines reported the highest mean injuries and fatalities, while plastic pipelines incurred the highest costs. Service lines exhibited the most severe consequences across all metrics compared to main lines. Failure modes related to joints and welds were associated with higher fatalities and costs than pipe body failures. Major causes of severe incidents included incorrect operations, natural forces, and excavation damage. The Zero-Inflated Poisson regression models demonstrated moderate accuracy, with better predictive performance for service lines.

The limited accessibility of a comprehensive dataset, including pipeline length, maintenance history, gas composition, and environmental conditions, restricts the model's ability to predict incident causality. It is advisable to explore more advanced models in future studies to enhance incident prediction and, consequently, improve the safety of gas distribution systems.

Acknowledgment

The first author gratefully acknowledges the support received from The Hong Kong Polytechnic University and McGill University through the Joint Postdoc Scheme.

References

Bachir-Bey, T., Belhaneche-Bensemra, N., 2020. Investigation of Polyethylene Pipeline Behavior after 30 Years of Use in Gas Distribution Network. J. of Materi Eng and Perform 29, 6652–6660. https://doi.org/10.1007/s11665-020-05118-9

Bianchini, A., Guzzini, A., Pellegrini, M., Saccani, C., 2018. Natural gas distribution system: A statistical analysis of accidents data. International Journal of Pressure Vessels and Piping 168, 24–38. https://doi.org/10.1016/j.ijpvp.2018.09.003

Cimellaro, G.P., Villa, O., Bruneau, M., 2015. Resilience-Based Design of Natural Gas Distribution Networks. J. Infrastruct. Syst. 21, 05014005. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000204

Coxe, S., West, S.G., Aiken, L.S., 2009. The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. Journal of Personality Assessment 91, 121–136. https://doi.org/10.1080/00223890802634175

Dunn, O.J., 1964. Multiple Comparisons Using Rank Sums. Technometrics 6, 241–252. https://doi.org/10.1080/00401706.1964.10490181

Fan, X., Wang, X., Zhang, X., ASCE Xiong (Bill) Yu, P.E.F., 2022. Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. Reliability Engineering & System Safety 219, 108185. https://doi.org/10.1016/j.ress.2021.108185

Halim, S.Z., Yu, M., Escobar, H., Quddus, N., 2020. Towards a causal model from pipeline incident data analysis. Process Safety and Environmental Protection 143, 348–360. https://doi.org/10.1016/j.psep.2020.06.047

Hall, D.B., 2000. Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. Biometrics 56, 1030–1039. https://doi.org/10.1111/j.0006-341X.2000.01030.x

Hendrick, M.F., Ackley, R., Sanaie-Movahed, B., Tang, X., Phillips, N.G., 2016. Fugitive methane emissions from leak-prone natural gas distribution infrastructure in urban environments. Environmental Pollution 213, 710–716. https://doi.org/10.1016/j.envpol.2016.01.094

Herrán-González, A., De La Cruz, J.M., De Andrés-Toro, B., Risco-Martín, J.L., 2009. Modeling and simulation of a gas distribution pipeline network. Applied Mathematical Modelling 33, 1584–1600. https://doi.org/10.1016/j.apm.2008.02.012

Hocine, A., Kara Achira, F.S., Habbar, G., Levent, A., Medjdoub, S.M., Maizia, A., Dhaou, M.H., Bezazi, A., 2024. Structural integrity assessment of corroded pipelines repaired with composite materials – Literature review. International Journal of Pressure Vessels and Piping 210, 105253. https://doi.org/10.1016/j.ijpvp.2024.105253

Jo, Y.-D., Ahn, B.J., 2005. A method of quantitative risk assessment for transmission pipeline carrying natural gas. Journal of Hazardous Materials 123, 1–12. https://doi.org/10.1016/j.jhazmat.2005.01.034

Khademi-Zahedi, R., 2019. Application of the finite element method for evaluating the stress distribution in buried damaged polyethylene gas pipes. Underground Space 4, 59–71. https://doi.org/10.1016/j.undsp.2018.05.002

Kruskal, W.H., Wallis, W.A., 1952. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association 47, 583–621. https://doi.org/10.1080/01621459.1952.10483441

Lam, C., Zhou, W., 2016. Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. International Journal of Pressure Vessels and Piping 145, 29–40. https://doi.org/10.1016/j.ijpvp.2016.06.003

Lambert, D., 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. Technometrics 34, 1. https://doi.org/10.2307/1269547

Li, X., Penmetsa, P., Liu, J., Hainen, A., Nambisan, S., 2021. Severity of emergency natural gas distribution pipeline incidents: Application of an integrated spatio-temporal approach fused with text mining. Journal of Loss Prevention in the Process Industries 69, 104383. https://doi.org/10.1016/j.jlp.2020.104383

Ma, L., Li, Y., Liang, L., Li, M., Cheng, L., 2013. A novel method of quantitative risk assessment based on grid difference of pipeline sections. Safety Science 59, 219–226. https://doi.org/10.1016/j.ssci.2013.04.012

Mann, H.B., 1945. Nonparametric Tests Against Trend. Econometrica 13, 245. https://doi.org/10.2307/1907187

McGeehan, P., 2015. Con Edison and New York City Are Faulted in East Harlem Explosion [WWW Document]. URL https://www.nytimes.com/2015/06/10/nyregion/consolidated-edison-is-largely-liable-in-deadly-east-harlem-explosion-regulators-find.html

Miller, R.G., 1966. Simultaneous statistical inference, 2nd ed. Springer International Publishing, Cham.

Nelder, J.A., 1974. Log Linear Models for Contingency Tables: A Generalization of Classical Least Squares. Applied Statistics 23, 323. https://doi.org/10.2307/2347125

Office of the Federal Register, 2023. Code of Federal Regulations. U.S. Government Publishing Office, Washington D.C.

PHMSA, 2023a. Annual Report Mileage Summary Statistics.

PHMSA, 2023b. Pipeline Incident 20 Year Trends.

PHMSA, 2023c. Pipeline Incident Flagged Files [WWW Document]. URL https://www.phmsa.dot.gov/data-and-statistics/pipeline/data-and-statistics-overview

PHMSA, 2017. History of PHMSA Incident Reporting Criteria [WWW Document]. URL https://www.phmsa.dot.gov/data-and-statistics/pipeline/history-phmsa-incident-reporting-criteria

Rahimi, F., Sadeghi-Niaraki, A., Ghodousi, M., Abuhmed, T., Choi, S.-M., 2024. Temporal dynamics of urban gas pipeline risks. Sci Rep 14, 5509. https://doi.org/10.1038/s41598-024-56136-9

Rajani, B., Kleiner, Y., 2001. Comprehensive review of structural deterioration of water mains: physically based models. Urban Water 3, 151–164. https://doi.org/10.1016/S1462-0758(01)00032-2

Ramírez-Camacho, J.G., Carbone, F., Pastor, E., Bubbico, R., Casal, J., 2017. Assessing the consequences of pipeline accidents to support land-use planning. Safety Science 97, 34–42. https://doi.org/10.1016/j.ssci.2016.01.021

Restrepo, C.E., Simonoff, J.S., Zimmerman, R., 2009. Causes, cost consequences, and risk implications of accidents in US hazardous liquid pipeline infrastructure. International Journal of Critical Infrastructure Protection 2, 38–50. https://doi.org/10.1016/j.ijcip.2008.09.001

Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation Coefficients: Appropriate Use and Interpretation. Anesthesia & Analgesia 126, 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Shan, K., Shuai, J., Xu, K., Zheng, W., 2018. Failure probability assessment of gas transmission pipelines based on historical failure-related data and modification factors. Journal of Natural Gas Science and Engineering 52, 356–366. https://doi.org/10.1016/j.jngse.2018.01.049

Shen, Y., Zhou, W., 2024. A comparison of onshore oil and gas transmission pipeline incident statistics in Canada and the United States. International Journal of Critical Infrastructure Protection 45, 100679. https://doi.org/10.1016/j.ijcip.2024.100679

Siler-Evans, K., Hanson, A., Sunday, C., Leonard, N., Tumminello, M., 2014. Analysis of pipeline accidents in the United States from 1968 to 2009. International Journal of Critical Infrastructure Protection 7, 257–269. https://doi.org/10.1016/j.ijcip.2014.09.002

Silva, R.C.C., Guerreiro, J.N.C., Loula, A.F.D., 2007. A study of pipe interacting corrosion defects using the FEM and neural networks. Advances in Engineering Software 38, 868–875. https://doi.org/10.1016/j.advengsoft.2006.08.047

Soomro, A.A., Mokhtar, A.A., Kurnia, J.C., Lashari, N., Lu, H., Sambo, C., 2022. Integrity assessment of corroded oil and gas pipelines using machine learning: A systematic review. Engineering Failure Analysis 131, 105810. https://doi.org/10.1016/j.engfailanal.2021.105810

Teng, L., Liu, X., Li, X., Li, Y., Lu, C., 2021. An approach of quantitative risk assessment for release of supercritical CO2 pipelines. Journal of Natural Gas Science and Engineering 94, 104131. https://doi.org/10.1016/j.jngse.2021.104131

Vetter, C.P., Kuebel, L.A., Natarajan, D., Mentzer, R.A., 2019. Review of failure trends in the US natural gas pipeline industry: An in-depth analysis of transmission and distribution system incidents. Journal of Loss Prevention in the Process Industries 60, 317–333. https://doi.org/10.1016/j.jlp.2019.04.014

Wang, W., Zhang, Y., Li, Y., Hu, Q., Liu, Chengsong, Liu, Cuiwei, 2022. Vulnerability analysis method based on risk assessment for gas transmission capabilities of natural gas pipeline networks. Reliability Engineering & System Safety 218, 108150. https://doi.org/10.1016/j.ress.2021.108150

Wikimedia, 2014. 2014 East Harlem gas explosion [WWW Document]. URL https://en.wikipedia.org/wiki/2014_East_Harlem_gas_explosion

Wikipedia, 2018. Merrimack Valley gas explosions [WWW Document]. URL https://en.wikipedia.org/wiki/Merrimack_Valley_gas_explosions

Wu, H., Zhao, H., Li, X., Feng, X., Chen, Y., 2022. Experimental and numerical studies on collapse of subsea pipelines with interacting corrosion defects. Ocean Engineering 260, 112066. https://doi.org/10.1016/j.oceaneng.2022.112066

Xiao, R., Zayed, T., Meguid, M.A., Sushama, L., 2024. Improving failure modeling for gas transmission pipelines: A survival analysis and machine learning integrated approach. Reliability Engineering & System Safety 241, 109672. https://doi.org/10.1016/j.ress.2023.109672

Xiao, R., Zayed, T., Meguid, M.A., Sushama, L., 2023. Understanding the factors and consequences of pipeline incidents: An analysis of gas transmission pipelines in the US. Engineering Failure Analysis 152, 107498. https://doi.org/10.1016/j.engfailanal.2023.107498

Xu, J., Jiang, F., Xie, Z., Wang, G., 2023. Risk Assessment Method for the Safe Operation of Long-Distance Pipeline Stations in High-Consequence Areas Based on Fault Tree Construction: Case Study of China–Myanmar Natural Gas Pipeline Branch Station.

ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng. 9, 05022003. https://doi.org/10.1061/AJRUA6.RUENG-960

Zakikhani, K., Nasiri, F., Zayed, T., 2021. A failure prediction model for corrosion in gas transmission pipelines. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 235, 374–390. https://doi.org/10.1177/1748006X20976802

Zakikhani, K., Nasiri, F., Zayed, T., 2020. A Review of Failure Prediction Models for Oil and Gas Pipelines. J. Pipeline Syst. Eng. Pract. 11, 03119001. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000407

Zamenian, H., Mannering, F.L., Abraham, D.M., Iseley, T., 2017. Modeling the Frequency of Water Main Breaks in Water Distribution Systems: Random-Parameters Negative-Binomial Approach. J. Infrastruct. Syst. 23, 04016035. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000336

Zerouali, B., Sahraoui, Y., Nahal, M., Chateauneuf, A., 2024. Reliability-based maintenance optimization of long-distance oil and gas transmission pipeline networks. Reliability Engineering & System Safety 249, 110236. https://doi.org/10.1016/j.ress.2024.110236

Highlights:

- Statistical analysis of natural gas distribution pipeline incidents in the U.S.
- Decreasing trend in incident frequency and severity over 30 years
- Incident costs show little correlation with number of incidents
- Pipeline material and attributes impact failure patterns and consequences

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: