+

U.S. Department of Transportation

Pipeline and Hazardous Materials
Safety Administration
Washington, DC 20590

Valuation of Crude Oil Spills in Transportation Incidents

Office of Hazardous Materials Safety

April 2023

Table of Contents

List of Tables

List of Figures

Executive Summary

The Pipeline and Hazardous Materials Safety Administration (PHMSA) conducts regulatory impact analyses and economic analyses of regulations, standards and other policies that address safety and the risk of a crude oil spill during transportation. This study aims to enhance PHMSA's ability to quantify the benefits of preventing releases of crude oil from pipelines and railcars through the development of defensible estimates of the social costs of such releases. Specifically, PHMSA uses historical crude oil spill incidents as the basis for developing empirical relationships between incident characteristics and the resulting social costs. These relationships have the following general form:

$$LogCost_i = f(SpillChar_i, SiteChar_i, CostMethod_i)$$

Where $LogCost_i$ is the log of total cost of incident i, $SpillChar_i$ is a vector that describes the characteristics of the spill, such as the quantity released, the spill source (pipeline vs. rail), or the spill mechanism (e.g., derailment or seepage), $SiteChar_i$ is a vector that describes the characteristics of the spill site such as population, proximity to water, whether the site is categorized as a high consequence area (HCA), and $CostMethod_i$ is a vector that describes the methodology used to estimate spill costs. For example, this might include whether or not particular categories of costs (e.g., commodity lost, operator damages) were accounted for in the cost estimate.

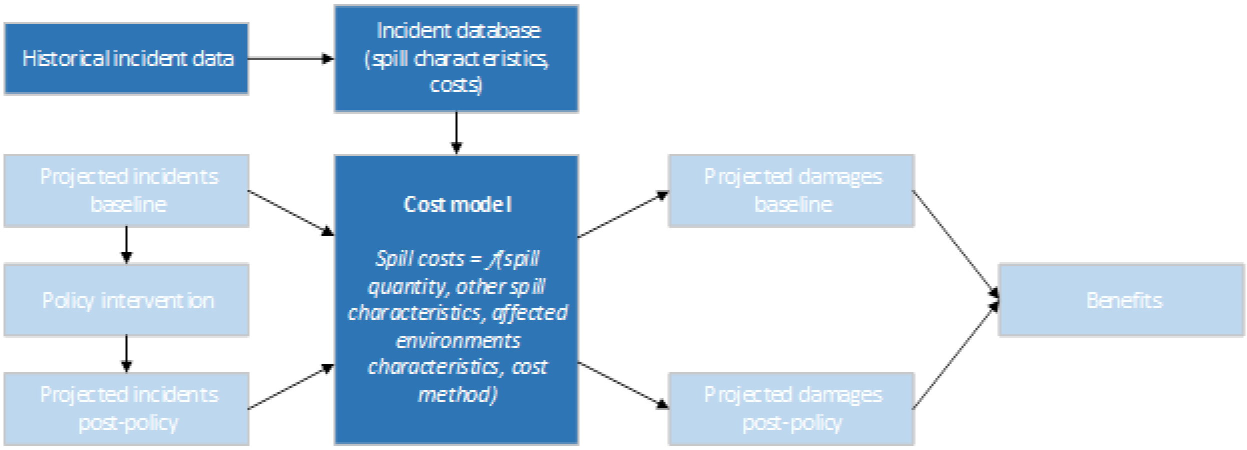
This study builds on a prior effort completed in 2017 and updates the empirical relationships using more recent data.

Intended Model Application

PHMSA intends for the cost models to be used to estimate the costs of an incident given information about the incident's characteristics and location. For example, PHMSA could use the cost models to calculate the expected costs of incidents projected to occur in the future under varying scenarios. Thus, one possible application of the cost models is to estimate the change in expected spill costs given a policy-induced change in the number, magnitude, or character of future incidents, relative to a pre-policy baseline. The expected costs will vary according to the transportation mode (rail or pipeline), crude oil quantity released, incident location (e.g., HCA, densely populated area), and other incident characteristics. The difference between the sum of total baseline incident costs and total post-regulation incident costs represents the avoided social costs (i.e., the benefits) of the policy.

Figure ES-1 illustrates the overall benefits analysis framework (elements addressed in this study are highlighted in the darker boxes in the diagram).

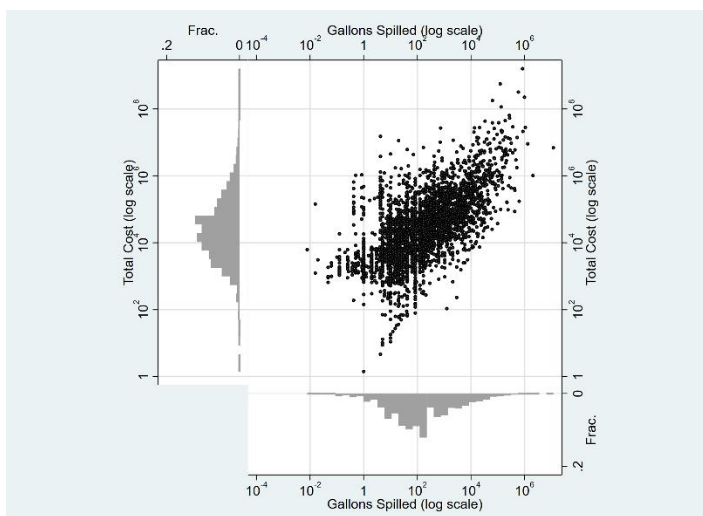
Figure ES-1: Overall benefits analysis framework with darker boxes highlighting the elements addressed in the study.



Incident Data

The study uses operator-reported data on pipeline and rail incidents resulting in the release of crude oil. PHMSA supplemented these primary datasets with data from the National Transportation Safety Board, U.S. Environmental Protection Agency, and other sources. For this updated analysis, PHMSA expanded the temporal scope of the incident database developed for the 2017 study to the end of 2022. The updated data cover the period of 2005 through 2022 and includes a total of 3,734 incidents: 3,250 pipeline releases and 484 rail releases. Of this total, 3,581 incidents (340 rail incidents and 3,241 pipeline incidents) reported some costs associated with the release.

Figure ES-2: Relationship between quantity released and total costs.



Exploratory analysis of the incident data shows a general trend of increasing total costs with increasing crude oil quantity released (Figure ES-2), but also shows that total costs are influenced by a multitude of other factors, including the transportation mode, incident location, and circumstances surrounding the release. PHMSA used a regression-based analysis to further examine and quantify the relationship between incident costs and potential influential characteristics, including the transportation mode (pipeline or rail), spill characteristics such as the volume of crude oil released and incident cause (e.g., train derailment, pipe corrosion), the incident location such as proximity to water or within a high consequence area (HCA), and incident circumstances (e.g., whether the spill was associated with a fire). Following the approach from the 2017 study, PHMSA specified separate regressions for rail and pipeline incidents.

Table ES-1 presents the two models developed in the study to relate total spill costs and characteristics of rail and pipeline incidents. The rail model shows the costs of rail incidents depending on the quantity of oil released, population living within a 800-meter radius of the spill site, and presence of wetlands or water bodies within a 800-meter radius. Additionally, spills caused by derailments tend to be associated with higher spill costs, whereas spills caused by components or aging equipment have relatively lower costs. Together, these parameters explain approximately 79 percent of the observed variability in total costs across rail incidents in the dataset. The pipeline model similarly shows the costs of pipeline incidents varying with the quantity of crude oil released, and proximity of the spill site to population and waterbodies. The pipeline incident costs also depend on the presence of sensitive environments and development density within a 800-meter radius of the site, and whether the area is categorized as an HCA. Regarding spill circumstances, pipeline incidents associated with a fire or underground sources tend to have higher total costs. The model parameters explain approximately 50 percent of the observed variability in total costs across pipeline incidents in the dataset.

The table also provides an example of a simple application of the pipeline model to a hypothetical incident involving the release of 5,000-gallon inside an HCA, associated with a fire, and caused by corrosion. The incident is assumed to occur in an area with 500 residents and where the land use is 40 percent high-density development. Based on these incident parameters, the model predicts the total costs of $5,872,506. See Section 2 for details on the cost models.

Table ES-1: Explanatory variables and coefficients for modeling total costs

Variable Type	Explanatory Variable		Regression Coefficient		Hypothetical Pipeline Incident Value
			Rail	Pipeline	
Intercept	intercept	—	7.336***	8.335***	1
Spilled crude oil quantity	ln_qreleased	ln(gallons)	0.162***	-0.0454	8.52
	ln_qreleased_sq	$\ln(\text{gallons})^2$	0.0200**	0.0425***	72.54
Spill site characteristics	pop800	number of people	0.0000876	0.000166***	500
	watwet800_bi	true/false	1.520**	0.405***	0
	spatial_missing	true/false	-0.885*	-1.353***	0
	et800_bi	true/false		0.500	0
	lu_hddev_800	% high density development		0.723*	0.4
	hca_bi	true/false		0.372***	1
Spill circumstances	fire_bi	true/false		2.449***	1
	underground_bi	true/false		1.059***	0
	rcause_derail	true/false	1.742***		
	rcause_component	true/false	-0.200		
	rcause_aging	true/false	-0.302*		
	pcause_corrosion	true/false		0.191*	1
	pcause_natlforces	true/false		0.530**	0
		Observations	340	2,465	
		R^2	0.785	0.498	
		Root MSE	0.878	1.525	1.525
A. Sum model coefficient times incident value					14.42
B. $(\text{Root MSE})^2/2$					1.17
C. Total cost $[e^{(A+B)}]$					$5,872,506
D. Unit cost [C/spill volume]					$1,175/gallon

Note: The dependent variable is the log of total cost. Each cell in the table shows a regression coefficient. Significance is denoted by asterisks: * p<0.05, ** p<0.01, *** p< 0.001.

Uncertainty and Limitations

The cost models provide a way to value damages, conditional on an incident occurring. To use the models, therefore, one needs to have information, actual or modeled, on incidents that are expected to occur in the baseline, and how these incidents would change, with respect to the cost model variables, as a result of the policy intervention. Development of the baseline and post-policy incident populations is beyond the scope of this study; in general, however, the incident population can be developed based on historical data (in which case the analysis implicitly assumes that history is a reasonable predictor of future conditions) by making assumptions regarding the likelihood of various representative incidents, using outputs from a fault-tree analysis, or other approaches.

The cost modeling methodology relies on empirical data and therefore inherits the uncertainty and limitations of the data in terms of the characteristics of the reported incidents or their costs. Because the reported costs do not capture all third-party costs or spill impacts, the modeled expected cost may understate actual costs especially for incidents with far reaching consequences. Conversely, to the degree that the incidents involve multiple events occurring concurrently with the crude oil release, the modeled expected costs may include damages that are not a direct result of the spill. Furthermore, factors not captured in the incident data and excluded from the cost models, such as response actions, could affect costs in either direction (e.g., increasing direct response costs while reducing overall damages and impacts). See Section 3 for a detailed discussion of the uncertainty and limitations.

1 Analysis Framework and Modeling Approach

1.1 Background

Analyses of federal regulations to protect and improve health, safety, and the environment are challenging, particularly when they involve estimating the benefits of avoiding relatively infrequent and far-reaching adverse outcomes. Analytical challenges in estimating the benefits of preventing oil spills include understanding the factors that determine the probability and size of a spill and the severity of the resulting damages. They also include the difficulty of fully quantifying, and then predicting, spill impacts, especially in monetary terms. With these challenges in mind, data from past incidents provide an empirical basis for quantifying damages from past spills and for estimating the potential benefits of preventing similar releases in the future.

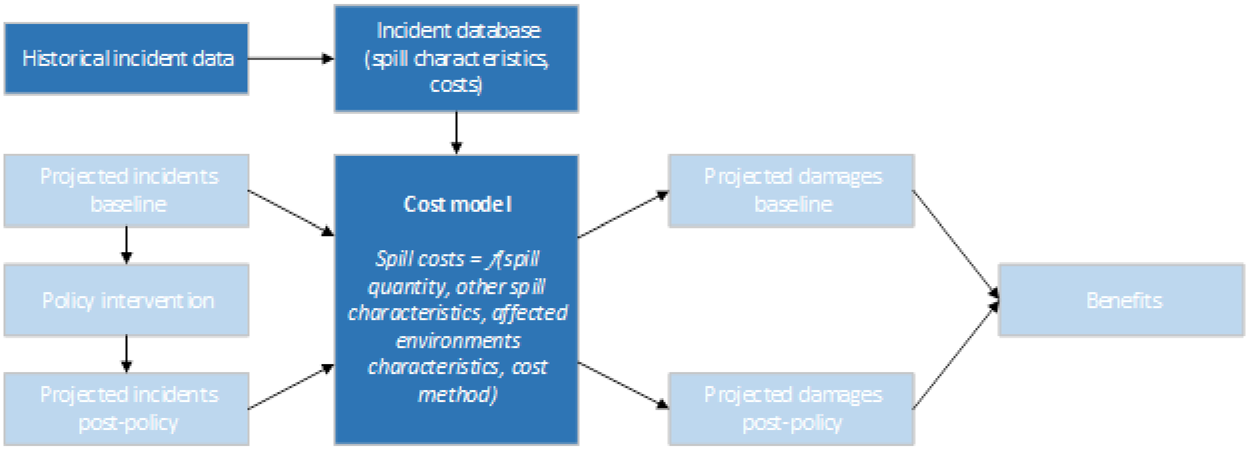
In analyzing the benefits of spill prevention regulations, PHMSA has typically relied on relatively simple measures of damages, such as cleanup and other costs incurred by operators per gallon of oil spilled. Not all spills are the same, however, and a regulation may address only a subset of spill circumstances. Furthermore, spill risk may change over time as more stringent safety measures are implemented. The ability to disentangle how different factors affect the probability of a spill and the resulting damages is an important step in improving PHMSA's ability to estimate benefits of its regulations. This study contributes to this effort by focusing on how PHMSA quantifies and monetizes damages.

The study builds on a prior study completed in 2017, titled "Valuation of Crude Oil Spilled in Transportation Incidents" that developed the original analysis framework (Abt Associates, 2017). Appendix A includes the original report. For the current effort, PHMSA updated the dataset previously compiled for the 2017 study to include more recent incidents. PHMSA then applied similar statistical methods as in the 2017 effort to update the empirical relationships and provide a model that can be used to estimate oil spill costs. The rest of this report highlights the key aspects of the incident data compilation and modeling approach and presents the updated relationships. For additional details, including a review of relevant literature on the valuation of crude oil release valuation, see Appendix A.

1.2 Conceptual Framework

Figure 1-1 illustrates the overall benefits analysis framework and how PHMSA may eventually use the elements addressed in this study and highlighted in the darker boxes in the diagram of Figure 1-1, to estimate the benefits of preventing crude oil spills. Specifically, the cost model developed by this study could be used to estimate the damages (costs) associated with incidents projected in the baseline (i.e., business-as-usual scenario in the absence of any intervention), and those associated with incidents projected following a policy intervention. The difference between costs projected under the two scenarios then represents the benefits of the policy intervention.

Figure 1-1: Overall benefits analysis framework with darker boxes highlighting the elements addressed in the study.



The cost model is designed to account for the interplay of factors that determine the magnitude of a crude oil release and the resulting damages and social costs. These factors include:

- **Source characteristics**, such as the transportation mode (pipe/rail), physical characteristics of the container such as the age, diameter, construction material, operating pressure, and characteristics of the transported product (e.g., viscosity);
- **Risk factors**, which are external factors such as environmental conditions (e.g., temperature and precipitation), operating procedures, and the speed at which the rail car is traveling, which may affect the probability and/or the consequences of an incident;
- **Incident causes**, which may lead to different failure modes and impacts;
- **Characteristics of the affected environment**, such as the affected media (e.g., water, soil), population density, proximity to sensitive ecological receptors; and
- **Effectiveness of the response**, which relates to actions that can limit the actual quantity released during the incident, contain the spread of the oil, and reduce resulting damages.

This study seeks to determine the relationship between the social costs of an incident and the multitude of factors that may contribute to the magnitude of these costs. We note that while many of these factors may also affect incident probabilities, the effort focuses strictly on how the factors influence the resulting costs of an incident, once it occurs. This is one of many steps in analyzing the benefits of regulations or other policies to improve the safety of the transportation of crude oil by rail or pipeline. The social costs of a release include those costs incurred by the operator (e.g., property damage, product loss, cleanup, reimbursement of government expenses, and third-party damages paid by the operator), as well as other costs to society (e.g., injuries, fatalities, human health effects, damage not compensated by the operator, natural resources damages, travel delays).[1]

1.3 Incident Data

1.3.1 Data Sources and Data Compilation Method

Section 3 in Abt Associates (2017) describes the data sources and data compilation methodology used to assemble the original crude oil spill incident database, while Section 4 describes the database in detail. The descriptions in Sections 3 and 4 of Abt Associates (2017) were based on pipeline and rail incidents for years 2005-mid-2016.[2]

For this updated analysis, PHMSA expanded the temporal scope of the incident data to the end of 2022 using PHMSA's pipeline and rail datasets.[3] The update added new incidents that occurred after June 2016 and added or revised data for incidents that occurred during the original dataset's timeframe.[4] The update also involved mapping the more recent data, which use a slightly different format due to changes in PHMSA's flagged incident and rail datasets, to the original database fields.[5]

Following the approach in Abt Associates (2017), PHMSA used several additional sources to supplement the primary PHMSA datasets, including updated value of statistical life data from the U.S. Department of Transportation (U.S. DOT, 2022), data on total crude oil transported via rail from the U.S. Energy Information Administration (U.S. EIA, 2023), and supplemental information available about specific oil spills from the National Transportation Safety Board (2017a, 2017b, 2021), the U.S. Environmental Protection Agency (U.S. EPA, 2016a, 2016b), and industry sources (Enbridge Inc., 2017, 2018; Plains All American Pipeline, 2020).

PHMSA also calculated spatial variable values for each incident that occurred after June 2016 and, as needed, revised values for incidents prior to June 2016. The spatial variables and

---

1 Some costs reported by operators, such as fines and penalties levied against the responsible party, are not social costs from a societal perspective but instead represent transfers to the collecting agencies. Additionally, whereas a crude oil release may temporarily reduce tourism in a particular community, it may increase tourism elsewhere as consumers shift their expenditures. In such cases, the net effect at a regional or national level is zero, even if local economic impacts are negative (or positive).

2 For the database described in Abt Associates (2017), the PHMSA pipeline and rail incident datasets were retrieved on 11/9/2016.

3 For the update, the PHMSA pipeline and rail incident datasets were retrieved on 2/20/2023 from: https://www.phmsa.dot.gov/data-and-statistics/pipeline/distribution-transmission-gathering-lng-and-liquid-accident-and-incident-data.

4 Nine additional pipeline incidents were reported in 2014 and one additional rail incident was reported in 2015. These differences are due to revisions to the datasets as new information became available.

5 This step required reorganizing the more recent data to match the old format, including reordering columns, renaming certain variables to their old names, creating blank variables for variables that are no longer recorded, and duplicating variables when multiple variables were combined into one. ICF created R scripts that perform the data reorganization steps for future reproducibility. The data formatting changes in recent years did not affect database fields critical to the cost modeling described in Section 2, such as fields related to costs and quantity released.

methodology are described in detail in Section 2.4 of Abt Associates (2017). For the latest update, PHMSA used the following data sources and methodology:

- **Census block-based population variables (BLK_POP10_300, BLK_POP10_800)**: Calculated values for incidents that occurred after June 2016. PHMSA continued to use 2010 Census data for consistency with other observations in the dataset and because block-level data products from the 2020 Census were not available at the time of the database updates. For this analysis, PHMSA downloaded state-level block shapefiles that contained 2010 Census population data (https://www.census.gov/geographies/mapping-files/2010/geo/tiger-data.html).
- **Land cover-based variables (LU_WATWET_300, LU_HDDEv_300, LU_ODEv_300, LU_WATWET_800, LU_HDDEv_800, LU_ODEv_800)**: Calculated landcover proportion values for *all* database observations, using 2006 National Land Cover Database (NLCD) data for incidents in years 2005-2010, 2011 NLCD data for incidents in years 2011-2015, and 2016 NLCD data for incidents in years 2016-2022. PHMSA did not use 2019 NLCD data because it is only available for conterminous United States (i.e., not for Alaska, where several new incidents occurred). PHMSA downloaded the 2006, 2011, and 2016 NLCD from https://www.mrlc.gov/data.
- **Threatened and endangered species variables (ET_300, ET_800)**: Calculated values for incidents that occurred after June 2016 using a dataset that aggregates critical habitat for all species into two shapefiles, one line and one polygon. The dataset, which was last updated on 2/20/23, was downloaded from https://ecos.fws.gov/ecp/report/table/critical-habitat.html.

1.3.2 Incident Data Summary

Table 1-1 summarizes the year and transportation mode for the 3,734 onshore incidents that reported release of crude oil from pipelines or railcars between January 2005 and December 2022. The updated data produced an additional 1,215 pipeline incidents and 32 rail incidents relative to the database summarized in Abt Associates (2017).

Table 1-1: Number of onshore incidents by year and transportation mode.

Incident Year	Pipeline Releases	Rail Releases	Total
2005	163	2	165
2006	157	1	158
2007	157	1	158
2008	151	8	159
2009	151	1	152
2010	150	9	159
2011	138	34	172
2012	184	88	272
2013	202	118	320
2014	235	144	379
2015	251	43	294
2016	201	6	207
2017	205	6	211
2018	215	6	221
2019	192	7	199
2020	165	5	170
2021	179	1	180
2022	154	4	158

Table 1-1: Number of onshore incidents by year and transportation mode.

Incident Year	Pipeline Releases	Rail Releases	Total
Total	**3,250**	**484**	**3,734**

Note: Includes all onshore pipeline and rail incidents (identified in PHMSA datasets) that involved the release of more than 0.0 gallons of crude oil and occurred between 01/01/2005 and 12/31/2022.

Table 1-2 summarizes spill size and cost information for all onshore crude oil spill incidents, including the number of incidents for each spill size category, total quantity released, and total costs.

Table 1-2: Number of incidents, quantity released, and costs of onshore incidents by spill size.

Spill Size (Gallons)	Number of Incidents	Total Gallons Released	Total Cost (Millions 2022$)
All	3,734	42,280,005	$4,563.2
>100,000	65	32,784,814	$3,381.4
50,000 to 99,999	41	2,790,494	$333.5
10,000 to 49,999	202	4,478,014	$387.4
1,000 to 9,999	532	1,841,479	$179.4
500 to 999	238	171,061	$81.4
100 to 499	697	158,455	$74.7
50 to 99	373	29,397	$22.1
5 to 49	1,144	25,687	$63.9
<5	442	604	$39.3

Note: Includes all onshore pipeline and rail incidents (identified in PHMSA datasets) that involved the release of more than 0.0 gallons of crude oil and occurred between 01/01/2005 and 12/31/2022.

Of the 3,734 incidents in the database with non-zero quantity released, 3,581 incidents (including 340 rail incidents and 3,241 pipeline incidents) reported some costs associated with the release. The remaining 153 incident reports did not include any costs, even though they did report non-zero quantity released. We exclude the 153 zero-reported-cost incidents from the remainder of this analysis.

Table 1-3 summarizes medians of the gallons spilled, total costs, and unit costs per gallon by incident cause. Most of the pipeline incidents (39 percent) were caused by equipment failure, with median total costs of $12,822. The largest median total costs for pipeline spills, $117,612, were associated with material failure, but incidents caused by material failure were relatively infrequent, representing 3.4 percent of all pipeline incidents. Regarding rail incidents, incidents caused by derailment, while relatively infrequent (6.6 percent of all rail incidents) had by far the largest median total costs, at $941,646, as well as much greater median gallons spilled (13,937 gallons). The remaining rail incidents tend to involve small quantities of crude oil and total costs less than $3,000.

Table 1-3: Median impacts by cause categories for pipeline and rail spills.

Cause Type	Count	Median Gallons Spilled	Median Total Cost	Median Unit Cost per Gallon
Pipeline Incidents				
Corrosion	979	252	$46,593	$148
Equipment Failure	1,265	70	$12,822	$136
Excavation Damage	118	4,515	$107,559	$33
Incorrect Operation	427	126	$14,568	$92
Material Failure	179	168	$117,612	$273
Natural Forces	136	126	$50,773	$194
Other Outside Forces	61	420	$67,750	$101
Other Incident Cause	76	252	$37,539	$147

Table 1-3: Median impacts by cause categories for pipeline and rail spills.

Cause Type	Count	Median Gallons Spilled	Median Total Cost	Median Unit Cost per Gallon
Total Pipeline	**3,241**	**126**	**$25,668**	**$134**
Rail Incidents				
Derailment	22	13,937	$941,646	$69
Loose, Missing, or Broken Component	175	2	$2,549	$1,570
Deterioration or Aging	31	1	$2,281	$2,513
Other Human Error	106	1	$2,850	$1,652
Total Rail	**340**	**2**	**$2,862**	**$1,566**

1.4 Modeling Approach

The modeling approach used to predict the total costs for pipeline- and rail-related onshore crude oil spills follows that presented in Abt Associates (2017). PHMSA estimated separate mode-specific models for releases from pipelines and railcars. The models used the same functional form and explanatory variables as in Abt Associates (2017) to estimate the relationship between incident characteristics and total costs. Refer to Section 5 in Abt Associates (2017) for a more detailed discussion regarding the development of the regression models.

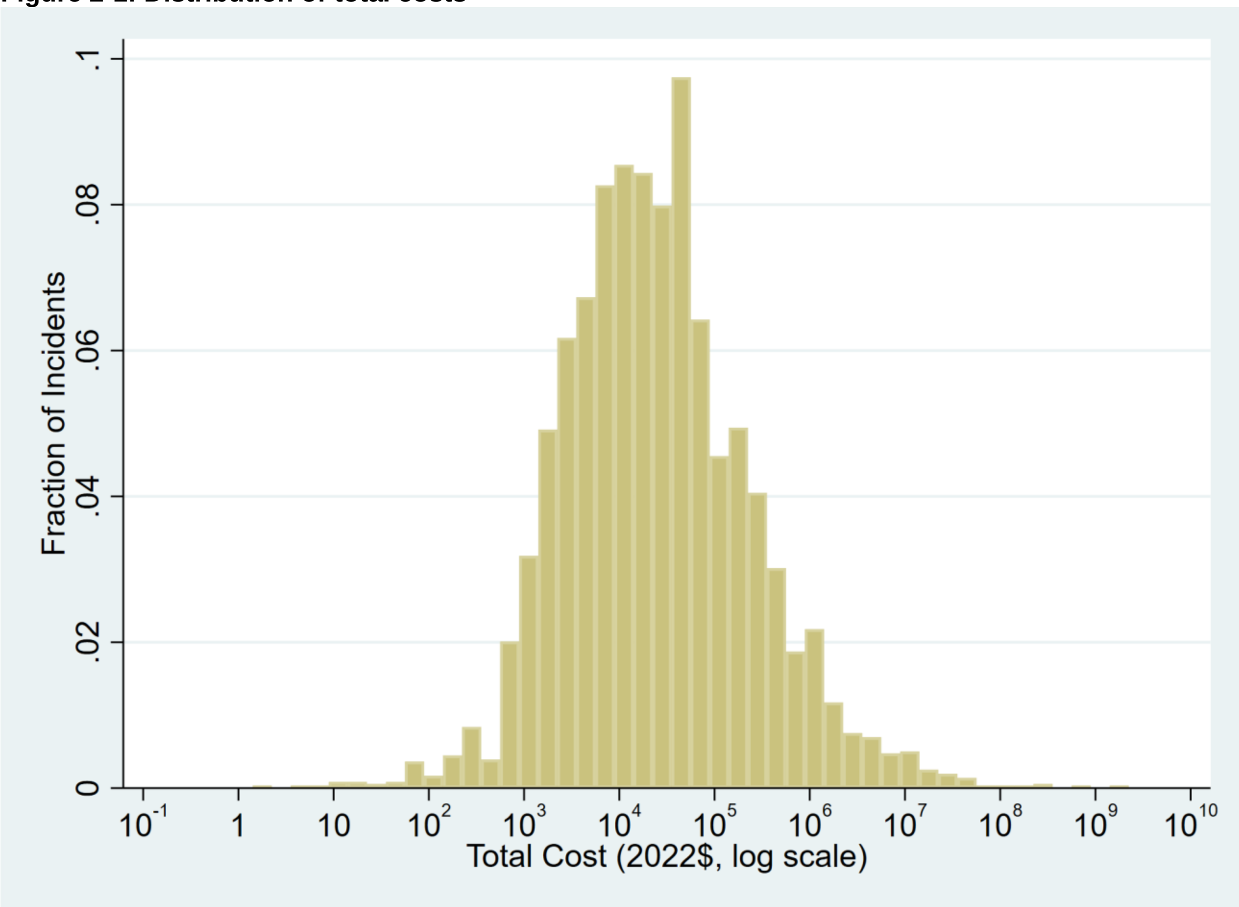
2 Crude Oil Spill Cost Modeling

2.1 Factors Affecting Crude Oil Spill Costs

This section presents costs graphically in relation to incident characteristics to provide insight into the effect of each characteristic on cost. The section begins by presenting basic information about the distribution of costs across incidents and continues by presenting a series of figures that explore how the distribution varies for different subsets of incidents (e.g., pipeline and train incidents).

Figure 2-1 presents a histogram of total costs and shows how the total costs of spills vary across all incidents in the database. The horizontal axis represents the total cost of spills (using a log scale) and each vertical bar represents the fraction of incidents associated with a particular range of costs. As shown in the histogram, total costs are right-skewed, with a few incidents having very high total costs. Total costs are expected to increase with the quantity of oil released (and a larger extent of contamination), greater response costs, and other factors.

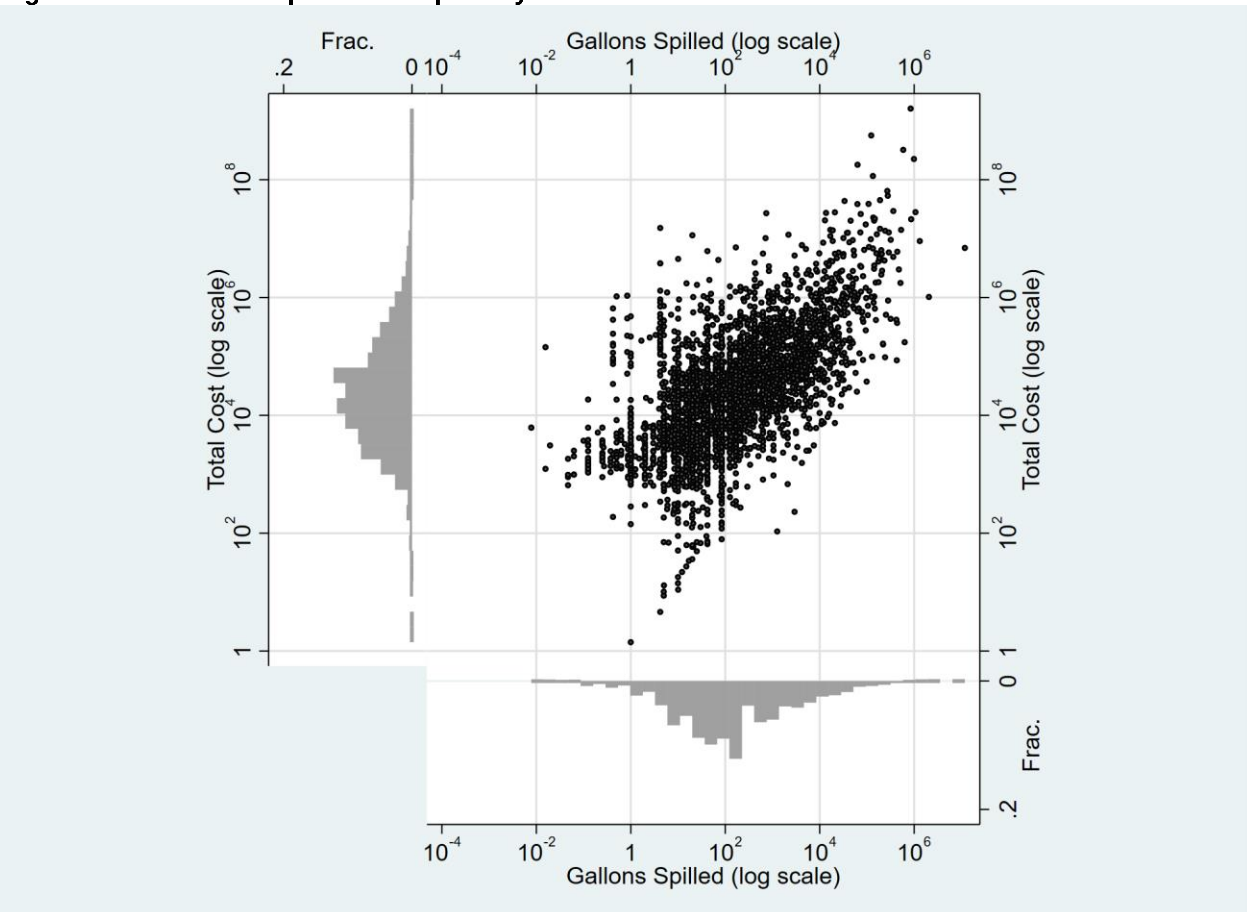
Figure 2-1: Distribution of total costs



Note: Includes 3,581 incidents with greater than zero quantity released and total costs.

The scatterplot in Figure 2-2 shows the relationship between quantity released and total costs. Each point in the scatterplot represents an incident. The figure also presents histograms of the total costs and quantity released. The figure shows a positive correlation between total costs and quantity released, but there is some noise in the data wherein some incidents with very low volume of oil released have relatively high costs and vice versa. As shown in the histograms, most incidents cost between $1,000 and $1,000,000 and involve between 10 and 1,000 gallons.

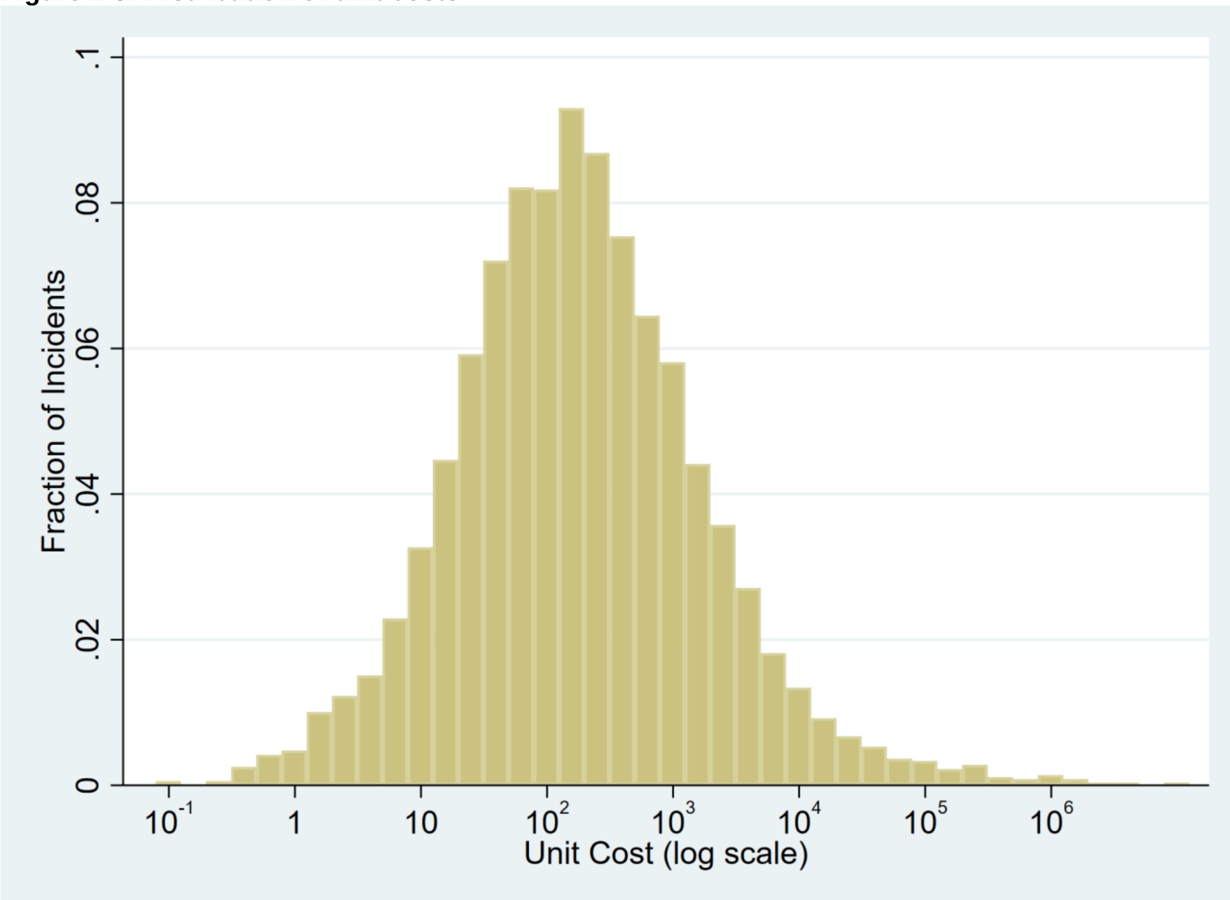
Figure 2-2: Relationship between quantity released and total costs



Note: Includes 3,581 incidents with greater than zero quantity released and total costs.

Figure 2-3 presents a histogram of unit costs, with the same interpretation as Figure 2-1. Unit costs are approximately normally distributed with skewness to the right (i.e., few incidents associated with large unit costs). The summary statistics confirm this: the median unit cost is $162 per gallon as compared to a mean unit cost of $10,215 per gallon.

Figure 2-3: Distribution of unit costs



Note: Includes 3,581 incidents with greater than zero quantity released and total costs.

Table 2-1 further investigates the relationship between unit costs and quantity released. The table compares the average and median of incident unit costs for all spills, spills of 10 or more gallons, spills of 42 or more gallons (i.e., one or more barrels), spills of 1,000 or more gallons, and spills greater than 50,000 gallons. The results are shown for all types of incidents, and, separately, for rail and pipeline incidents. The table shows that the average unit costs of incidents decline as the quantity released increases. The average spill has a unit cost of $10,215, based on all incidents, whereas for spills greater than 1,000 gallons, the average unit cost is only $116. This pattern of lower unit costs for spills involving larger volumes generally holds for both pipeline and rail incidents. One exception is the subset of pipeline incidents involving more than 50,000 gallons, which have higher unit costs ($158 per gallon).

Table 2-1: Average unit costs for incidents based on quantity released

Release Quantity	Number of Incidents	Average Unit Costs	Median Unit Costs
All Incidents			
Any release quantity	3,581	$10,215	$162
10 or more gallons	3,069	$1,163	$118
42 or more gallons	2,392	$526	$82
1,000 or more gallons	839	$116	$26
50,000 or more gallons	106	$149	$17
Pipeline Incidents			
Any release quantity	3,241	$6,688	$134
10 or more gallons	2,998	$1,184	$117
42 or more gallons	2,352	$532	$83
1,000 or more gallons	811	$116	$26
50,000 or more gallons	98	$158	$17
Rail Incidents			
Any release quantity	340	$43,835	$1,566
10 or more gallons	71	$268	$154
42 or more gallons	40	$169	$65
1,000 or more gallons	28	$110	$58
50,000 or more gallons	8	$37	$33

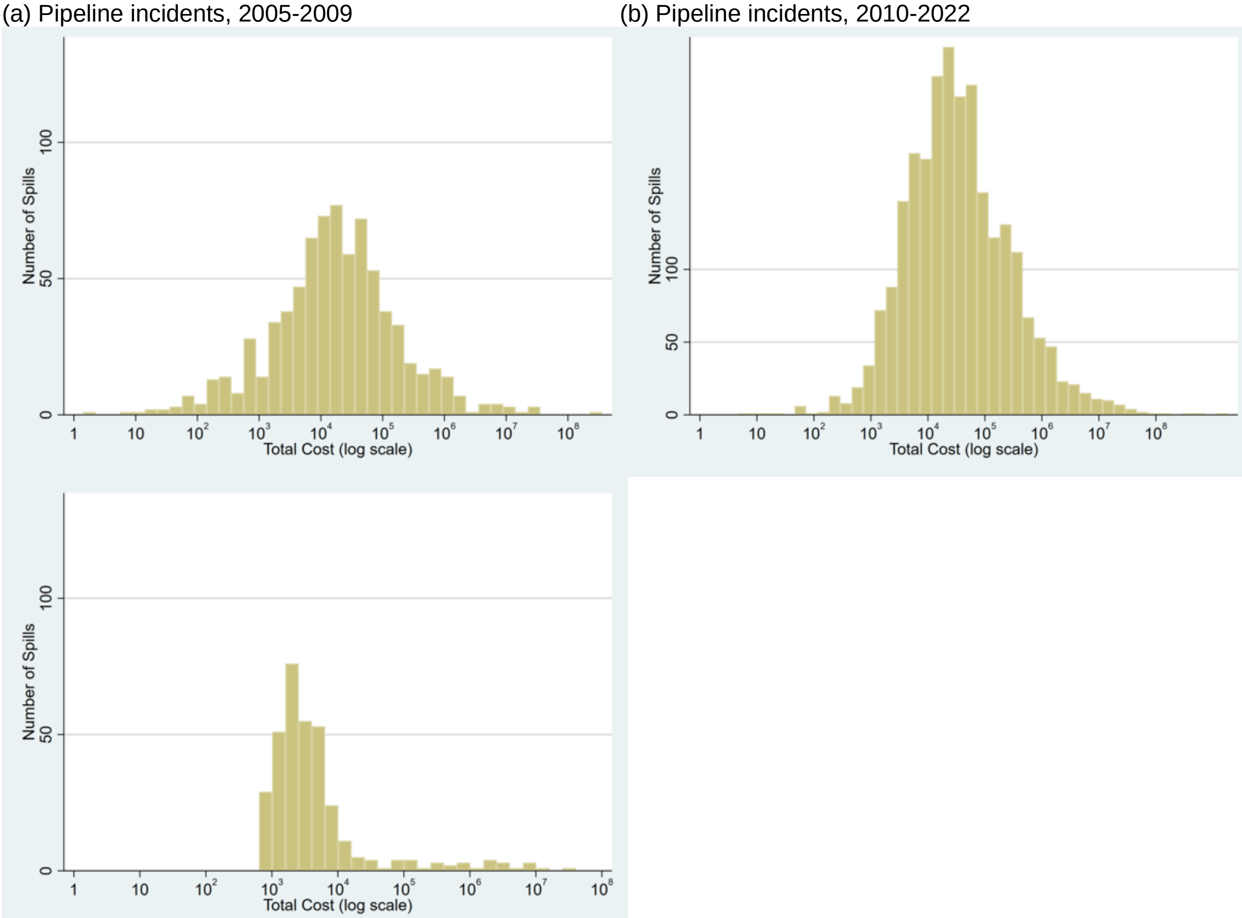
The remainder of this section explores some of the factors contributing to total incident costs (in contrast to the prior section which focused on unit costs). Figure 2-4 presents histograms of total costs by transportation mode. Total costs are generally lowest in the 2005-2009 pipeline dataset and highest in the 2010-2022 pipeline dataset.[6] The figure shows that the rail costs are relatively more skewed compared with the pipeline incident costs, with fewer incidents below $1,000 in total costs and a larger number of higher-cost incidents.

---

6 As explained further in Abt Associates (2017), the format of the pipeline data changed in 2010, making direct comparisons between incidents before and after that change difficult. The database therefore differentiates between incidents that occurred in 2005-2009 and those that occurred after 2010. There is no need to separate out rail incidents by period since the format of these data remained consistent. Further, there were relatively few rail incidents involving crude oil releases reported before 2010. Of the 315 incidents in the database, only 10 occurred before 2010.

Figure 2-4: Distribution of total costs, by transportation mode

(a) Pipeline incidents, 2005-2009 (b) Pipeline incidents, 2010-2022

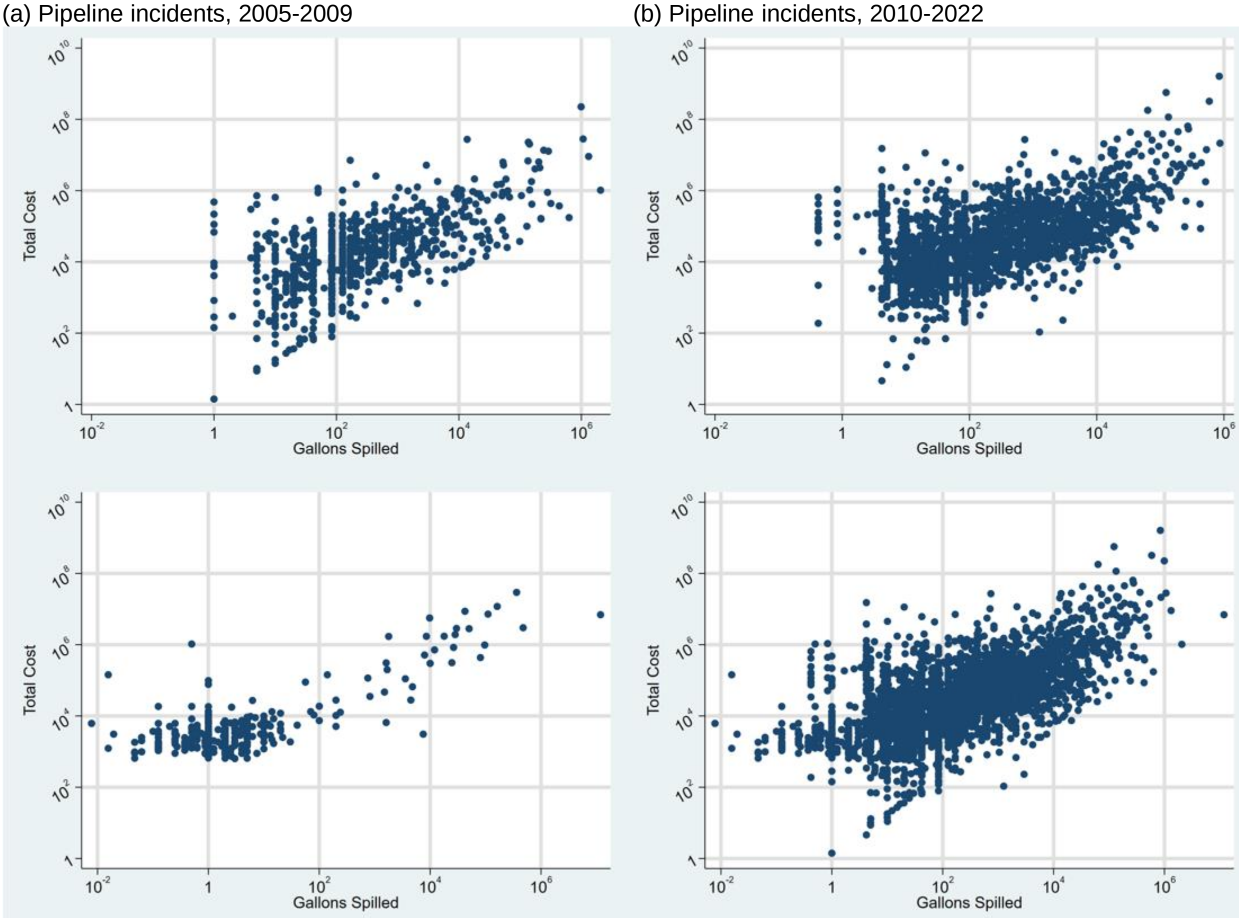


(c) Rail incidents

Figure 2-5 presents scatterplots comparing total costs to the quantity of oil spills. The figure provides additional information about the influence of quantity released and transportation mode on total costs. The four panels in the figure show the relationship separately for pipeline incidents in 2005-2009, pipeline incidents in 2010-2022, rail incidents, and all incidents combined. All four panels show total costs increasing with quantity released. Comparing panels (a), (b), and (c) shows that the slope of this relationship appears relatively consistent across the datasets. However, for smaller volume spills (below about 100 gallons) rail incidents tend to have higher total costs, while pipeline incident costs are more variable in that range. Additionally, relative to pipeline incidents, there are more rail incidents with very small (under a gallon) quantities released.

Figure 2-5: Quantity released and total costs

(a) Pipeline incidents, 2005-2009 (b) Pipeline incidents, 2010-2022



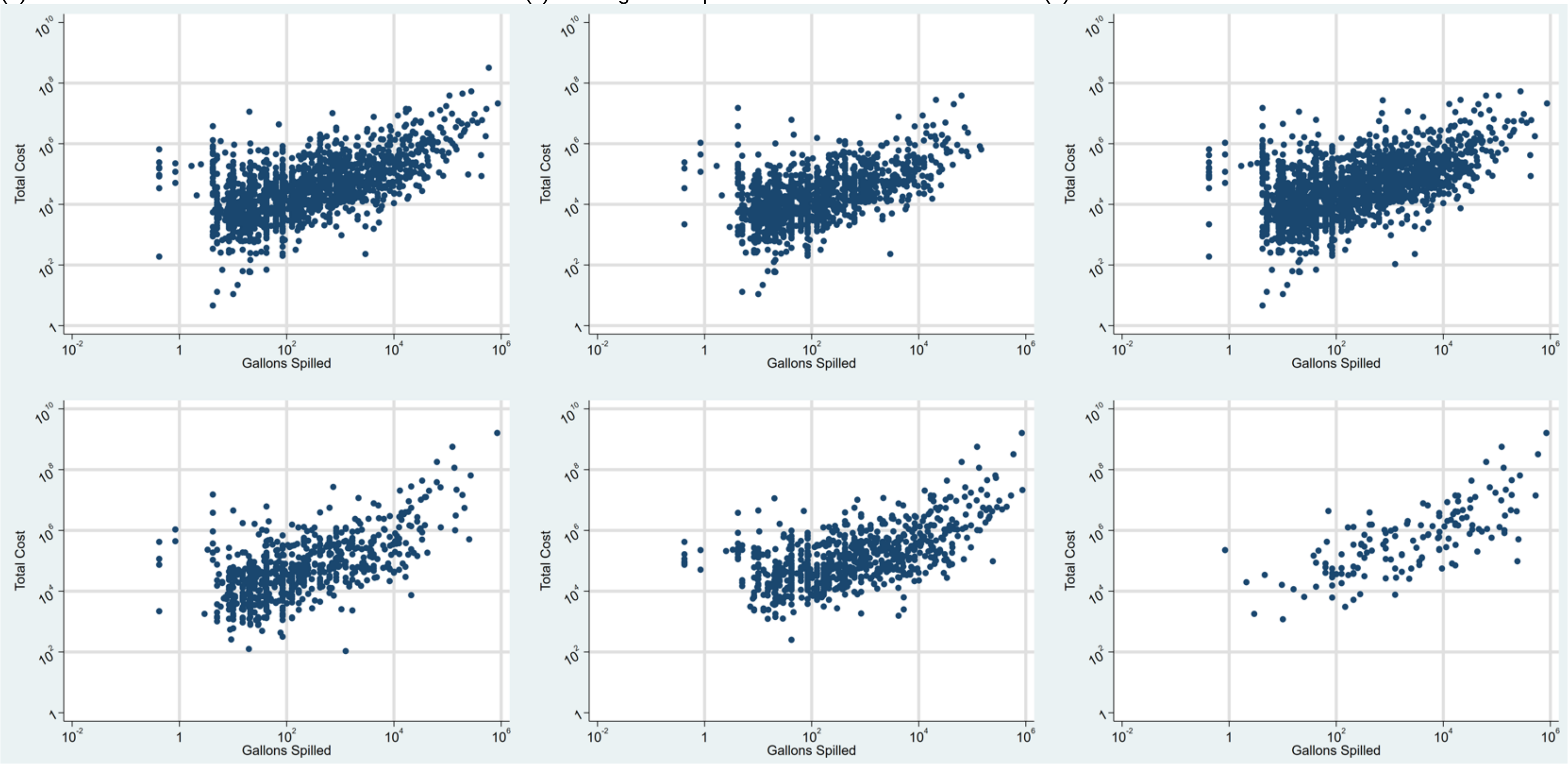
(c) Rail incidents (d) All incidents

Many other potentially relevant factors may influence total incident costs such as the size of the population living within the vicinity of the spill, and whether the spill occurs near surface water. Figure 2-6 shows how the relationship between costs and spill quantity varies across three categorical variables: (1) whether the incident occurred within or outside of a high consequence area (HCA), (2) whether the release occurred above or below the surface, and (3) whether the release affected surface water. To control for differences in outcomes and dataset completeness, the figure shows results only for pipeline incidents from 2010 to 2022.

Panels (a) and (b) present the relationship between total costs and quantity for incidents that are located outside and inside HCAs. The figure shows that HCA and non-HCA incidents have similar costs with the exception of incidents involving larger spills, where HCA-incidents appear to have higher costs. Panels (c) and (d) present a similar comparison for incidents that involve aboveground and underground releases. The figure shows that incidents involving underground spills are associated with higher costs than aboveground spills. Lastly, panels (e) and (f) present a comparison of releases that did and did not affect surface water. The figure shows that incidents affecting surface water contamination have higher costs than incidents that do not. Additionally, the differences in costs are most pronounced for higher quantity releases.

Figure 2-6: Quantity released and total costs, for pipeline incidents 2010-2022, by category

(a) Outside HCA area (c) Aboveground spill (e) Surface water not affected



(b) Inside HCA area (d) Underground spill (f) Surface water affected

2.2 Regression-Based Analysis of Predictors of Incident Costs

PHMSA used a regression-based analysis to examine the relationship between incident costs and potential influential characteristics, including the transportation mode, volume of crude oil released, and incident location. The model predicts per incident cost as a function of the incident characteristics and has the following general form:

$$LogCost_i = f(SpillChar_i, SiteChar_i, CostMethod_i)$$

where

- $LogCost_i$ is the log of total cost of incident i.
- $SpillChar_i$ is a vector that describes the characteristics of the spill for incident i. This includes variables such as the quantity released, the spill source (pipeline vs. rail), the spill mechanism (derailment vs. seepage), and other spill characteristics.
- $SiteChar_i$ is a vector that describes the characteristics of the spill site for incident i. This includes variables such as the time of year, the population within the block group or within 800 m of where the spill occurred, proximity to water, whether the spill occurred in an HCA, etc.
- $CostMethod_i$ is a vector that describes the methodology used to estimate spill costs for incident i. For example, this might include whether or not particular categories of costs (e.g., commodity lost, operator damages) were accounted for in the cost estimate.

Following the approach from the 2017 study, PHMSA specified separate regressions for rail and pipeline incidents. Results from these regressions are presented in Table 2-2. The regression uses an ordinary least squares (OLS) regression approach. This means that the function $f()$ has a linear functional form. Using an OLS approach is standard in the literature and provides considerable flexibility in model specifications. For example, based on the relationship apparent in Figure 2-1, the logarithm of the cost appears to be the most appropriate dependent variable for the model. This approach can also accommodate the use of interactions between different explanatory variables—e.g., an interaction between the quantity of oil released and the transportation mode or categories of costs with non-zero values.

Table 2-2: Regression-based analysis of total costs

Variable	Regression Specification	
	Rail	Pipeline[1]
intercept	7.336*** (0.556)	8.335*** (0.181)
ln_qreleased	0.162*** (0.0347)	-0.0454 (0.0621)
ln_qreleased_sq	0.0200** (0.00653)	0.0425*** (0.00488)
pop800	0.0000876 (0.000240)	0.000166*** (0.0000375)
watwet800_bi	1.520** (0.539)	0.405*** (0.0622)
spatial_missing	-0.885* (0.353)	-1.353*** (0.102)
et800_bi		0.500 (0.256)
lu_hddev_800		0.723* (0.369)
hca_bi		0.372*** (0.0766)
fire_bi		2.449*** (0.473)
underground_bi		1.059*** (0.0831)
rcause_derail	1.742*** (0.518)	
rcause_component	-0.200 (0.105)	
rcause_aging	-0.302* (0.134)	

Table 2-2: Regression-based analysis of total costs

Variable	Regression Specification	
	Rail	Pipeline[1]
pcause_corrosion		0.191* (0.0826)
pcause_natlforces		0.530** (0.184)
Observations	340	2,465
R^2	0.785	0.498
Root MSE	0.878	1.525

Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.
[1] Based on data from pipeline incidents between 2010 and 2022.

In general, the statistical fit of the estimated models is good. Model results suggest that a significant systematic component of total cost variation is associated with transportation mode, spill, and location characteristics. The rail model has an R^2 of 0.785 while the pipeline model has an R^2 of 0.498. These R^2 values indicate that the models are respectively predicting 79 percent and 50 percent of the variation in total costs.

The coefficient values of the intercept terms can, roughly, be interpreted as the expected natural log of total cost for spills that are one gallon in size for incidents within the transportation mode. For example, the coefficient of 8.335 on the pipeline model intercept indicates that the natural log of total costs is expected to be 8.335 for a pipeline spill of one gallon; Hence, the total cost of a pipeline spill of one gallon is $4,167.

The variables *ln_qreleased* and *ln_qreleased_sq* represent the natural log of the quantity released and the natural log of the quantity released squared, respectively. Together these two variables capture the influence of spill size on incident costs. In both models, the coefficients for the quadratic term (*ln_qreleased_sq*) is positive and significant which indicates that incident costs increase with quantity, but do so at a decreasing rate (i.e., the marginal cost per gallon spilled decreases as the spilled volume increases).

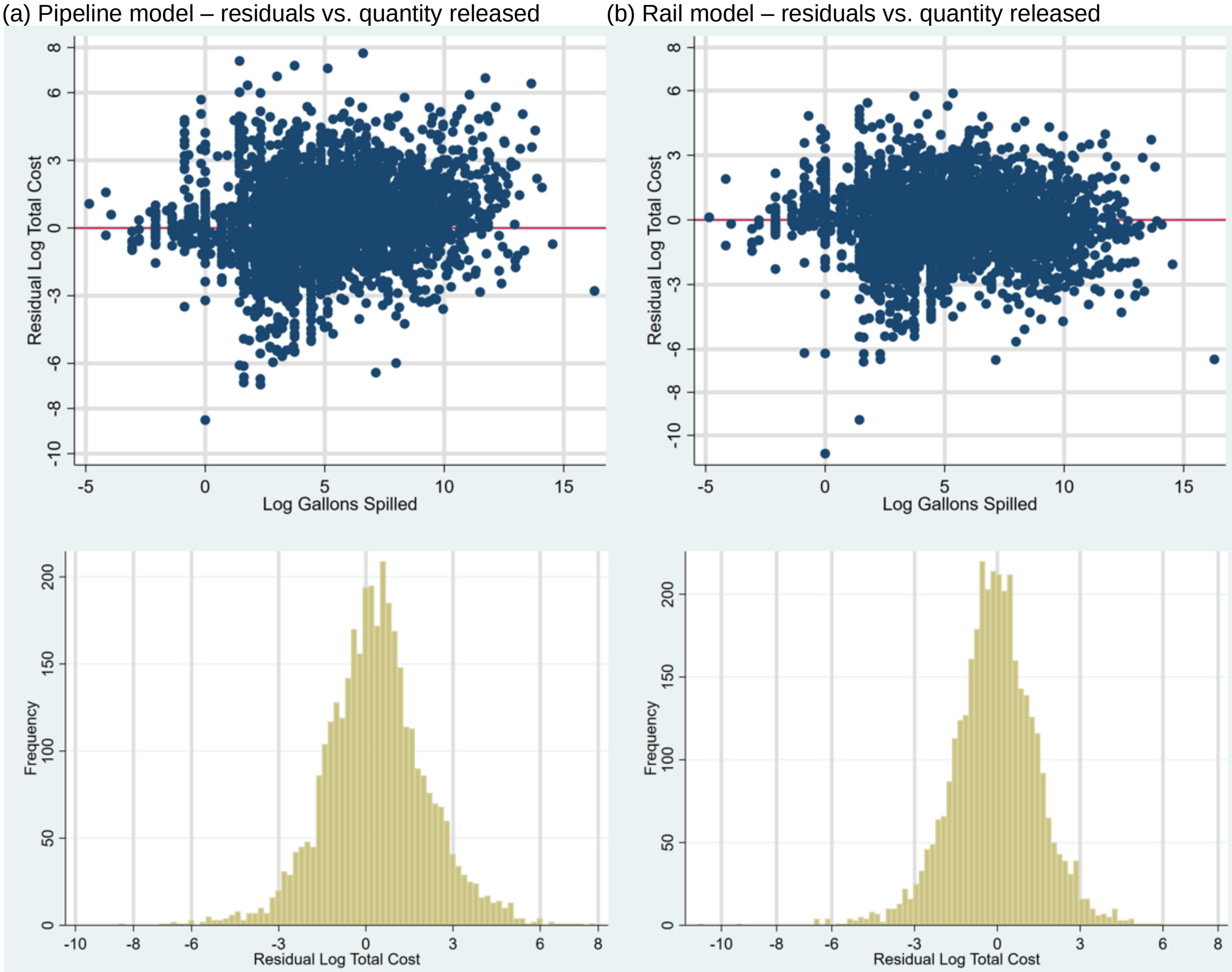
The pipeline model shows that costs are statistically higher in HCAs (*hca_bi*), in areas with surface waters or wetland present within 800 meters of the incident (*watwet800_bi*), in areas with high intensity development within 800 meters (*lu_hddev_800*), in incidents that involve fires (*fire_bi*), and in incidents that originate underground (*underground_bi*). The model also shows incident costs increasing where there are more people living within 800 meters of the incident (*pop800*), and when the cause of the release is categorized as “corrosion” (*pcause_corrosion*) or “natural forces” (*pcause_natlforces*).

The rail model shows that incidents involving derailments (*rcause_derail*) have significantly and substantially higher costs than other types of releases from railcars. And, in contrast, the model shows incident costs decreasing when the cause of the release is categorized as “aging” (*rcause_aging*). Lastly, the rail model also shows that costs are statistically higher when incidents occur in an area where surface waters or wetlands are present.

Figure 2-7 presents graphs that evaluate the robustness of the regression results from Table 2-2. Panels (a) and (b) present graphs of the model residuals against the quantity released. The panels confirm that the models fit the data well, since the residuals do not exhibit a pattern and are balanced around zero. However, both panels show some signs of heteroskedasticity (i.e., relationship between the variability of residuals and spill size) for incidents associated with low quantities released. Additionally, panels (c) and (d) present the distribution of residuals from each model. One of the OLS regression assumptions is that the model residuals are normally distributed. Both panels show that the residuals resemble a bell shape and indicate that this is a reasonable assumption.

Figure 2-7: Regression diagnostics

(a) Pipeline model – residuals vs. quantity released (b) Rail model – residuals vs. quantity released



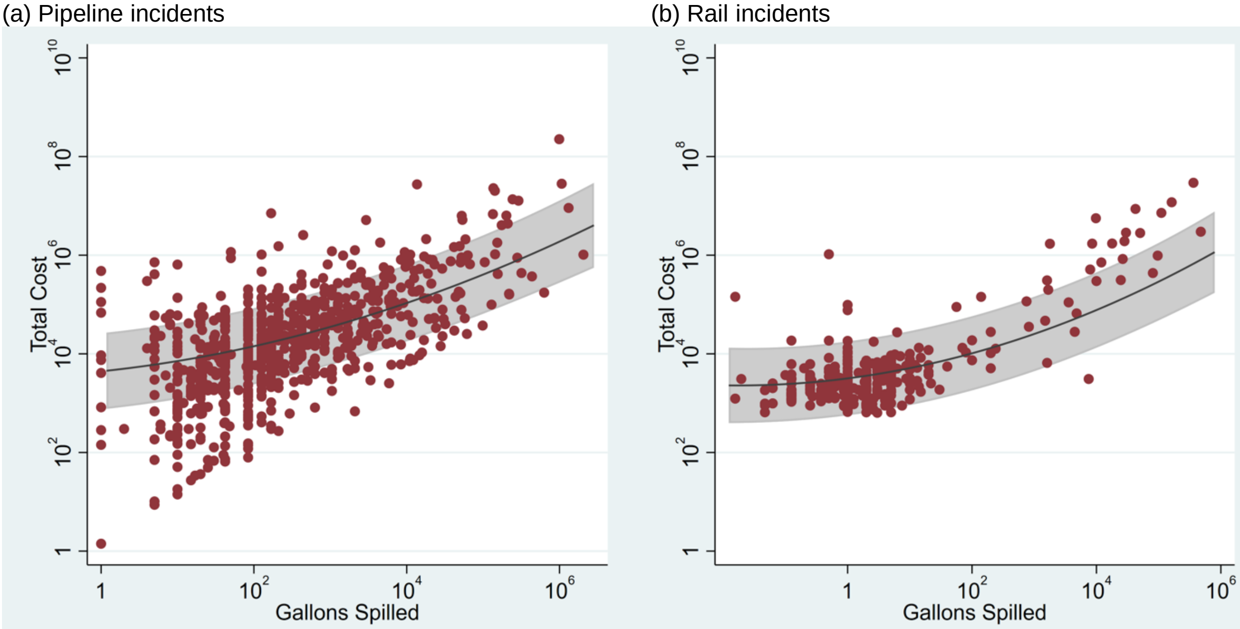
(c) Pipeline model – distribution of residuals (d) Rail model – distribution of residuals

2.3 Predicting Costs of Incidents

This section demonstrates how well the regression models predict the costs of an incident, given the incident's characteristics. Figure 2-8 presents graphs of the predicted and actual costs of incidents for the regression models described in Section 2.2. In each panel, the points represent actual total costs for incidents in the dataset. The dark line shows the average predicted total costs as a function of quantity released while the gray region represents the 95% confidence band. When predicting the total costs, all other variables in the dataset were set to their average values for the specified transportation mode. Panel (a) shows that for pipeline spills, when the spill is greater than 100 gallons the total costs start to increase substantially as the quantity released increases. Panel (b) shows a similar pattern for rail incidents where total costs begin to substantially increase when the quantity released is above 10 gallons. The panel also shows that the model tends to underpredict the total costs of larger spills (greater than 100 gallons). However, of note is that the figures predict the costs of typical incidents in all variables (which, except for quantity released, are set to their averages). The regression models allow for the specification of values for all variables to better predict the cost of atypical incidents. As such, the regression models are better able to predict costs than the simplified examples in the figures might suggest.

Figure 2-8: Predicted vs. actual total costs, primary regression models

(a) Pipeline incidents (b) Rail incidents



Note: The central line in each graph is based on the regression equation from Table 2-2 assuming average values for all variables except quantity released.

Table 2-3 demonstrates how the regression model results from Table 2-2 can be used to calculate oil spill costs for incidents with specific characteristics. The table presents examples of modeled results for four spill scenarios, two using the rail model and two using the pipeline model.

For each scenario, the variables (e.g., spill volume, population density, etc.) are set based on the characteristics of the incident. The rows at the bottom of the table illustrate the steps in using the model to calculate the predicted total costs for each incident. The difference in costs across scenarios is not only due to the quantity released and transportation modes, but also to the site characteristics.

Scenarios 1 and 2 focus on rail-related incidents. Scenario 1 represents an incident involving a 100-gallon spill from a railcar that occurs within an 800-meter radius of 400 residents and near the presence of surface water. The model predicts the total costs of the incident to be $34,465.

Scenario 2 represents a 10,000-gallon spill resulting from a derailment. The incident involves only 250 people living within an 800-meter radius and without the presence of surface water nearby. The model predicts the total costs of the incident to be $319,391. Predicted costs are substantially larger when compared to Scenario 1 because of the difference in spill volume and the incident involving a derailment, both of which are associated with greater costs.

Scenarios 3 and 4 focus on pipeline-related incidents. Scenario 3 represents an incident involving a 100-gallon spill inside an HCA, originating underground. The incident occurs in an area within 800-meters of 50 residents, critical habitats for threatened or endangered species, wetlands present, and 7 percent high-density development. The model predicts the total costs of the incident to be $294,340. Scenario 4 represents a 5,000-gallon spill from a pipeline inside an HCA that resulted in a fire and was caused by corrosion. The incident occurs within 800-meters of 500 residents and is 40 percent high-density development. The model predicts the total costs of the incident to be $5,872,506. Predicted costs are substantially larger when compared to Scenario 3 because of the difference in spill volume, the presence of a fire, the relatively large percentage of high-density development affected, and the incident being caused by corrosion.

Table 2-3: Example application of regression models

Variable	Rail Incident Model			Pipeline Incident Model		
	Model Coefficient	Scenario 1 Values	Scenario 2 Values	Model Coefficient	Scenario 3 Values	Scenario 4 Values
intercept	*7.336*	1	1	*8.335*	1	1
ln_qreleased	*0.162*	4.61	9.21	*-0.0454*	4.61	8.52
ln_qreleased_sq	*0.02*	21.21	84.83	*0.0425*	21.21	72.54
pop800	*0.0000876*	400	250	*0.000166*	50	500
watwet800_bi	*1.52*	1	0	*0.405*	1	0
spatial_missing	*-0.885*	0	0	*-1.353*	0	0
et800_bi				*0.5*	1	0
lu_hddev_800				*0.723*	0.07	0.4
hca_bi				*0.372*	1	1
fire_bi				*2.449*	0	1
underground_bi				*1.059*	1	0
rcause_derail	*1.742*	0	1			
rcause_component	*-0.2*	0	0			
rcause_aging	*-0.302*	0	0			
pcause_corrosion				*0.191*	0	1
pcause_natlforces				*0.53*	0	0
Root MSE	*0.878*			*1.525*		
A. Sum Coefficient times value		10.06	12.29		11.42	14.42
B. (Root MSE)^2/2		0.39	0.39		1.17	1.17
C. Total Cost [$e^{(A+B)}$]		$34,465	$319,391		$294,340	$5,872,506
D. Unit Cost [C/spill volume]		$345	$32		$2,943	$1,175

3 Uncertainty and Limitations

The analysis presented in this report includes several limitations and uncertainties regarding the social costs of crude oil releases. This section discusses some of the key limitations of the models described in Section 2.2 and explores various sources of uncertainty may affect model coefficients and resulting relationships.

3.1 Issues Related to Missing Cost Data

One of the difficulties in estimating a regression model of incident costs is distinguishing between incidents which have cost categories equal to zero and incidents which have missing cost data. The regression models could underestimate total costs if missing cost categories are treated as zero.

Table 3-1 summarizes the cost categories that are present in the three primary datasets used. Total costs are a combination of operator costs (e.g., property damage, emergency response costs, environmental remediation costs), and costs accruing to the public (e.g., property damage, environmental remediation costs). The figure shows substantial variation in the completeness of reporting across datasets.

Table 3-1: Percentage of categories with non-zero values, by dataset

	Rail		Pipeline 2002-2009		Pipeline 2010 to 2022	
Variable	Present	% Non-Zero	Present	% Non-Zero	Present	% Non-Zero
Public property damages	*	3%	*	11%	*	10%
Public environmental remediation costs	~	2%	*	43%	~	1%
Other public costs	X	0%	*	4%	X	0%
Costs of commodity lost	*	31%	*	82%	*	86%
Operator property damages	*	3%	*	27%	*	82%
Operator emergency response costs	*	97%	X	0%	*	81%
Operator environmental remediation costs	*	30%	X	0%	*	64%
Other operator costs	*	0%	*	73%	*	10%
Additional operator costs	X	0%	X	0%	~	0%

Note: Values represent the percent of observations that have a non-zero value for each cost category, across all observations from that dataset. Excludes imputed values for fatalities, injuries, value of product lost, and value of travel delays.
* = data reported
X = data not reported in underlying dataset
~ = data not reported in underlying dataset, but some supplementary data is used for a subset of incidents.

Table 3-2 shows the percentages of per-gallon costs in each different cost category, based on incidents reporting non-zero damages in the category. As shown in the table, the majority of incidents have positive costs for commodity lost, operator emergency response costs, and operatory property damages.

Table 3-2: Unit cost by cost category

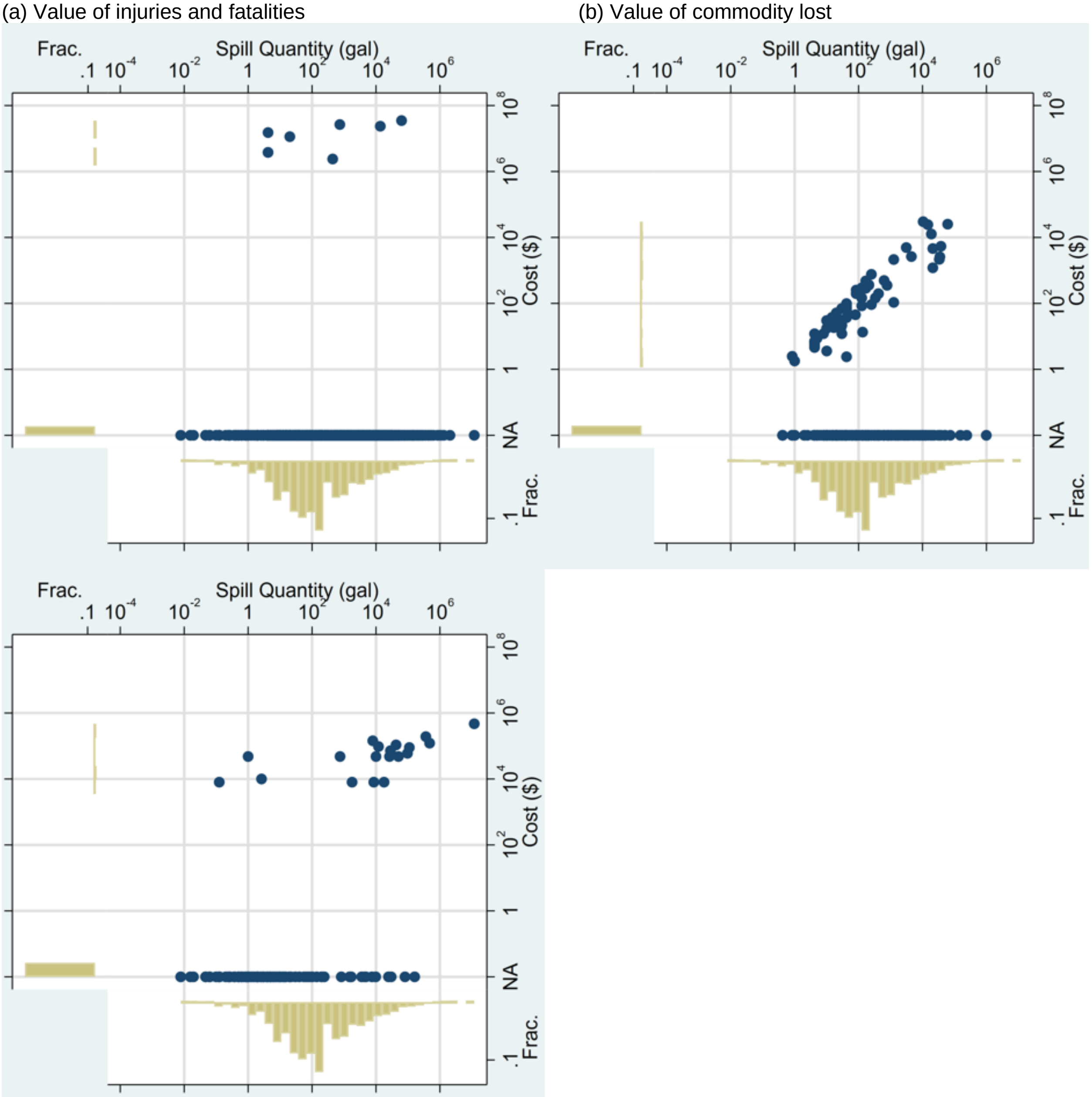
Cost Category	Percentage of Incidents, by Range of Cost per Gallon							
	$0 or missing	$0 to $1	$1 to $10	$10 to $100	$100 to $1,000	$1,000 to $10,000	$10,000 to $100,000	>$100,000
Public property damages	90.8%	2.1%	3.5%	2.2%	1.1%	0.3%	0.1%	0.0%
Public environmental remediation costs	90.0%	0.7%	2.2%	4.4%	2.2%	0.5%	0.1%	0.0%
Other public costs	99.1%	0.1%	0.3%	0.3%	0.1%	0.0%	0.0%	0.0%
Costs of commodity lost	20.4%	19.9%	58.0%	1.5%	0.1%	0.1%	0.0%	0.0%
Operator property damages	37.6%	7.3%	13.9%	20.5%	14.3%	4.6%	1.1%	0.6%
Operator emergency response costs	34.8%	2.8%	10.8%	22.4%	20.0%	7.1%	1.9%	0.2%
Operator environmental remed. costs	53.3%	1.0%	8.6%	19.6%	14.2%	2.8%	0.4%	0.0%
Other operator costs	77.2%	2.5%	4.3%	8.4%	6.1%	1.3%	0.2%	0.1%
Additional operator costs	99.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 3-1 presents scatterplots of imputed costs against quantity released, by category. The figure displays slight upward trends in costs for larger sized spills for the value of injuries and fatalities (panel a) and the value of travel delays (panel c) categories. In contrast, the figure displays a more pronounced upward trend in costs for larger sized spills for the value of commodity lost (panel b) category.

Figure 3-1: Costs vs. spill quantity, by cost category, for imputed costs

(a) Value of injuries and fatalities

(b) Value of commodity lost



(c) Value of travel delays

To mitigate concerns about under-reporting of cost categories in the database underlying the analysis, PHMSA ran several regression models which used each cost category as the dependent variable. Table 3-3 and Table 3-4 show these model specifications for rail and pipeline incidents, respectively. Summing across the estimated costs for each of these categories for an incident provides an upper-bound estimate of the total costs, assuming that all cost categories would be incurred.

Table 3-3: Regression-based analysis of costs categories for rail

Variable	Regression Specification				
	Public Property Damage	Commodity Loss Cost	Op. Property Damage	Op. Emergency Response Cost	Op. Env. Remediation
Intercept	7.635**	0.997***	12.35***	7.666***	6.540***
ln_qreleased	-0.178	1.122***	0.348*	0.0743	0.605***
ln_qreleased_sq	0.0485	-0.0168*	-0.0174	0.0320***	-0.00164
Observations	11	106	10	331	102
R^2	0.614	0.936	0.568	0.552	0.752
Root MSE	2.659	0.958	1.156	1.08	1.386

Note: Each column represents a different regression model, where the dependent variable is the log of cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

Table 3-4: Regression-based analysis of cost categories for pipeline

Variable	Regression Specification						
	Public Property Damage	Public Response/ Remediation	Commodity Loss Cost	Op. Property Damage	Op. Emergency Resp.	Op. Env. Remed.	Op. Other Costs
intercept	7.895***	9.124	1.441***	8.597***	7.591***	7.272***	8.151***
ln_qreleased	-0.174	-0.595	0.772***	-0.337***	0.0248	0.170*	-0.239
ln_qreleased _sq	0.0437***	0.0697	0.00179	0.0546***	0.0374***	0.0312***	0.0437**
hca_bi	0.971**	1.327	0.0400	0.491***	0.712***	0.363***	1.594***
fire_bi	0.749***	-3.275***	0.640*	2.861***	0.562	-0.0959	0.740
Observations	234	18	2,110	2,017	2,004	1,570	253
R^2	0.345	0.460	0.803	0.186	0.349	0.441	0.365
Root MSE	1.979	2.092	0.980	2.176	1.755	1.664	1.787

Note: Each column represents a different regression model, where the dependent variable is the log of cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

3.2 Issues Related to Spill Causality

The causal chain that connects incident characteristics to incident costs is not straightforward. For example, incidents involving larger spill volumes result in higher incident response costs and remediation costs. However, there exists the possibility of reverse causation, where rapid, resource-intensive responses can reduce the quantity of oil released and mitigate remediation costs.

This issue was explored using an instrumental variables (IV) regression. The IV regression uses response time as the instrumental variable.[7] Table 3-5 presents a comparison of the pipeline regression model results from Table 2-2 to the second-stage IV model results. The results from the main regression are similar to that of the second-stage IV, which suggests that reverse-causality between costs and quantity released has little impact on the estimation of the coefficients in the pipeline model.

Table 3-5: Regression-based analysis of total costs (instrumental variable analysis comparison)

Variable	Regression Specification	
	Primary Pipeline Regression	Second-Stage IV Regression
intercept	8.335	0.0444
ln_qreleased	-0.0454	NA
ln_qreleased_sq	0.0425	0.0444
pop800	0.000166	0.000135
watwet800_bi	0.405	0.459
spatial_missing	-1.353	0
et800_bi	0.500	0.467
lu_hddev_800	0.723	0.944
hca_bi	0.372	0.345
fire_bi	2.449	2.568
underground_bi	1.059	0.952
pcause_corrosion	0.191	0.230
pcause_natlforces	0.530	0.597
Observations	2,465	2,112
R^2	0.498	0.461

Note: The instrumental variable regression uses the first-stage to predict a single independent variable value, in this case *ln_qreleased_sq*. The models exclude incidents with response time exceeding two hours due to high variance and inconsistencies in reporting.

3.3 Alternative Pipeline Model with Full Incident Dataset

The main pipeline regression model described in Section 2.2 is based on the pipeline incidents that occurred between 2010 and 2022. This set of incidents has the most complete data available, following changes PHMSA made to reporting requirements. Table 3-6 presents the results of that regression using the full pipeline dataset, including incidents between 2005 and 2009.

This model has different slope terms for the two pipeline datasets. Separating these variables allows the costs across the datasets to have different relationships with quantity released. The

[7] An instrumental variable is a variable that represents an exogeneous source of variation in the independent variable. Response time directly influences the quantity released but does not directly affect remediation cost outside its effects on quantity released.

intercept of 8.336 is similar to the intercept of 8.335 in the main pipeline model. In contrast, the intercept is relatively lower for older pipeline incidents which would have a 1 value for the "*transp_mode_pipe09*" variable, shifting the intercept down by 1.415.

The rest of the variables in the regression match those from the main model, and the coefficients for the variables are generally similar in direction and magnitude.

Table 3-6: Regression-based analysis of total costs (full pipeline dataset)

Variable	Regression Specification
intercept	8.336*** (0.181)
transp_mode_pipe09	-1.415** (0.488)
ln_qreleased_pipe09	0.238 (0.137)
ln_qreleased_pipe22	-0.0528 (0.0619)
ln_qreleased_sq_pipe09	0.0197* (0.00951)
ln_qreleased_sq_pipe22	0.0430*** (0.00486)
pop800	0.000181*** (0.0000329)
watwet800_bi	0.417*** (0.0568)
spatial_missing	-0.604* (0.244)
et800_bi	0.576** (0.191)
lu_hddev_800	0.658* (0.323)
hca_bi	0.410*** (0.0721)
fire_bi	2.170*** (0.479)
underground_bi	1.067*** (0.0730)
pcause_corrosion	0.190** (0.0698)
pcause_natlforces	0.639*** (0.164)
Observations	3,241
R^2	0.502
Root MSE	1.574

Note: The dependent variable is the log of total cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

3.4 "Kitchen Sink" Models

Following the approach in the 2017 study, PHMSA also ran additional model specifications that includes additional explanatory variables, as an additional robustness check on the primary regression results; if the addition of more variables significantly impacts the direction and magnitude of the coefficients that are included in the primary regression model, it is an indicator that omitted variable bias may be driving the results.

Table 3-7 shows the results of the main regression models with additional explanatory variables selected based on variables available in the incident dataset and expert judgement on those factors that may influence the severity of the damages, the extent of contamination, and complexity of the spill response activities. These additional explanatory variables include an indicator for soil contamination (*soilcont_bi*), train speed (*train_speed*), and an indicator for incidents occurring during winter months (*winter*). The coefficients shared between the main models in Table 2-2 and Table 3-7 have the same direction and generally the same magnitude, indicating that omitted variable bias is not a significant concern in the regression analysis.

Table 3-7: Regression-based analysis of total costs (kitchen sink models)

Variable	Regression Specification	
	Rail	Pipeline
intercept	7.405*** (0.537)	8.152*** (0.186)
ln_qreleased	0.149*** (0.0326)	-0.106 (0.0623)
ln_qreleased_sq	0.0179** (0.00597)	0.0449*** (0.00489)
pop800	0.0000406 (0.000219)	0.000162*** (0.0000367)
watwet800_bi	1.417** (0.515)	0.410*** (0.0612)
spatial_missing	-0.876* (0.350)	-1.215*** (0.113)
et800_bi	-0.704 (1.557)	0.551* (0.258)
lu_hddev_800		0.779* (0.361)
hca_bi		0.392*** (0.0760)
fire_bi		2.540*** (0.468)
underground_bi		1.005*** (0.0825)
soilcont_bi		0.550*** (0.0728)
winter	0.0998 (0.128)	0.132 (0.0682)
train_speed	0.0312* (0.0143)	
rcause_derail	1.259* (0.499)	
rcause_component	-0.180 (0.105)	
rcause_aging	-0.287* (0.132)	
pcause_corrosion		0.185* (0.0819)
pcause_natlforces		0.554** (0.182)
Observations	340	2,465
R^2	0.792	0.510
Root MSE	0.867	1.507

Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

3.5 Incident Cause Models

As described in Section 1.3, reported incident causes were consolidated into four categories for rail incidents and eight categories for pipeline incidents. Some incident causes may result in systematically different incident costs. To examine the relationship between incident causes and costs, regression analyses were conducted predicting total costs based on the incident cause. These regressions allowed both the intercept and slope (i.e., the relationship between total cost and quantity released) to vary by cause. Table 3-8 summarizes the results of this analysis.

For pipeline incidents, the model excludes an indicator for “other” causes, meaning that when the other six indicator cause variables are set to zero, the model defaults to the “other” cause category. Similarly, the rail model excludes an indicator for incidents caused by error and when the other three indicator cause variables are set to zero, the model defaults to the “error” cause.

Focusing on the pipeline model, relative to the “other” cause category, other incident causes do not result in substantially different total costs (as indicated by the statistical insignificance of each). However, the growth in total costs with increases in quantity released is relatively faster for each of the other cause categories when compared to the “other” cause category.

Focusing on the rail model, relative to the error cause category, derailments (*rcause_derail*) result in substantially higher total costs, while the component (*rcause_component*) and aging (*rcause_aging*) causes result in relatively lower costs. Additionally, incidents caused by aging result in the lowest total cost, and the increase in total costs with quantity released for this cause (*ln_qr_rcause_aging*) category occurs relatively slowly. In contrast, total costs increase fastest with quantity released for derailments (*ln_qr_rcause_derail*).

Table 3-8: Regression-based analysis of total costs (cause models)

Variable	Regression Specification	
	Rail	Pipeline
intercept	8.333*** (0.164)	8.579*** (0.704)
rcause_derail	3.348*** (0.604)	
rcause_component	-0.481** (0.173)	
rcause_aging	-0.591** (0.189)	
pcause_corrosion		-0.142 (0.723)
pcause_equipment		-0.843 (0.717)
pcause_excavation		-1.337 (0.908)
pcause_incorrectopp		-1.116 (0.770)
pcause_mwfail		1.359 (0.798)
pcause_natlforces		0.606 (0.909)
pcause_outforce		-0.804 (1.200)
ln_qr_rcause_derail	0.235*** (0.0552)	
ln_qr_rcause_component	0.223*** (0.0366)	
ln_qr_rcause_aging	0.0653 (0.0731)	
ln_qr_rcause_error	0.246*** (0.0644)	
ln_qr_pcause_corrosion		0.469*** (0.0266)
ln_qr_pcause_equipment		0.418*** (0.0258)
ln_qr_pcause_excavation		0.573*** (0.0650)
ln_qr_pcause_incorrectopp		0.491*** (0.0550)
ln_qr_pcause_mwfail		0.384*** (0.0621)
ln_qr_pcause_natlforces		0.361*** (0.102)
ln_qr_pcause_outforce		0.595*** (0.125)
ln_qr_pcause_other		0.533*** (0.115)
Observations	340	2,465
R^2	0.616	0.416
Root MSE	1.172	1.645

Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

3.6 Combined Quadratic Log-Log Model

Table 3-9 shows the results of the regression analysis based on a pooled dataset of rail and pipeline releases (including pipeline incidents since 2010). When compared to the separate rail and pipeline regression models in Table 2-2, the coefficients shown in Table 3-9 have the same direction and generally the same magnitude.

Table 3-9: Regression-based analysis of total costs (combined)

Variable	Regression Specification
intercept_pipe	8.368*** (0.180)
intercept_rail	8.022*** (0.295)
ln_qreleased_pipe	-0.0553 (0.0619)
ln_qreleased_rail	0.170*** (0.0364)
ln_qreleased_sq_pipe	0.0432*** (0.00487)
ln_qreleased_sq_rail	0.0125 (0.00689)
pop800	0.000170*** (0.0000374)
watwet800_bi	0.409*** (0.0621)
spatial_missing	-0.423 (0.285)

Table 3-9: Regression-based analysis of total costs (combined)

Variable	Regression Specification
et800_bi	0.413 (0.252)
lu_hddev_800	0.616 (0.362)
hca_bi	0.381*** (0.0764)
fire_bi underground_bi	2.165*** (0.420) 1.059*** (0.0832)
rcause_derail	1.627** (0.546)
rcause_component	-0.208* (0.106)
rcause_aging	-0.320* (0.133)
pcause_corrosion	0.190* (0.0826)
pcause_natlforces	0.542** (0.183)
Observations	2,805
R^2	0.563
Root MSE	1.465

Note: The dependent variable is the log of total cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

3.7 Unit Cost Model

An alternative modeling approach would be to use the log of unit cost as the dependent variable. However, this approach is mathematically equivalent to the main regression specification and, with the exception of quantity released, would produce the same coefficients on all variables (see Abt Associates (2017) for additional details). Table 3-10 illustrates this by presenting results for the combined quadratic log-log model (Table 3-9), where the dependent variable is the log of unit costs.

Table 3-10: Regression-based analysis of unit costs (combined)

Variable	Regression Specification
Intercept_pipe	8.368*** (0.180)
Intercept_rail	8.022*** (0.295)
ln_qreleased_pipe	-1.055*** (0.0619)
ln_qreleased_rail	-0.830*** (0.0364)
ln_qreleased_sq_pipe	0.0432*** (0.00487)
ln_qreleased_sq_rail	0.0125 (0.00689)
pop800	0.000170*** (0.0000374)
watwet800_bi	0.409*** (0.0621)
spatial_missing	-0.423 (0.285)
et800_bi	0.413 (0.252)
lu_hddev_800	0.616 (0.362)
hca_bi	0.381*** (0.0764)
fire_bi	2.165*** (0.420)
underground_bi	1.059*** (0.0832)
rcause_derail	1.627** (0.546)
rcause_component	-0.208* (0.106)
rcause_aging	-0.320* (0.133)
pcause_corrosion	0.190* (0.0826)
pcause_natlforces	0.542** (0.183)
Observations	2,805
R^2	0.578
Root MSE	1.465

Note: The dependent variable is the log of unit cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

3.8 Alternative Functional Forms

The quadratic log-log functional form (which includes the log of quantity released and the log of quantity released squared) is the best fit for the rail and pipeline incidents costs in the database. However, alternative functional forms were also explored, including a linear form and a “binned” form. These alternative functional form models are specified similarly to those models presented in Table 2-2, except that the quantity variables take on different forms.

Table 3-11 presents the results of rail and pipeline models using a linear form (i.e., excludes *ln_qreleased_sq*).

Table 3-11: Regression-based analysis of total costs (linear log-log)

Variable	Regression Specification	
	Rail	Pipeline
intercept	7.458*** (0.576)	7.056*** (0.0981)
ln_qreleased	0.285*** (0.0349)	0.469*** (0.0156)
pop800	-0.0000485 (0.000250)	0.000159*** (0.0000392)
watwet800_bi	1.734** (0.545)	0.432*** (0.0633)
spatial_missing	-1.178** (0.397)	-0.822*** (0.0777)
et800_bi		0.613* (0.300)
lu_hddev_800		0.671 (0.367)
hca_bi		0.380*** (0.0787)
fire_bi		2.911*** (0.471)
underground_bi		1.175*** (0.0874)
rcause_derail	2.326*** (0.487)	
rcause_component	-0.217 (0.112)	
rcause_aging	-0.325* (0.147)	
pcause_corrosion		0.0346 (0.0853)
pcause_natlforces		0.542** (0.197)
Observations	340	2,465
R^2	0.761	0.474
Root MSE	0.924	1.560

Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

Next, Table 3-12 presents regressions using a “binned” specification, in which total costs are modeled independently for a number of ranges of gallons spilled for each transportation mode. This model allows flexibility in the relationship between incident costs and different ranges of gallons spilled. However, this model cannot be used to examine how total costs may change *within* each range of gallons spilled. This limits the model’s explanatory power. Both the rail and pipeline models reveal an increase in total incident costs for larger gallons spilled (i.e., when comparing coefficient magnitudes between bins representing larger quantities released).[8]

Table 3-12: Regression-based analysis of total costs (binned functional form)

Variable	Regression Specification	
	Rail	Pipeline
intercept	12.65*** (0.789)	9.863*** (0.493)
pop800	0.0000640 (0.000259)	0.000159*** (0.0000382)
watwet800_bi	1.396* (0.552)	0.385*** (0.0633)
spatial_missing	-0.868* (0.336)	-1.381*** (0.152)
et800_bi		0.471 (0.255)
lu_hddev_800		0.679 (0.372)
hca_bi		0.366*** (0.0777)
fire_bi		2.286*** (0.488)
underground_bi		1.113*** (0.0841)
rcause_derail	1.708** (0.525)	
rcause_component	-0.148 (0.104)	
rcause_aging	-0.266* (0.127)	
pcause_corrosion		0.179* (0.0832)
pcause_natlforces		0.528** (0.184)
Ln_qreleased_binA (0.001 to 0.009 gal)	-4.448*** (0.555)	
Ln_qreleased_binB (0.01 to 0.09 gal)	-5.719*** (0.594)	
Ln_qreleased_binC (0.1 to 0.9 gal)	-5.344*** (0.562)	
Ln_qreleased_binD (1 to 9 gal)	-5.142*** (0.546)	-1.442** (0.523)
Ln_qreleased_binE (10 to 99 gal)	-4.731*** (0.548)	-1.252* (0.494)
Ln_qreleased_binF (100 to 999 gal)	-3.501*** (0.567)	-0.194 (0.495)
Ln_qreleased_binG (1,000 to 9,999 gal)	-2.391*** (0.522)	0.689 (0.495)
Ln_qreleased_binH (10,000 to 99,999 gal)	-1.268** (0.418)	2.287*** (0.505)
Ln_qreleased_binI (100,000 to 999,999 gal)	0.709 (0.729)	4.423*** (0.574)
Observations	340	2,465
R^2	0.791	0.483
Root MSE	0.874	1.548

Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient with the corresponding standard errors in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with zero costs are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p< 0.001.

3.9 General Uncertainties in the Analysis

The analysis has several sources of uncertainty that may impact predictions of rail and pipeline incident costs. Table 3-13 presents potential sources of uncertainty and their directional effects on costs.

8 Note that the intercepts are included in both the rail and pipeline models. The large and positively signed intercepts must be considered in conjunction with the coefficients on each bin to determine the overall impact on total costs. For example, when focusing on the pipeline model, the effect on log total costs for spills between 1 and 9 gallons is (9.863 + -1.442), *ceteris paribus*.

Table 3-13: Sources of uncertainty in the analysis of incident costs

Source of Uncertainty	Description of Uncertainty	Directional Effect on Cost
Varying levels of detail and quality in data characterizing crude oil spill incidents	The database consists of data assembled from different datasets, with varying quality and levels of detail. Much of the incident data were self-reported, leading to potential differences in data availability and quality across incidents. For example, operators may not account for all cost categories in their reporting, which may lead to inconsistent cost estimates.	Uncertain
Missing information in the rail dataset	The rail dataset includes less data on potentially important variables, such as location (e.g., latitude and longitude) and surface water impact indicators. For some incidents, PHMSA supplemented the rail data using external sources (e.g., POLREP, Federal Rail Administration data). However, our ability to fill in the missing information was limited by the availability of matching records and data in these other datasets.	Underestimate
Sourcing damages to the release of crude oil	Not all damages may be directly attributable to the release of crude oil from rail and pipeline incidents. Some damages may have been caused by other precipitating events (e.g., fire, natural catastrophe). Including all costs may overstate the costs associated directly with a crude oil release.	Overestimate
Under-representation of costs associated with the release of crude oil	Operator reports to PHMSA, which constituted the primary source of data used in the analysis, are not a comprehensive accounting of the costs associated with an incident. For example, costs omit potential reductions in nearby property values which have been found to be substantial (Cheng *et al.*, 2021). As a result, costs may be understated.	Underestimate

4 Conclusions and Recommendations

This study updates the original Abt Associates (2017) analysis of the social costs of onshore releases of crude oil from pipelines and railcars with six additional years of incident data (rail and pipeline incidents that occurred between 2017 and 2022). The updated regression models presented in Section 2.2 (whose results closely align with the original models from Abt Associates, 2017) allow PHMSA to estimate the expected costs from crude oil releases while accounting for incident characteristics such as the transportation mode, quantity released, affected environments, incident cause, and others. Incident costs can vary significantly depending on incident characteristics so the ability to account for these factors should improve the accuracy of cost predictions compared to simpler cost analyses (e.g., cost predictions estimated from historical averages). Additionally, the new database provides a rich set of information about historical incidents that could support future analyses.

Although the updated models produce regression results that are largely consistent with the results presented in Abt Associates (2017), cost predictions can differ substantially. In general, larger and more comprehensive datasets produce more accurate models, so continuing to update the dataset with additional incidents over time could further improve the models.

Although the models presented in this report predict the cost of crude oil spills, similar models could be developed for spills involving other hazardous materials such as natural gas. Additionally, it may be worth investigating other sources of costs not captured in the analysis such as the cost of oil spills on nearby property values.[9] Lastly, although the models can be used to predict the cost of large outlier oil spill incidents, they are not well suited for this purpose. In the updated database, the incident with the largest cost is $1.42 billion, several orders of magnitude above the average of $1.2 million. It may be worth exploring more appropriate, alternative statistical approaches, such as worst-case scenario extrapolation methods (Chen *et al.*, 2015; Malevergne *et al.*, 2006), to predict the costs of such events.

[9] Cheng *et al.* (2021) found average property value reductions of 7.4 percent for homes within 1-kilometer of a spill and 27 percent of incidents in the updated database occur within an HCA.

5 References

Abt Associates. (2017). *Valuation of Crude Oil Spilled in Transportation Incidents*. Retrieved from

Chen, J., Lei, X., Zhang, L., & Peng, B. (2015). Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China. *PLoS ONE, 10*(2), e0118521.

Cheng, N., Li, M., Liu, P., Luo, Q., Tang, C., & Zhang, W. (2021). *Pipeline and Property Values: A Nationwide Hedonic Analysis of Pipeline Accidents Over the Past Three Decades.* Paper presented at the 2021 Annual Meeting, August 1-3, Austin, Texas.

Enbridge Inc. (2017). *Annual Report: 2016*. Retrieved from https://www.enbridge.com/investment-center/~/media/Enb/Documents/Investor%20Relations/2017/2016_ENB_EI_AnnualReport_Web.pdf

Enbridge Inc. (2018). *Annual Report: 2017*. Retrieved from https://www.enbridge.com/~/media/Enb/Documents/Investor%20Relations/2018/2017_ENB_AnnualReport_Full_optimized.pdf

Malevergne, Y., Pisarenko, V., & Sornette, D. (2006). On the power of generalized extreme value (GEV) and generalized Pareto distribution (GPD) estimators for empirical distributions of stock returns. *Applied Financial Economics*, *16*(3), 271-289.

National Transportation Safety Board. (2017a). *National Transportation Safety Board Railroad Accident Brief: BNSF Railway Crude Oil Unit Train Derailment, Heimdal, North Dakota*. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/RAB1712.pdf

National Transportation Safety Board. (2017b). *NSF Railway Train Derailment and Subsequent Train Collision, Release of Hazardous Materials, and Fire, Casselton, North Dakota*. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/RAB1701.pdf

National Transportation Safety Board. (2021). *December 22, 2020, Derailment of BNSF Railway Freight Train U-TNDFEP-0-36T and Subsequent Hazardous Materials Release*. Retrieved from https://data.ntsb.gov/Docket/?NTSBNumber=HMD21LR001

Plains All American Pipeline. (2020). Form 10-K (Annual Report) Filed 2021-03-01 for the fiscal year ended December 31, 2020.

U.S Environmental Protection Agency. (2016a). *Ash Coulee Spill. December 2016*. Retrieved from https://response.epa.gov/site/site_profile.aspx?site_id=11952

U.S Environmental Protection Agency. (2016b). *Pollution/Situation Report. ENTERPRISE PIPELINE SPILL- E17602 -- Polrep #3*. Retrieved from https://response.epa.gov/site/sitrep_profile.aspx?site_id=11890&counter=27892

U.S. Department of Transportation. (2022). *Departmental Guidance on Valuation of a Statistical Life in Economic Analysis*. Retrieved from https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-on-valuation-of-a-statistical-life-in-economic-analysis

United States Energy Information Administration (U.S. EIA). (2023). *Total Crude Oil by Rail (Thousand Barrels)*. Retrieved from https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=ESM_EPC0_RAIL_ZAMN-ZAMN_MBBL&f=M

Appendix A: 2017 Study on the Valuation of Crude Oil Spilled in Transportation Incidents

This appendix provides the report from the 2017 study, including more details on the approach for modeling the relationship between the expected costs of crude oil releases and incident characteristics.

Valuation of Crude Oil Spilled in Transportation Incidents

Contract GS-10F-0086K
Task Order #DTPH5614F00037
Technical Task Directive #007

March 31, 2017

Prepared for:
Pipeline and Hazardous Materials Safety Administration
PHH-60
1200 New Jersey Avenue, SE
Washington, DC 20590

Submitted by:
Abt Associates
4550 Montgomery Avenue
Suite 800 North
Bethesda, MD 20814

Table of Contents

Executive Summary

Study Objectives and Results

The Pipeline and Hazardous Materials Safety Administration (PHMSA) conducts regulatory impact analyses and economic analyses of regulations, standards and other policies that address safety issues and the risk of a crude oil spill during transportation. The purpose of this study is to improve PHMSA's ability to estimate the benefits of preventing onshore releases of crude oil from pipelines and railcars, through the development of rigorous, defensible estimates of the social costs of such releases. The study results include mode-specific models to estimate expected costs from crude oil releases, accounting for incident characteristics.

Incident Data

We compiled data on onshore releases of crude oil from railcars and pipelines from a variety of data sources. The main sources of data are incident reports submitted to PHMSA by pipeline or rail operators. We supplemented these data with information from U.S. Coast Guard's National Response Center and National Pollution Fund Center, National Transportation Safety Board investigation reports, natural resource damage assessment settlements, company financial reports, spatial analysis of incident locations, and other sources. The incident database compiled for this study contains information for a total of 2,487 onshore incidents that involved the release of crude oil from pipelines (2,035 incidents) or railcars (452 incidents) between January 2005 and June 2016. 2,341 of these incidents (2,026 pipeline and 315 rail incidents) included some reported costs. For each incident, the database provides information on the quantity released, source characteristics, risk factors, incident cause, response actions, affected environment, and costs. Exhibit ES-1 shows the distribution of total costs for all incidents in the database, whereas Exhibit ES-2 shows the relationship between the quantity released and total costs.

Exhibit ES-1: Distribution of incidents by total costs (2015$)

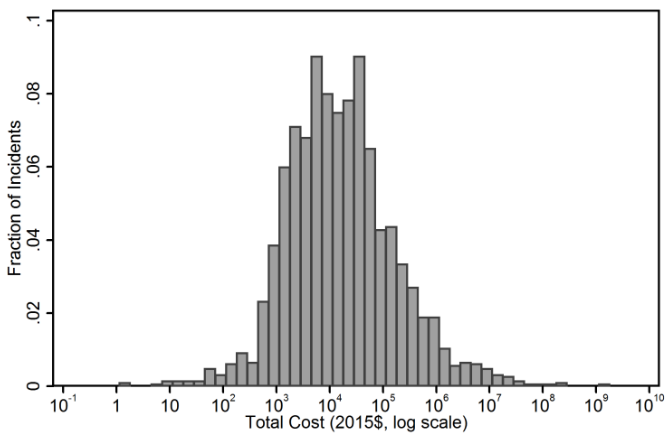
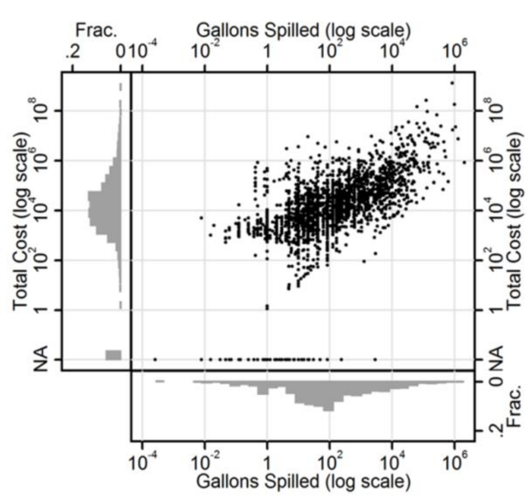


Exhibit ES-2: Relationship between crude oil quantity released and total costs



Crude Oil Spill Cost Models

The database provides an empirical basis to model the relationship between incident-specific factors and costs of a crude oil spill. The model takes on the general form $logCost_i = f(SpillCharacteristics_i, SiteCharacteristics_i, CostMethod_i)$ for a given incident i.

Exhibit ES-3 shows the variable and coefficients for the two main regression models of expected spill costs for rail or pipeline incidents developed in the study. See Section 5 of this report for details. The models allow estimates of the expected costs of crude oil spills for different types of incidents and quantities released.

Variables included in the two models differ according to the transportation mode and reflect differences in the relevant factors contributing to an incident, and data availability. As expected, the quantity released is a key determinant of the expected cost of an incident, but characteristics of the affected environment, incident cause, and other factors also influence the costs. For example, all else being equal, the expected costs for a pipeline spill occurring in a high consequence area (HCA) are 1.6 times larger than those for an otherwise similar spill outside of an HCA; expected costs for a spill associated with a train derailment are 9 to 14 times larger than costs for other causes. The study provides additional models that address uncertainty and sensitivity of expected costs to the assumed functional form, included explanatory variables, or data gaps. See details in Appendix D.

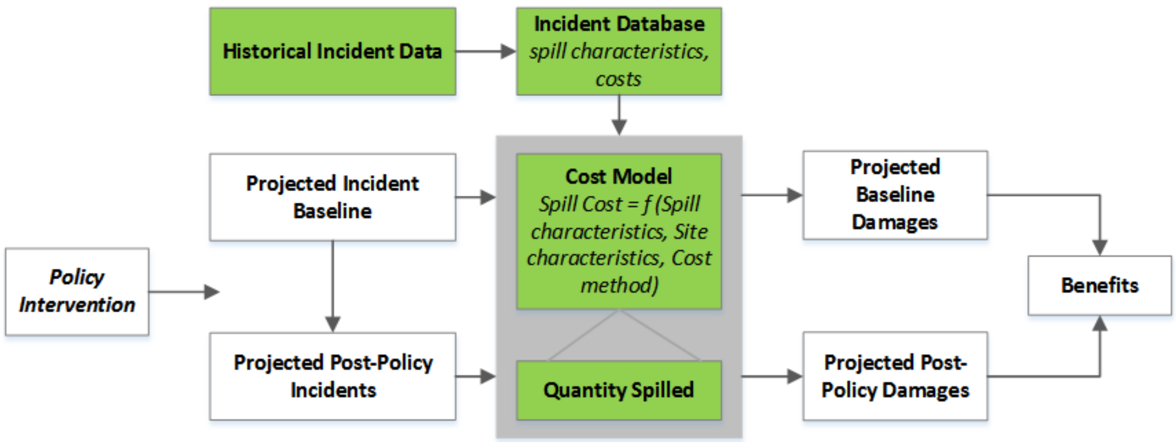
Exhibit ES-3: Specification of regressions used to estimate *ln(costs)* for crude oil spill incidents from rail and pipeline.

Variable Category	Variable	Regression Model	
		Rail	Pipeline
	Intercept	7.687***	8.113***
Quantity released	Ln(quantity released)	0.141***	-0.0769
	Ln(quantity released)^2	0.0171***	0.0464***
Affected environment	Population within 800 meters	0.000195	0.000166***
	Water within 800 meters (binary)	1.009	0.298**
	Spatial data missing (binary)	-0.859	-1.195***
	E&T habitat within 800 meters (binary)		0.708
	% high intensity development within 800 meters		0.839
	High Consequence Area (binary)		0.478***
Incident cause	Corrosion (cause binary)		0.0337
	Natural forces (cause binary)		0.653*
	Derailment cause (binary)	2.147***	
	Loose, missing, broken component cause (cause binary)	-0.277*	
	Aging (cause binary)	-0.469**	
Spill impacts	Fire (binary)		2.857***
	Underground (binary)		1.143***
	Root MSE	*0.857*	*1.604*
	R^2	*0.761*	*0.497*

Model Application in Policy Analyses

One possible application of the cost models is to estimate the change in expected spill costs given a policy-induced change in the number, magnitude, or character of future incidents, relative to a pre-policy baseline. Exhibit ES-4 shows main steps in this type of analysis. The study describes two example analyses that apply the cost models to estimate the benefits of hypothetical policy interventions designed to reduce the risk transporting crude oil by pipeline or rail. The first example estimates the benefits a hypothetical reduction in the delay between a pipeline rupture and subsequent detection and response actions; the second example estimates the benefits of a hypothetical strengthening of railcars to reduce the probability of the car releasing its content during a derailment. See Section 6 of this report for details of these two illustrative analyses.

Exhibit ES-4: Application of the cost models to analyses of the benefits of policy interventions. The green boxes on the diagram highlight the elements of the analysis addressed in this study.



Uncertainty and Limitations

The cost modeling methodology relies on empirical data and therefore inherits the uncertainty and limitations of the data in terms of the characteristics of the reported incidents or their costs. Because the reported costs do not capture all third-party costs or spill impacts, the modeled expected cost may understate actual costs especially for incidents with far reaching consequences. Conversely, to the degree that the incidents involve multiple events occurring concurrently with the crude oil release, the modeled expected costs may include damages that are not a direct result of the spill. Furthermore, factors not captured in the incident data and excluded from the cost models, such as response actions, could affect costs in either direction (e.g., increasing direct response costs while reducing overall damages and impacts). See Section 7 for a detailed discussion of the uncertainty and limitations.

Conclusions and Recommendations

The study provides a comprehensive database and cost models for use in estimating the benefits of safety regulations or other policy interventions that reduce the risk of a crude oil spill. The incident database provides a rich set of information to support a variety of future analyses. Since costs of crude oil spills can vary significantly across incidents, the cost models' ability to account for the differences in incident characteristics will improve the validity of the estimates as compared to prior analyses that relied on uniform unit costs.

More generally, the study provides a methodology for relating the cost of incidents to their characteristics. PHMSA can update the crude oil cost models as new data become available, or use the same approach for estimating models for other hazardous liquids. PHMSA may also want to consider developing a companion methodology for estimating incident probabilities and characteristics that could serve as input for valuing benefits.

1 Introduction

The Pipeline and Hazardous Materials Safety Administration (PHMSA) conducts regulatory impact analyses and economic analyses of:

- Proposed and final regulations and standards,
- Revisions to regulations and standards, and
- Regulatory and non-regulatory alternatives that address safety issues.

These analyses must be consistent with requirements and guidelines contained in various public laws, executive orders, court decisions, and other documents.[1] In addition, PHMSA engages in surveys and statistical analyses, evaluations, reviews of hazard data, and the development of analytical tools. These efforts help PHMSA to identify risks and their causes, prioritize regulatory and non-regulatory activities, and produce effective approaches and solutions for improving pipeline and rail safety.

The objective of this study is to inform these analyses by providing PHMSA with an evidence base and methods for estimating avoided damages from hazardous material transportation safety regulations related to crude oil. Specifically, the study develops monetized damage estimates for crude oil spills based on past incidents, for use in calculating the benefits of avoiding future releases. The study also assembles non-quantitative and non-monetized information to characterize the types of damages resulting from crude oil releases. Finally, the study illustrates how this information can be used to support policy analysis, by applying the cost models to estimate the benefits of two hypothetical interventions designed to enhance the safety of transportation of crude oil by pipeline and rail.

1.1 Purpose and Scope

Analyses of federal regulations to protect and improve health, safety, and the environment are challenging, particularly when they involve estimating the benefits of avoiding relatively infrequent and far-reaching adverse outcomes. In the case of the benefits of preventing oil spills, analytical challenges include understanding the factors that contribute to the incidents in the first place and affect the probability and size of a spill. The severity of damages depends on the human and environmental receptors affected and on the effectiveness of containment and cleanup actions following the spill. It is not always easy, or even possible, to quantify the impacts of a spill, especially in monetary terms. With these challenges in mind, data from past incidents provide an empirical basis for quantifying damages from past spills and for estimating the potential benefits of preventing similar releases in the future.

In analyzing the benefits of spill prevention regulations, PHMSA has typically relied on relatively simple measures of damages, such as costs incurred by operators per gallon of oil spilled. Not all spills are the same, however, and a regulation may address only a subset of spill circumstances. Furthermore, spill risks

[1] These requirements include Executive Order 12866: Regulatory Planning and Review, the Regulatory Flexibility Act, the Unfunded Mandates Reform Act, the Paperwork Reduction Act, and Environmental Impact Statements.

may change over time as more stringent safety measures are implemented. The ability to disentangle the contribution of multiple factors to the probability of a spill and its resulting damages is an important step in improving PHMSA's ability to estimate benefits of its regulations. This study contributes to this effort by focusing on the damages aspect.

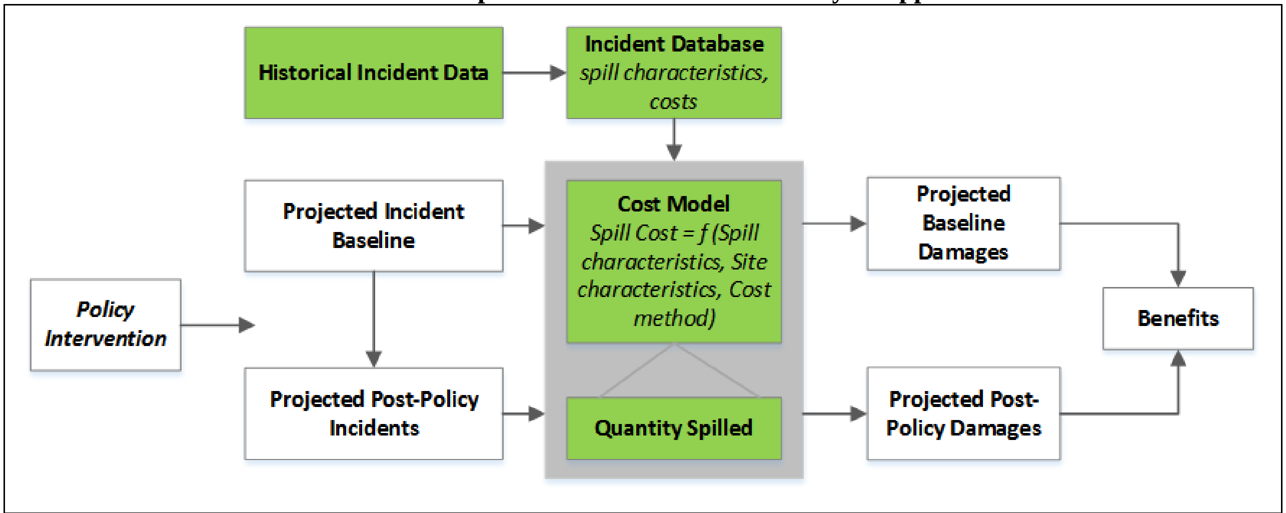
1.2 General Analysis Approach

The purpose of this study is to improve PHMSA's ability to estimate the benefits of preventing onshore releases of crude oil from pipelines and railcars, through the development of rigorous, defensible estimates of the social costs of such releases. The analysis estimates expected costs from crude oil releases, accounting for incident characteristics that affect the magnitude of damages.

To develop these estimates, we first compiled a database of onshore releases of crude oil from railcars and pipelines from a variety of federal datasets, incident reports, natural resource damage assessments, and other sources. We reduced, validated, and cleaned the data to assemble an analytic database of incident characteristics, context, and costs in a consistent format.

We then used this database to develop models of the factors driving the costs associated with oil released during transportation incidents. These models allow estimates of the expected costs of crude oil spills for different types of incidents and quantities released. The green boxes in the flow diagram of Exhibit 1-1 represent the steps in the analysis that are the main focus of this report.

Exhibit 1-1: Simplified Outline of Benefits Analysis Approach



Source: Abt Associates

1.3 Organization of Report

This report is organized as follows:

- **Section 2: Background and Conceptual Framework** – provides information about the conceptual analytic framework as well as an overview of a literature review of similar or related studies and of the impacts of crude oil spills.

- **Section 3: Incident Data Sources and Compilation Methodology** – documents the datasets from which we assembled the incident database, and our methods for extracting, cleaning, and consolidating the data.
- **Section 4: Database and Crude Oil Incidents Description** – provides an overview of the database contents and summary statistics for key variables.
- **Section 5: Analysis of Crude Oil Incident Costs** – documents the relationship between incident costs and key explanatory variables, and describes the development of a regression predicting costs for different types of incidents.
- **Section 6: Model Application** – describes the use of the model in policy analysis and provides illustrative examples for hypothetical policy interventions.
- **Section 7: Sensitivity Analyses and Uncertainties** – discusses uncertainties in the analysis and presents sensitivity analyses exploring the impact of key uncertainties.
- **Section 8: Conclusions and Recommendations –** summarizes key insight from the study and suggested next steps.
- **Section 9: References** – provides the references used in the analysis.
- **Appendices** provide additional supplementary information and data.

2 Background and Conceptual Framework

Expansion in domestic oil production has led to increasing volumes of crude oil being transported to refineries within the United States. Traditionally, pipelines and oceangoing tankers have delivered the vast majority of crude oil to U.S. refineries, accounting for approximately 93 percent of total receipts in 2012 (U.S. Energy Information Administration (U.S. EIA), 2015). In recent years, however, other modes of transportation have accounted for a growing share of the volume. Rail transportation, in particular, has increased significantly, with the volume of crude oil carried by rail growing by over 2,000 percent between 2011 and 2014, from 6.5 million to nearly 160 million barrels (Frittelli et al., 2014; U.S. EIA, 2015).

The growing volume of crude oil transported by rail raises the prospect of increased risk of train accidents that could release crude oil to the environment. These releases may cause significant damages to the public, property, and natural resources. However, information about the costs associated with the releases of oil is limited and highly variable.

Section 2.1 describes literature and past efforts to quantify and monetize the damages associated with crude oil releases generally. Section 2.2 describes the conceptual framework underlying our database and analysis. Sections 2.3 and 2.4 review the literature on two selected topics pertinent to the purpose of this study: the types of costs resulting from crude oil releases and incident causes, respectively.

2.1 Past Efforts to Value Crude Oil Releases

Several studies have sought to place a value on damages associated with crude oil releases during transportation, including:

- Chang et al. (2014) developed a general framework to evaluate a range of biophysical and social characteristics that influence the potential consequences of oil spills. Although the study’s focus was primarily on spills in marine environments, the central premises have some applicability to onshore releases as well. The authors focused on factors including the occurrence of a spill event, characteristics of the particular oil spill, short-term and long-term consequences to ecosystems, and impacts to society in terms of economic, human health, and social consequences. With respect to the oil spill event itself, the authors note that location of a spill is one of the most important predictors of its impacts. Other determinants include the amount of oil spilled. Characteristics of the physical environment determine impacts on ecosystems and society.
- Etkin (2004) developed the Basic Oil Spill Cost Estimation Model (BOSCEM) for the U.S. Environmental Protection Agency (U.S. EPA) to estimate the costs and damages for different types of oil spills. The model uses information about the oil type, quantity, and primary response methodology to assign a base cost to a spill, then applies cost modifiers based on year spilled and location-specific variables such as socio-economic/cultural value, freshwater use, and habitat and wildlife sensitivity. This yields a per-gallon unit cost for a particular release. For example, the model provides base crude oil spill response and cleanup (using mechanical oil recovery) of $199 per gallon, socioeconomic costs of $50 per gallon, and base environmental costs of $90 per gallon for spills of less than 500 gallons. Modifiers to the base costs include response cost modifiers ranging between 0.5 (for pavement) and 1.6 (for wetlands) with a default of 1.0 for open water or shoreline; socioeconomic

and cultural value modifiers ranging between 0.1 for industrial areas to 2.0 for sites that support subsistence or commercial fishing (default is 0.7 for residential areas), vulnerability cost modifiers ranging between 0.4 for industrial areas to 1.7 for areas used by wildlife (default of 0.9), and habitat sensitivity modifiers ranging between 0.4 (for urban areas) to 4.0 (for wetlands), with a default of 1.5 for rivers and streams.

- Helton and Penn (1999) evaluated natural resource damage assessments (NRDAs) in the context of the total costs associated with oil release incidents. For 48 incidents with available data, the authors compiled data on responsible party response costs, federal and state response costs, assessment costs, natural resource damages, third party claims, penalties, and other costs (including salvage and repair costs, delay and additional operating costs, and lost or damaged cargo costs). Since the focus of the study was to put NRDA into context, the authors focused on incidents with NRDA data available, potentially skewing NRDA costs upward compared with other categories where the authors had incomplete information. They found that NRDA costs are typically about 25 percent of the known incident costs, while response costs are over 1.5 times higher.

In estimating the benefits associated with avoided releases of crude oil from railcars or other modes of transportation, the U.S. Department of Transportation (U.S. DOT) has relied on operator-reported costs from past incidents. For example, in the 2015 Regulatory Impact Analysis (RIA) of the final HM-251 rule (Hazardous Materials: Enhanced Tank Car Standards and Operational Controls for High-Hazard Flammable Trains (HHFTs)), PHMSA (2015) drew on the studies above and other sources, including data from reported incidents, to gather costs associated with crude oil (and ethanol) releases. The agency calculated an overall average unit cost of approximately $200 per gallon spilled, based on historical data from pipeline releases, and used this value to estimate the benefits associated with avoided incidents resulting from HM-251 by multiplying the average per gallon unit costs by the average volume released by a mainline derailment (83,602 gallons). See Appendix A for details on the unit cost calculations in PHMSA (2015). In the 2015 RIA, PHMSA also summarized the unit costs reported in the literature. For crude oil spills, these unit costs ranged from less than $10 per gallon to nearly $900 per gallon (in 2013 dollars).[2]

PHMSA also applied a per-gallon benefit estimate in a subsequent analysis of the benefits associated with a rule expanding the applicability of requirements for oil spill response plans (OSRPs) for HHFTs,[3] which the agency expects will reduce the magnitude and severity of spills (PHMSA, 2016a). Although the rule applies to rail transportation only, PHMSA (2016a) notes that sufficient data to estimate the costs of releases from rail cars are not available, and that “pipeline incidents are the closest analog available.” As such, the agency applied a per-gallon cost of $211 to oil released due to rail incidents, based on the crude oil analysis presented in PHMSA (2015). Since available data suggests that rail incidents may have a higher rate of water contamination compared with pipeline releases, PHMSA (2016a) also evaluated the impact of water contamination on per-gallon costs. The agency found that spills involving water contamination (n=95) had significant higher unit costs than those without water contamination (n=878), at $467 per gallon compared with $49 per gallon.

---

2 See Table C1 in PHMSA (2015).

3 HHFT is defined as a train transporting 20 or more loaded tank cars of liquid petroleum oil in a continuous block or a single train carrying 35 or more loaded tank cars of liquid petroleum oil throughout the train consist.

As noted in PHMSA (2015; 2016a) and elsewhere, site- and incident-specific factors are important determinants of the costs associated with a particular spill. Additionally, despite the large costs associated with certain incidents, there is relatively little literature exploring the determinants of these costs or models to predict the extent and costs of damages resulting from the releases.[4] This information is critical for evaluating the benefits of preventing or reducing the magnitude of future incidents through more stringent safety regulations or preventive measures.

2.2 Conceptual Framework

The diagram shown in Exhibit 2-1 illustrates the interplay of factors that contribute to the magnitude and costs of a crude oil release. These factors include:

- **Source characteristics**, such as the transportation mode (pipe/rail), physical characteristics of the container such as the age, diameter, construction material, operating pressure, and characteristics of the transported product (e.g., viscosity);
- **Risk factors**, which are external factors such as environmental conditions (e.g., temperature and precipitation), operating procedures, and the speed at which the rail car is traveling, which may affect the probability and/or the consequences of an incident;
- **Incident causes**, which may lead to different failure modes and impacts;
- **Characteristics of the affected environment**, such as the affected media (e.g., water, soil), population density, proximity to sensitive ecological receptors; and
- **Effectiveness of the response**, which relates to actions that can limit the actual quantity released during the incident, contain the spread of the oil, and reduce resulting damages.

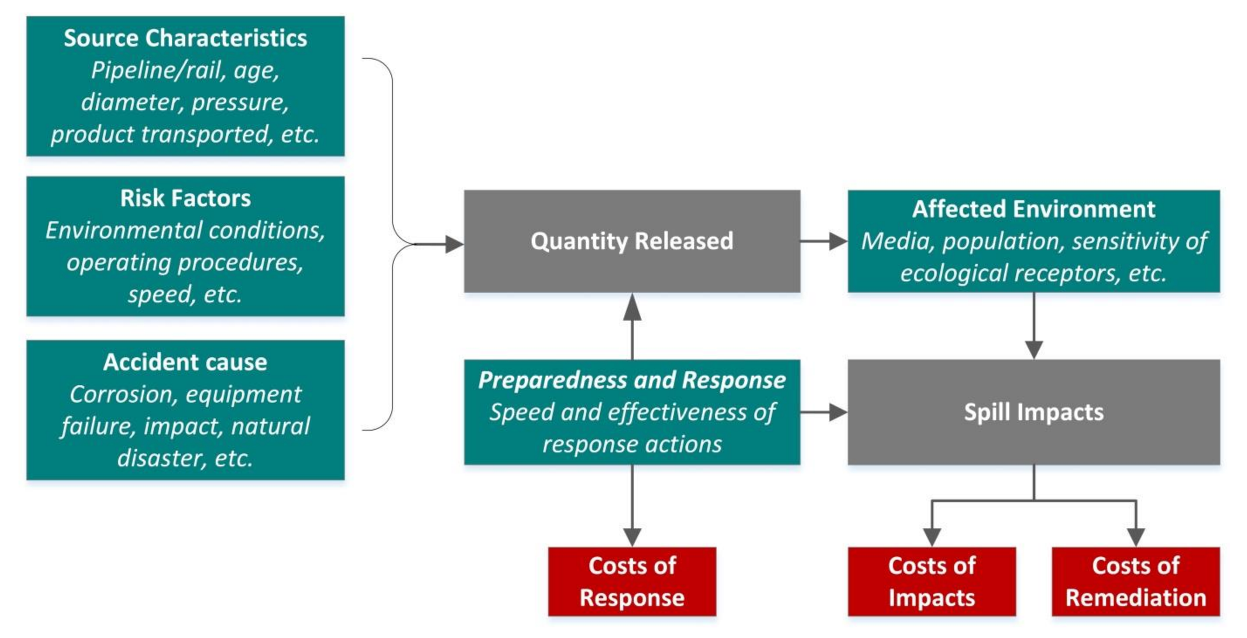
The diagram shows those relationships as simple lines between discrete factors, but the actual interplay of factors may be significantly more complex in how they affect the probability of an incident, its magnitude, and consequences. For example, material defects may be more likely for older pipes, regular assessments to detect corrosion may be more frequent for pipelines located in more densely populated areas, etc.

Through this study, we seek to determine the relationship between the costs (red boxes at the bottom of the diagram) and the multitude of factors that may contribute to the magnitude of these costs. We note that while many of these factors may also affect incident probabilities, the study focuses strictly on how the factors influence the resulting costs of an incident, once it occurs. This is one of many steps in analyzing the benefits of regulations or other policies to improve the safety of the transportation of crude oil by rail or pipeline.

---

[4] There is more literature available documenting and modeling the costs associated with marine releases of oil; however, we expect the damages and cleanup costs associated with marine spills to be potentially very different from those associated with onshore releases from railcars, and as such, this study is focused on onshore releases only.

Exhibit 2-1: Crude Oil Release Causality and Costs Flowchart



Source: Abt Associates

As noted above, the ability to relate incident characteristics to costs is one of the many elements of analyzing the benefits of safety regulation and other policies. In general, we expect policy analyses to involve the following steps:

1. **Characterizing baseline risk**: This first step involves describing the risk posed by pipeline or rail sources, for example in terms of the probability of an incident, volume of oil released, and damages. The analysis of the baseline risk may help prioritize policy interventions to reduce population and environmental exposure to crude oil and improve the efficiency of crude oil transportation. We note that the baseline risk may be driven by factors that change over time (e.g., age of the infrastructure or amount of oil transported) such that future risk may be different than that indicated by historical data.
2. **Estimating the change in risk due to the implementation of safety measures**: This second step of the analysis seeks to determine how safety measures will reduce crude oil incidents risk in terms of their probability (i.e., reduce their frequency as compared to the baseline) or consequences (e.g., reduce the amount of oil released or the resulting damages). This requires an understanding of the factors that contribute to the incidents, as well as the mechanisms by which the measures change those factors. For example, a regulation that improves preparedness through prepositioning of response equipment, training, and other measures would not be expected to change the probability of an incident occurring but may reduce the consequences of this incident through deploying oil containment booms faster, limiting the spread of oil in the environment, and reducing the extent of the contamination.
3. **Quantifying the benefits of the risk reduction**: The change in the total social costs of crude oil spills is a measure of the benefits of the risk reduction. Based on predicted changes in the frequency or characteristics of incidents over time, this step involves calculating the changes in the social costs of crude oil releases. Monetized benefit estimates facilitate evaluation of various policy options by allowing the comparison of the costs of implementing the measures (e.g., inspections, preparedness

training), to what they accomplish on behalf of society. This last step in the analysis is the focus of this study. Specifically, the models described herein provide an empirically-based approach to estimating the changes in social costs that accounts for the characteristics of the types of incidents addressed by a given rule.

Section 6.2 of this report illustrates the approach through two example analyses of hypothetical policy interventions.

2.3 Types of Costs from Crude Oil Releases

Crude oil releases from pipelines and rail cars have direct impacts in the vicinity of the release, including property damages, commodity losses, and in some cases injuries and fatalities. In addition to these direct damages, releases of crude oil from pipelines and rail cars may cause natural resource damages and have adverse economic impacts, including short-term impacts such as reduced business activity, closure of transportation routes (including highways, railways, and others), and loss of productivity. Longer-term adverse economic impacts may include depressed residential and commercial property values, decreased tourism and recreational spending, and others.

The benefits of preventing oil spills include avoiding the total social costs of the damages they cause, whether these damages accrue to the private party responsible for the spill or to a third party or society as whole and uncompensated. Exhibit 2-2 shows types of costs relevant to estimating the full damages of oil spills.

Exhibit 2-2: Examples of Types of Damages and Costs Incurred by Operators and Society as a Result of Crude Oil Releases

	Operator Expenditures	Other Costs to Society[1]
Social cost	• Property damages • Product loss • Cleanup • Government expenses reimbursement • Third-party damages paid by operator (e.g., relocation)	• Injuries • Fatalities • Human health effects from exposure to crude oil pollutants • Damage to third-party/public property (not compensated by operator) • Natural resource damages • Travel delays
Transfers	• Fines and penalties	
1. Additional costs to society that either overlap with the social costs or "net out" at a national level may include decreased property values and decreased revenues for businesses affected by releases. See discussions in Sections 2.3.4 through 2.3.6.		

As discussed in Section 1.1, an objective of this study is to estimate the social costs of crude oil releases to support the evaluation of national-level regulations and other policies. As such, it is important to differentiate between *social costs* and *transfers*. Some costs reported by operators, such as fines and penalties levied against the responsible party, are not social costs from a societal perspective but instead represent transfers to the collecting agencies. Additionally, whereas a crude oil release may temporarily reduce tourism in a particular community, it may increase tourism elsewhere as consumers shift their expenditures rather than eliminate them. In such cases, the net effect at a regional or national level is zero, even if local economic impacts are negative (or positive).

This section provides a summary of the literature documenting the impacts of oil spills, as well as information about the potential for placing monetary values on the impacts.[5] Sections 2.3.1 and 2.3.2 briefly summarize the direct expenditures by the operator and the public as a direct result of the incident, while the remaining sections discuss other types of costs arising from crude oil spills that may not be reflected in estimates of direct costs.

2.3.1 Operator Costs and Expenditures

Operator costs and expenditures on oil spills include the damage to operator-owned equipment, cleanup and remediation costs, value of the lost product, and other expenditures directly related to the release. Operators provide this information directly to PHMSA in incident reports for incidents that meet reporting criteria. Operator incident reports are a key source of the data described more fully in Section 3. While fairly comprehensive as far as operator expenditures are concerned, these reports do not include costs incurred by other parties affected by the release or impacts that have no direct monetary value.

2.3.2 Public Costs and Expenditures

Crude oil releases from pipelines and railcars often entail response actions (and associated expenditures) by local, state, and/or federal emergency response and environmental protection agencies. These entities may incur costs associated with responding to the emergency (e.g., responding to fires, coordinating evacuations) and evaluating, minimizing, and remediating environmental damages. In some cases, expenditures by public agencies are included in operator reports to PHMSA (as described further in Section 3). Additionally, crude oil releases (and associated fires) can damage public property.

2.3.3 Natural Resource Damages

According to U.S. EPA (2016), oil spills are more common and can be more destructive in freshwater bodies compared with marine environments. Because freshwater is typically less dense than ocean water, spilled crude oil has a higher chance of sinking below the surface of the waterbody (Steen et al., 1999), making cleanup more challenging. In standing water, such as wetlands and lakes, oil tends to “pool” in the water, where it remains for long periods of time, affecting habitats for years (U.S. EPA, 2016). Oil and its associated contaminants can settle into bottom sediments; cause toxic effects to exposed fish, waterfowl, reptiles, and other wildlife; cover surface, submerged, and shoreline vegetation in oil; and bioaccumulate throughout the food chain. In flowing waters, impacts may be shorter-term since currents can provide a “natural cleansing mechanism”; however, impacts to instream aquatic communities and river banks can be similar to those occurring in standing waters (U.S. EPA, 2016).

The nature and extent of impact of crude oil released into freshwater habitats depends on many factors, including the characteristics of the receiving system, the quantity of oil, and others (Steen et al., 1999). For example, coarse and organically enriched soils experience higher penetration of oil, leading to increased impacts to the roots of plants. In addition to direct and apparent impacts to ecosystems from oil releases (such as bird, fish, and mammal mortality), the impacts of oil on the ecosystem can be very complex. For example, bacteria populations may shift to spilled oil as their primary carbon source,

---

[5] Section 3.5 describes how we accounted for various additional cost categories in the compilation of the crude oil incident database.

leading to decreased decomposition of vegetation – which in turn may lead to planktonic population reductions. Removal of oiled shoreline plants can reduce the stability of streambeds and banks, increasing erosion.

Natural resource damages, including those caused by oil spills, are relatively difficult to value monetarily compared with property damages and other direct costs. One method employed in NRDAs (see Section 3.3.6 and NOAA (n.d.)) is the habitat equivalency value analysis (HEA) approach, which "seeks to determine the restoration projects that would provide ecosystem or other related services (including capital investments such as boat docks) sufficient to compensate for a loss from a natural-resource injury" (U.S. EPA Science Advisory Board (SAB), 2009; p. 52). A similar approach is the resource equivalency analysis (REA). Abt Associates and Bear Creek Economics (2016; Chapter 4) summarize the equivalency approach in three steps: 1) sum the quantity of natural resource injuries over space and time, 2) determine the amount and timing of natural resource benefits or services expected per unit of restoration, and 3) divide the total losses by the benefit per restored unit to calculate the quantity of required restoration.

However, HEA has several limitations to its utility in assigning values to lost or damaged resources (U.S. EPA SAB, 2009). The costs expended to "make the public whole" through restoration is not necessarily equivalent to the damages suffered. Additionally, while HEA should use a value-to-value approach, it is often applied on a simplified basis, such as acres-to-acres, given the difficulties in identifying, quantifying, and replicating lost ecosystem services. Overall, the U.S. EPA SAB urges caution in any attempt to use replacement costs as a proxy for value.

2.3.4 Decreased Property Values

The value of homes in close proximity to pipelines involved in a high-profile incident may be significantly and adversely affected following the release, at least in the short-term, presumably indicating adverse impacts to homeowners. Some studies have attempted to quantify such impacts.

Simons et al. (2001) conducted a hedonic analysis after a pipeline rupture in Maryland that released oil into a river, affecting 10 miles of shoreline. The study of 2,300 single-family homes found that the rupture resulted in a statistically significant reduction in the price of waterfront properties of over 10 percent in the six months following the spill.

Simons (1999) conducted a case study on the effects of a 1993 legal settlement for a BP pipeline rupture from 1962 that contaminated residential property in Summit County, Ohio. The author found that single-family homes contaminated by a well-publicized oil pipeline rupture experienced a price reduction of approximately 25 percent while remediation was underway.

The effects on property values resulting from the release of other substances may be similar to those of crude oil and vice versa, especially in cases where the release is highly publicized or causes documented environmental damages. Researchers found that other types of environmental releases also affect property values. For example, Hansen et al. (nd) conducted a hedonic analysis of home sales in Bellingham, Washington including five years of data before and after a high-profile gas pipeline release in 1999. Before the incident, there was no significant relationship between distance to the pipeline and home price; after the incident, homes closer to the pipeline had a statistically significant lower value (with the effect diminishing over time). As another example, Simons et al. (1996) conducted a study on the effect of leaking hazardous liquid underground storage tanks on residential property values in Cuyahoga County, OH. The study showed that a home located "on the same block or within 300 feet" of a registered leaking tank site sold for 17 percent less holding all else constant.[6]

2.3.5 Decreased Revenues for Local Businesses

Oil releases may affect local businesses by reducing revenues over the short term as a result of closures, decreased operating times, or decreased number of visitors to the community (for example due to injury to a resource that attracted visitors).

Zhang (2009) examined the vulnerability of businesses to environmental disasters. The study showed that certain factors can intensify the effects of environmental disasters on business including: capital immobility, inflexible labor organization, inter-business dependence, and lack of market diversification. Based on these factors, the author concluded that small businesses are more vulnerable to disaster effects than are larger businesses.

There are limited accounts that quantify commercial impacts of crude oil releases which occurred inland, and many more examples of impacts from offshore spills.[7] The May, 2015 Refugio spill in California had some documented economic impacts in the form of decreased revenues for local businesses.[8] For example, Kacik (2016) reports that several business owners filed a class action suit against the operator (Plains All American Pipelines) for lost revenues. One commercial sablefish fishing business saw a decline of over $100,000 in income between 2014 and 2015, while a sea cucumber harvester was forced to throw much of its harvest away due to contamination fears, losing about $18,000 in profits.

Decreased revenues and other adverse impacts on businesses are likely to be localized. At the national level, the net effect of such impacts are likely zero; as such, for this study, which is concerned with developing a methodology to assess the benefits of national-level policy interventions, we treat these impacts as transfers rather than social costs (see Section 2.2).

2.3.6 Decreased Recreational Expenditures

The ecosystem impacts described in Section 2.3.3 may have short- and long-term economic impacts on local communities through decreased recreational opportunities. Petroleum products in crude oil can accumulate in freshwater fish for at least 40 days after a spill, and beyond that period, fish can continue to be contaminated through bioaccumulation (Puckett, 2015; Montana Fish, Wildlife, and Parks (FWP), 2015). As such, government authorities will sometimes close recreational fishing sites or issue fish consumption advisories following oil spills.

---

6 Note that temporal differences between (ongoing) releases of underground storage tanks and a (discrete) incident involving a pipeline or rail may influence property value impacts.

7 Several articles document the business impacts of offshore releases (e.g. the BP Gulf spill). For example, Jepson and Colburn (2013) developed a set of social and economic indicators that establish the baseline economic and social characteristics of coastal communities, for use in assessing the potential impacts to these communities from fishery disruptions and management changes.

8 In addition to forgone revenues for local businesses, this incident could have substantial economic impacts to Santa Barbara County; According to Kacik (2015), if the pipeline remains dormant for three years following the release, the county would lose out on $37 million in property taxes, 155 jobs, $32 million in worker income, and $5 million in federal royalties.

For example, following a spill into the Yellowstone River in 2015, Montana FWP received many inquiries from recreational anglers regarding the safety of their catch. The agency issued a fish consumption advisory, advising anglers not to consume any fish caught in the river (Puckett, 2015). The advisory affected a stretch of the river starting upstream of Glendive and ending at the North Dakota state line[9] (Freshwater Fishing News, 2015a) and lasted from the time of the spill in January until mid-April (O'Brien, 2015). The advisory led to concerns about the impact of the spill on visitation to the region during the outdoors season (Francingues, 2015). This spill affected many popular sport fish species including paddlefish, pallid sturgeon, channel catfish, and walleye, steelhead rainbow trout, northern pike, smallmouth bass, and many others (Puckett, 2015; Freshwater Fishing News, 2015a; Montana FWP, 2015). Montana FWP personnel had a difficult time assessing impacts to sport fish due to the river being iced over. Officials requested that sport fishers donate their catch so that the agency could conduct testing to assess the extent of contamination (Puckett, 2015).

Another example is a February 2015 train derailment in Fayette County, West Virginia that resulted in the release of Bakken crude oil into the Kanawha River. According to Freshwater Fishing News (2015b), the river is "a popular destination for freshwater fishing, kayaking, canoeing, and site seeing" and "is home to smallmouth bass, musky, walleye, sauger, channel catfish, flathead catfish, sunfish, suckers, and other species."

The Kalamazoo River release in July 2010 caused extensive natural resource damages that resulted in significant losses in recreational opportunities. According to the U.S. Fish and Wildlife Service (USFWS) and others, approximately 100,000 recreational user-days were lost as a result of the incident (USFWS et al., 2015). The Trustees responsible for selecting restoration options ultimately selected projects that they believed would compensate the public for these recreational losses; these projects were expected to cost approximately $1.7 million.

Even in cases where no official fish consumption or recreational advisories are issued, publicized oil releases into freshwater ecosystems may cause decreases in visitation to recreational sites due to public perceptions of safety and cleanliness.

Abt Associates and Bear Creek Economics (2016) conducted a study of the economic impact of lost recreational fishing opportunities due to a potential "worst-case" oil spill scenario on the Columbia River in Washington State. The expected economic losses to local businesses were estimated by multiplying the number of lost recreational fishing trips times avoided recreational fishing-associated expenditures. Abt Associates and Bear Creek Economics (2016) applied avoided expenditures of $46.98 per fishing day, which was derived from USFWS and U.S. Census Bureau (2014) based on data from Oregon and Washington. As noted in the report, avoided expenditures do not represent the value of lost recreational fishing days[10]; rather, "the loss of spending by anglers can represent a disruption to local economic activity, particularly for businesses close to the affected areas and those businesses that provide services specifically for anglers, such as bait shops and marinas."

The 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (USFWS and U.S. Census Bureau, 2014) includes data on the number of days spent participating in fishing, hunting, and

---

9 Based on maps of the region, this stretch of the river appears to be at least 60 miles.

10 To estimate the value of lost fishing days to anglers, the study applies a value of $58 per fishing day based on a report prepared for Washington State.

other wildlife-related recreation, as well as associated expenditures. According to the data, Americans spent 553.8 million days fishing in 2011 (including 57.5 million days by visitors from out-of-state), spending $41.8 billion in total expenditures and $21.8 billion in trip-related expenditures. On a per-fishing-day basis, this comes out to $75 per fishing day in total expenditures and $39 in trip-related expenditures.

Other resource-dependent activities may also be adversely impacted by oil releases, including hunting and wildlife viewing. Nationally, hunters spent 281.9 million days hunting (including 20.3 million days by visitors from out-of-state), with 23.3 million days spent hunting migratory birds. These hunters spent $33.7 billion including $10.4 billion in trip-related expenditures (USFWS and U.S. Census Bureau, 2014). On a per-hunting-day basis, this comes out to $120 in total expenditures per day and $37 in trip-related expenditures.

For wildlife viewing away from home, participants spent $54.9 billion including $17.3 billion for trip-related expenditures, over a total of 335.6 million days spent wildlife viewing. 67.2 million of the days were by visitors from out-of-state (USFWS and U.S. Census Bureau, 2014). On a per-day basis, wildlife viewing expenditures come out to $163 per trip in total expenditures, including $51 in trip-related expenditures.

To the extent that anglers, hunters, and wildlife watchers visit alternative sites rather than canceling trips to affected resources, decreased trip-related expenditures represent transfers rather than social costs. Additionally, even in cases where travelers cancel trips and forgo trip-related expenditures, the social costs are not equal to avoided expenditures. Rather, the expenditures themselves are transfers between sellers and buyers, and the consumer surplus that is lost as a result of avoided expenditures is a cost to society.

2.3.7 Travel Disruptions

Oil spills, particularly those occurring on and around rail tracks, may cause travel delays and disruptions for passengers and freight carriers.

An example is a 2015 train derailment in Culbertson, Montana that involved the release of approximately 35,000 gallons of crude oil and caused extensive passenger train delays. According to Pholphiboun (2015), Amtrak passengers across the state each faced increased travel times of up to 40 or 50 hours as a result. That same incident also resulted in the detour of all vehicular traffic within a mile of the site (Klemann, 2015).

These types of delays may have substantial economic impacts, both to passengers facing delays, and to industries that rely on rail and truck shipments to move their products. The U.S. DOT[11] issues guidance regarding the value of travel time for passengers traveling locally and inter-city (U.S. DOT, 2015a), based on a variety of factors such as trip purpose (business or personal reasons),[12] income, comfort during

11 Other federal agencies may also develop estimates of the value of time spent traveling (such as the United States Army Corps of Engineers in the context of flood modeling); however, we have not explored these alternatives further, since, as described in Section 3.5.3 below, we do not have incident-specific data to apply such estimates.

12 For example, U.S. DOT (2015a) estimates that 95.4% of local travel by surface modes is personal, and the remaining 4.6% is business-related. For intercity travel, the breakout is 78.6% personal and 21.4% business.

travel, and personal characteristics. The guidance yields estimates of $13 per hour for all-purpose local surface travel, and $19 per hour for intercity surface travel. Professional drivers (truck, drivers, bus drivers, and transit rail operators) have higher travel time values, ranging from $26 to $46 per hour.

The U.S. DOT developed its guidance for the purpose of valuing travel time *savings* associated with regulatory actions. Consumers may have different values for unexpected travel *delays*. However, applying a rough estimate of the value of travel time of $19 per hour for a typical intercity traveler to the Culbertson incident described above, the economic impacts of travel delays may be in the range of up to $760 to $950 per passenger (assuming a delay of 40 to 50 hours).

The U.S. DOT does not have guidance on the value of time in freight transportation, which is considerably more complex than passenger travel (U.S. DOT, 2015a). Some studies have attempted to place a value on freight travel delays. For example, Mesa-Arango, et al. (2013) developed economic models to estimate the economic impacts of freight disruptions, including direct, indirect, and induced impacts. They applied their economic impacts model to a case study of four days of highway closures in northern Indiana in 2008 from flooding. Inputs to the model included identification of the highway stretches closed, values of travel time, vehicle operating cost (including cost of fuel, shipping inventory, vehicle repair and maintenance, and other factors), network and traffic flows in the study region, proportion of trucks per link in each commodity, and others. Using the model, the authors estimated that the total economic loss during the delay in shipments of commodities in the study region was $11.2 million in output, 60 jobs, $3.4 million in labor income, $1.2 million in taxes, and $54.5 million in value added.[13]

Another example is Weinstein and Clower (1998), a study of the effects of Union Pacific Railroad Service Disruptions on the Texan economy between July 1997 and January 1998. The disruptions and delays were caused by a variety of factors including derailments, crashes, and general congestion. The study estimates that these disruptions resulted in short-term costs of $1.093 billion to Texas producers and growers due to higher shipping costs.[14] These costs are based on delays attributable to various causes over an entire rail network and not a specific incident involving the release of crude oil; however, they still provide a sense of potentially far reaching impacts of railroad transportation disruptions. A spill resulting in extensive clean-up and delays on a particularly critical rail segment may have broad-ranging impacts.

Lovett, et al. (2015) estimated the costs of freight train delays to various stakeholders, including railroads (broken out into crew, locomotives, fuel, railcars, and lading), shippers (inventory devaluation and holding costs), and the public (increased cost of rail transportation, costs of purchased goods, emissions, and level crossing delays). The extent of each cost component varies based on numerous factors including the number of locomotives/railcars, railcar length, cargo transported, whether the train is a unit, manifest, or intermodal train, loading/unloading frequency and time. Exhibit 2-3 summarizes the costs associated with train delays based on the train type and route length, including all cost categories accruing to railroads, shippers, and the public.

---

13 “Value added” included employee compensation, proprietor income, indirect business taxes, and other property type income such as payments from interest, rents, royalties, dividends, and profits.

14 These costs are broken down as follows, $400 million for the chemical industry, $150 million for agriculture, $292 million for paper and forest products, $146 million for building materials, $25 million for electric utilities, and $80 million for retail trade.

Exhibit 2-3: Variable Hourly Railroad Train Delay Costs for Different Route Lengths

Cost Category	805 km (500 miles)	2,012 km (1,250 miles)	3,219 km (2,000 miles)
Bulk trains	$834	$834	$834
Manifest trains (no lading cost)	$900	$900	$900
Manifest trains (with lading cost)	$1,423	$1,423	$1,423
Intermodal (no lading cost)	$1,062	$1,138	$1,258
Intermodal (with lading cost)	$2,215	$2,291	$2,411
Source: Lovett, et al. (2015; Table 11)			

Lovett, et al. (2015) provided estimates of the costs associated with "level crossing" delays – or the time that drivers are delayed by trains blocking road crossings. This estimate is based on a series of equations using the track speed, average train length, average time the crossing is activated before a train reaches the crossing and after is passes, vehicle arrival rates, and others. For a four-hour delay, these calculated costs ranged from $2,255 to $3,239 for different train types and route lengths.[15]

Hallenbeck, et al. (2014) also estimated the costs associated with road closures; they conducted a travel-cost study associated with the closure of state highways in Washington State that could occur as a result of flooding. Their route-specific models account for the added costs of time and vehicle mileage associated with available detour routes, as well as the costs associated with canceled trips. The study does not account for economic losses associated with delayed delivery of goods or services, losses in economic activity attributable to travelers being unable to reach their destinations, or economic losses associated with goods not being delivered. Exhibit 2-4 summarizes the results of the Hallenbeck et al. (2014) study on each road, as well as a calculation of per-hour costs.

Exhibit 2-4: Travel Costs Due to Road Closures Resulting from Potential 100-Year Flood in Centralia/Chehalis Basin

Road	Estimated Closure Length and Duration	Travel Costs Due to Closure	Per-Hour Travel Costs Due to Closure[1]
I-5	20 miles for 123 hours	$11,872,000	$96,520
US-12	Approximately 2 miles[2] for152 hours	$340,000	$2,237
SR-6	Multiple portions for 51 hours	$114,000	$2,235
Source: Hallenbeck, et al. (2014) *1. Travel costs due to closure divided by estimated closure duration.* *2. Estimated based on map of closure.*			

2.4 Incident Cause

Incident cause is an important consideration in estimating the social costs of crude oil spills. The effect is twofold. First, different incident causes may lend themselves to different policy remedies to prevent the spills in the first place. For example, incidents caused by corrosion, material defects and other detectable conditions may be prevented through better inspection technologies and more stringent inspection requirements. Second, the cause of the incident may determine the release mechanism and quantity released, thereby affecting the magnitude of the damages. For example, train derailments may result in the

15 Assuming a railcar length of 19.8 m, a road user cost of $12.98 per person-hour, 3.2 km level crossing spacing, a vehicle arrival rate of 500 vehicles/day, average crossing length of 9.1 m, average crossing activation time of 30 seconds, and average vehicle occupancy of 1.36.

release of relatively greater quantities of oil but are more likely to be detected, as compared to a leaking valve.

The literature provides limited insight into the relationships between incident causes and their consequences and damages. The majority of literature on oil spill causes focuses on pipeline releases rather than releases from rail cars.[16] Many of the studies consider spill cause as it relates to incident probability, with such studies generally falling into two categories. The most common category of studies consider failure cause probability based on historical data, with the goal of improving maintenance practices (Dawotola et al., 2012; Hovey & Farmer, 1993). These types of studies develop estimates primarily focused on a single failure cause, with corrosion being most common. The second common category consists of studies that combine historical data with theoretical models of failure, while applying a variety of econometric and statistical techniques including fuzzy-based bow tie analysis (Shahriar et al., 2012), combined analytical hierarchy process-fault tree analysis (Dawotola et al., 2009), and Monte Carlo analysis (Caleyo et al., 2009). These studies are similarly concerned with better understanding the probability of spills by cause with the goal of improving pipeline safety through more efficient and effective maintenance. Although a few of the second type of studies connect spill cause to impacts theoretically, none of the studies we reviewed does so quantitatively.

Two studies link pipeline spill cause to resulting damage costs. First, Restrepo et al. (2009) analyze PHMSA data on hazardous liquid pipeline accidents between January 2002 and 2005. Crude oil represents the most frequent hazardous liquid in the data the authors use, accounting for more than a third of 1,582 accidents analyzed. The authors sort the spills with available cause data into 25 cause categories. The authors then regress these spills against three different available impact measures – product loss cost, property damage cost, and cleanup and recovery cost. In addition to the cause category, the monetary impact regressions include spill characteristics (gallons lost, location onshore or offshore, system part involved, presence of an explosion, and location within a high consequence area).[17]

Overall, the authors find that there is “significant variation in the consequence measures examined depending on the factors affecting the accident” (Restrepo et al., 2009). Many of the cause categories that the authors analyze do not have a statistically significant impact on the costs. However, some cause categories have a high statistical significance ($p < 0.01$) effect for certain costs. Causes that tend to increase cleanup and recovery costs include earth movement, fire/explosion, and rupture of previously damaged pipes. Causes associated with lower cleanup and recovery costs include corrosion, incorrect operation, and broken couplings/seals. Spills caused by internal corrosion tend have lower product loss costs. These results are presented in Exhibit 2-5.

---

16 Relatively few studies explore quantitatively the cause of failure or impact for crude oil spills from rail transportation. The comparatively lower number studies may be because of the smaller volume of crude oil transported by rail as compared to pipeline (Atkin, 2015; Young, 2014; Frittelli et al., 2014).

17 Note that the authors do not appear to control for the type of hazardous liquid involved in the incident.

Exhibit 2-5: Effect of Spill Cause on Product Loss, Property Damage, and Cleanup/Recovery Costs from Restrepo et al. (2009)

Cause	Product Loss	Property Damage	Cleanup/Recovery
Corrosion, external	+	-	-***
Corrosion, internal	-***	-	-***
Earth movement	+	+	+***
Lightening	-	-	-
Heavy rains/floods	-	-	-
Temperature	-	+	+
High winds	+	+*	+*
Operator excavation damage	-	-*	-*
Third party excavation damage	+*	-	-
Fire/explosion as primary cause	+	n/a	+***
Car, truck or other vehicle not related to excavation activity	+	+	+
Rupture of previously damaged pipe	+	+	+***
Vandalism	+	-	-
Body of pipe	+	+*	+*
Component	-	-*	-
Joint	+	-	-
Butt weld	-	+	+
Fillet weld	-	+	+
Pipe seam weld	+**	+	+*
Malfunction of control / relief equipment	-	-**	-***
Threads stripped, broken pipe coupling	-	-**	-***
Ruptured or leaking seal / pump packing	-	-	-***
Incorrect operation	-	-	-***
Miscellaneous	-	-	-**
Unknown	+	+	-

Note: Exact values are not presented, as the coefficients are subject to the specified functional form.
"-" = relationship with negative slope, "+" = relationship with positive slope
** = statistically significant at $p < 0.1$*
*** = statistically significant at $p < 0.05$*
**** = statistically significant at $p < 0.01$*

Girgin and Krausmann (2016) focus on the impact of spills caused by natural hazards to the crude oil pipeline system across different transportation stages, including the pipeline, storage facilities, stations, and terminals. The dataset is based on PHMSA data from 1986-2012, supplemented with National Response Center (NRC) data. The authors further refine the data by analyzing spill narratives to capture additional natural hazard spills.[18] The authors supplement the cost data with National Transportation Safety Board (NTSB) reports and lawsuit settlements, and note a high degree of uncertainty in cost estimates with frequent and substantial underestimates. Using this dataset, the authors analyze how

18 Overall, the authors found that 76 percent of natural hazard incidents were correctly identified in the original database. However, cause identification was more accurate in more recent data, and 62 percent of misidentified natural hazard incidents occurred between 1986 and 2001.

different natural hazard incident causes correspond to different transportation stages, geographic locations, and incident characteristics such as spill volume, cost, and indicated significance. Overall, the study finds that while natural hazard triggered incidents are less frequent than incidents from other causes, the "consequences are comparatively more serious," including more frequently being flagged as 'significant' incidents and with higher total damages.

Within the different natural disaster causes, Girgin and Krausmann (2016) find that for pipelines, geological hazards such as earthquakes or frost heave were the most frequent triggers, accounting for 43 percent of all natural hazard incidents. The largest share of volume spilled across all transportation phases is due to hydrological hazards such as floods or stream erosion, which are responsible for 110,000 barrels (bbl), out of 318,000 bbl released. Meteorological hazards such as heavy rainfall, tropical cyclone, or lightning are the largest category with respect to economic costs, accounting for $319 million out of $596 million across all natural hazard incidents. The authors also note substantial geographic variation across types and severity of natural hazard spills, and that natural hazard spills are different from other causes, since natural hazards may increase the spread of oil spilled and/or hinder response operations.

3 Incident Data Sources and Compilation Methodology

The database includes information about the circumstances, characteristics, and costs or damages of historical crude oil spill incidents from either rail or pipeline. In assembling the database, our objective was to compile data for a representative set of crude oil spill incidents to support analysis of the factors that contribute to spill costs.[19] As such, we searched for a variety of data sources that may provide information about the characteristics of incidents (e.g., transportation mode and train speed at the time of the incident), the incident location and affected receptors (e.g., land use and population density in the immediate vicinity of the release), costs and damages resulting from the incident (e.g., emergency response costs, public and private property damage, and injuries), and other factors.

In compiling the database, we searched for the type information summarized in Exhibit 3-1 for onshore releases of crude oil from rail cars and pipelines between 2005 and June 2016. In collecting and parsing data on the costs and damages resulting from crude oil releases, we distinguished between *social costs* and *transfers*, and only included those incident costs that are truly costs to society considering the national perspective for the analyses. For more discussion of the difference between social costs and transfers, see Section 2.3.

The PHMSA pipeline and rail incident datasets contribute the universe of incidents included in the database. In using these datasets as the basis for defining the universe of relevant incidents, we assumed that all incidents involving the release of crude oil from pipeline or rail have been reported to PHMSA, as required under Title 49 of the Code of Federal Regulations. Operator-reported data comprise the primary source of information about the incidents. However, operator-reported data do not fully reflect the damage to society that can result from a release of crude oil. To maximize the information included in the database about each incident and to provide a more comprehensive value for crude oil spills, we supplemented the PHMSA datasets with additional sources of information, such as other federal government datasets and incident-specific reports and accounts.

The rest of this section describes the data obtained from primary and supplementary data sources, which we used to build the database. Section 3.1 describes the primary pipeline incident datasets from PHMSA and the selection of relevant incidents involving crude oil releases. Section 3.2 describes the PHMSA rail dataset. Section 3.3 describes the other sources of information that supplement the PHMSA datasets in the database. Section 3.4 describes additional data derived from a spatial analysis of the incidents. Finally, Section 3.5 describes data adjustments and quality assurance processes employed in compiling the database. For a discussion of data gaps and uncertainties, see Section 7.

[19] This dataset is not meant to support estimates of incident probabilities, but is instead meant to support estimates of the costs resulting from a spill, conditional on a spill occurring.

Exhibit 3-1: Desired Information about Rail and Pipeline Incidents

Category	Incident Data
Basic Information	• Transportation mode (rail or pipeline) • Incident date • Incident location (state, county, zip code, latitude/longitude) • Quantity of crude oil released
Incident Context	• Incident cause • Train type, number of cars, and speed • Fire indicator • Weather conditions • High Consequence Area (HCA) indicator • Response time • Surface water, groundwater contamination, and/or soil contamination • Pipe diameter, pressure, thickness, and manufacture year • Population density in vicinity • Land use in vicinity
Non-monetary Damages	• Fatalities • Injuries • Type and duration of travel disruptions • People evacuated • Miles of river damaged or acres of wetland impacted • Wildlife impacted
Monetary Damages	• Property damages (public and operator) • Cost of commodity lost • Emergency response costs (public and operator) • Environmental remediation costs (public and operator) • Additional costs (public and operator)

3.1 PHMSA Pipeline Incident Data

Under Title 49 of the Code of Federal Regulations (49 CFR Parts 191 and 195), pipeline operators must report incidents involving any release of hazardous liquids that are 5 gallons or more, or that result in an unintentional fire.[20] Operators are required to report information to PHMSA within 30 days of an incident, and PHMSA compiles the information into publically available datasets.

We compiled data from two separate PHMSA datasets of pipeline incidents. One dataset includes incidents reported for calendar years 2005 through 2009, and the other includes incidents between 2010 and mid-2016.[21] The two separate datasets reflect changes in the reporting requirements that became effective in 2010 and that expanded the scope or details of the data operators must provide about each incident. Whereas most of the data fields in the two incident datasets are the same, starting in 2010 the reporting forms include some information not previously requested, such as additional information about incident causes and consequences.

20 See 49 CFR Part 195.50 for additional information and details about the reporting requirements.

21 Both data sets were retrieved on 11/9/2016 from: http://www.phmsa.dot.gov/pipeline/library/data-stats/distribution-transmission-and-gathering-lng-and-liquid-accident-and-incident-data

To construct the database, we first downloaded the available data in Excel spreadsheet format for each of the two datasets. We then filtered the full sets of all hazardous liquid releases to retain only those incidents involving the release of crude oil (based on the "Commodity_Released_Type" field in the 2010 to 2016 dataset and the "Class_Text" field in the 2005 to 2009 dataset). We then excluded pipeline incidents that occurred offshore (i.e., those affecting marine waters away from land), since we expect offshore incidents to involve different response actions or to cause fundamentally different damages than onshore incidents.[22] The original PHMSA datasets identified 61 incidents as occurring offshore.[23]

3.1.1 PHMSA Primary Pipeline Incident Data

For each relevant incident that occurred between 2005 and 2009, we pulled data from the PHMSA incident report dataset, including the date, location, commodity spilled, volume released, and monetary losses associated with the incident[24] (Part A); pipe age, thickness and other characteristics (Part D); fatalities, injuries, and other consequence information such as wildlife, water, and soil impacts (Part F); incident cause (Part H); and a narrative description of the incident (Part I).

Similarly, for each relevant incident that occurred in 2010 or later, we pulled data from the PHMSA incident report dataset, including the date, time, location, commodity spilled, volume released, and associated injuries and fatalities (Part A); pipe age, thickness, and other characteristics (Part C); additional damage and consequence information such as wildlife impacts, high consequence area (HCA) indicators, and monetary estimates of damages[25] (Part D); operating conditions at the time of the incident (Part E); the cause of the incident (Part G); and a narrative description of the incident (Part H).

3.1.2 Additional Pipeline Data from PHMSA Flagged Files

For the purposes of conducting trend analyses, PHMSA supplements the reported data (from both datasets) in "Incident Flagged Files."[26] In some cases, these flagged files provide updated, revised, or supplementary information about incidents. For each of the onshore crude oil incidents that we pulled in the two primary datasets, we extracted the corresponding data from the flagged files.

---

22 Offshore incidents are included in the raw data files but we did not seek data on these incidents from other sources to complement information provided to PHMSA, nor did we include these incidents in the data summaries or analyses presented in Sections 4 and 5 of this report.

23 See Section 3.4 for a description of a spatial analysis of incidents that identified an additional 19 offshore incidents.

24 Monetary losses include public/community losses reimbursed by the operator (including property damage, emergency response, environmental remediation, and others), and operator losses (including value of product lost, property damage, and other costs).

25 Damages include public and non-operator private property damage, cost of commodity lost, cost of operator's property damage and repairs, cost of operator's emergency response, cost of operator's environmental remediation, and other costs.

26 Available from: http://www.phmsa.dot.gov/pipeline/library/data-stats/flagged-data-files

3.1.3 Pipeline Incidents

The PHMSA dataset for incidents occurring between 2010 and June 2016 includes 1,276 onshore releases of crude oil. The pre-2010 dataset includes 796 onshore releases of crude oil. All 2,072 of these pipeline incidents had data in the flagged files (which includes updated information for a limited number of incidents for a limited number of fields, but otherwise reflects the data provided in the underlying datasets) and as such, we used the flagged files as primary source.

3.2 PHMSA Rail Incident Data

The Hazardous Materials Regulations at 49 CFR Parts 171-180 require that certain types of incidents be reported to PHMSA. Under §171.16, operators are required to report information to PHMSA within 30 days of an incident that meets broadly defined criteria that include: "(1) Any of the circumstances set forth in §171.15(b); (2) An unintentional release of a hazardous material or the discharge of any quantity of hazardous waste; (3) A specification cargo tank with a capacity of 1,000 gallons or greater containing any hazardous material suffers structural damage to the lading retention system or damage that requires repair to a system intended to protect the lading retention system, even if there is no release of hazardous material; or (4) An undeclared hazardous material is discovered."[27] PHMSA compiles the information provided by operators in Form DOT F5800.1 (OMB Approval No. 2137-0039) into publically available datasets.

3.2.1 PHMSA Primary Rail Incident Data

The PHMSA rail incident data set is the primary source for rail incidents included in the database. We downloaded all incidents that occurred in calendar years 2005 through June 2016,[28] and identified those incidents that involved the release of crude oil based on the United Nations commodity identification number ("UN1267"). Following discussion with PHMSA staff, we also included one additional incident that occurred on January 31, 2014 in Perry County, Illinois that involved the release of crude oil but did not use the UN1267 identifier (PHMSA incident report X-2014020292).[29]

For each relevant incident, we pulled data including the date, location, train speed, quantity released, incident cause, fatalities and injuries, monetary damages,[30] and a narrative description of the incident. In the rail dataset, incidents involving more than one train car have multiple observations, with the quantity

[27] Under §171.16(c), operators must update their incident reports within one year of the date of occurrence of the incident whenever: (1) A death results from injury caused by a hazardous material; (2) There was a misidentification of the hazardous material or packaging information on a prior incident report; (3) Damage, loss or related cost that was not known when the initial incident report was filed becomes known; or (4) Damage, loss, or related cost changes by $25,000 or more, or 10% of the prior total estimate, whichever is greater.

[28] Obtained from: https://hazmatonline.phmsa.dot.gov/IncidentReportsSearch/IncrSearch.aspx.The database reflects incident reports received as of the retrieval date of 11/9/2016.

[29] See discussion of New Augusta incident in PHMSA (2015; p. 85).

[30] Monetary damages include public property damage, cost of commodity lost, operator property damage, operator emergency response cost, operator environmental remediation costs, and other operator costs.

released from each train car reported separately. As such, we consolidated the data into a single observation for each incident by summing across the quantity released for each car to calculate a total amount released.[31]

3.2.2 Rail Incidents

The PHMSA rail dataset includes data on 456 releases of crude oil between 2005 and June 2016.

3.3 Supplementary Incident Data

After identifying relevant pipeline and rail incidents based on the PHMSA datasets, we supplemented the database with information from other datasets and incident-specific reports. Supplementary data sources include the Coast Guard's National Response Center (NRC), Pollution reports (POLREPs) prepared by EPA's On-Scene Coordinators (OSCs), the Federal Rail Administration (FRA), settlements from the U.S. Coast Guard's National Pollution Funds Center (NPFC), NRDAs prepared by federal and state agencies or responsible parties, and public financial statements and other reports of companies responsible for the incidents. We describe the data obtained from each of these other sources below.

We note that reporting requirements vary across data sources and that some incidents reported to PHMSA are not included in other datasets that have different scope or reporting requirements. For these incidents, we rely only on the information reported to PHMSA in the incident report.

Conversely, some incidents may be reported in other data sources, but not in the PHMSA datasets. For the purpose of this study, we assumed that the PHMSA datasets provide a complete inventory of the incidents relevant to this analysis – i.e., they cover all crude oil releases from rail and regulated pipelines. Accordingly, we assumed that additional incidents reported in other sources (such as NRC) and therein described to involve "pipelines" involved lines or equipment not within PHMSA's jurisdiction.

3.3.1 National Response Center

NRC is the federal point of contact for reporting all hazardous substances releases and oil spills that trigger notification requirements under several laws. The NRC database compiles information about reported hazardous substance releases and oil spills, including those associated with pipeline and rail transportation. This dataset includes information about the incident cause, weather conditions, quantity of material released, remediation efforts undertaken, and additional contextual and response information. The data reflect information available at the time NRC was notified of the release. The NRC database does not capture actual or estimated spill costs. Although remediation efforts and involved agencies are sometimes noted, the main purpose of the NRC database is to record the initial incident notification and the information is typically not updated to reflect details that become known only later. Because of these limitations, for this effort, we used the NRC data primarily to verify or complement data provided in the PHMSA datasets.

[31] Note, however, that the same total incident damages are reported for each rail car; as such, we did not sum across cars when we pulled in the damage estimates.

Of the 2,528 onshore crude oil releases in the database, 607 incident reports included an identifier for the NRC database. For those incidents, we pulled available NRC data on the location of the incident and the quantity released.

3.3.2 EPA Pollution Reports

POLREPs are prepared by EPA's OSCs who monitor or direct responses to oil spills or hazardous substance releases reported to the federal government and occurring in inland areas and waters. The POLREPs contain information about the incident location (including latitude and longitude for some), background and context for the incident, current and planned removal activities, and in some cases, estimates of cleanup costs and expenditures by responsible parties. The EPA OSC updates these reports periodically during the response and cleanup effort. For this study, we reviewed POLREPs for responses initiated under the authority of the Oil Pollution Act (OPA) of 1990 to identify those incidents originating from rail or pipeline and involving crude oil. We matched POLREPs to incidents in the database based on the location and date of the incident, as well as other details as needed.

We found 44 POLREPs corresponding to incidents in the database. Of those, 19 reports include supplementary information for the incident that goes beyond information included in the PHMSA datasets, particularly with regard to emergency response expenditures (for which 8 reported information). These reported expenditures are generally presented in the form of approved ceilings on expenditures by EPA. Some POLREPs also provide information about affected surface waters and/or the latitude and longitude of the incident site. For each incident with a POLREP, we pulled this data from the latest (i.e., most up-to-date) available report. The POLREPs also include the federal project number (FPN) associated with the incident, which corresponds to NPFC claims as described in Section 3.3.4.

3.3.3 Federal Railroad Administration

FRA receives reports from rail carriers about "all accidents and incidents resulting in injury or death to an individual or damage to equipment or a roadbed arising from the carrier's operation" under the Accident Reports Act (Public Law No. 165). Since FRA data are not focused on the release of oil (or other substances) resulting from incidents, they typically do not provide detailed information about the release or the associated damages. Further, FRA has substantially different reporting requirements for incidents, and we expect that the majority of incidents involving stationary trains are not reflected in the FRA data.

However, for those incidents reported in both the FRA and PHSMA data sets, FRA provides additional information about the context of the incident, including the latitude and longitude (which is not included in the PHMSA rail data). Since there is no common identifier or crosswalk between the FRA and PHMSA datasets, to match incidents between the two datasets, we generated a list of all possible matches based on the recorded state, year, and month of the incident. This produced 33 different possible configurations, including multiple possible matches for some FRA incidents. We then reviewed the suggested matches manually to identify 15 common incidents based on consistency across narratives, cause of spill, location, size of spill, material spilled, and emergency response. Of those 15 incidents, 11 had latitude and longitude data in the FRA dataset; we pulled these coordinates into the database.

3.3.4 National Pollution Funds Center

The U.S. Coast Guard's NPFC manages the Oil Spill Liability Trust Fund to facilitate cleanup and compensation for oil spills.[32] For this study, we used data provided by NPFC summarizing the "limit of liability" for all incidents with an FPN. These data include all removal and response costs expended by Federal OSCs plus claims paid related to an FPN. We matched incidents in the PHMSA database to FPNs (based on POLREPs, as described in Section 3.3.2) and ultimately to payments from the NPFC to reimburse federal agencies for emergency response and/or environmental remediation expenditures, based on the total expenditures. The spreadsheet provided by NPFC does not include details about the nature of the expenditures; we assumed that they represent expenditures on emergency response and/or environmental remediation.

Of the incidents for which we identified FPNs, 23 of them corresponded to a record in the NPFC dataset.

3.3.5 National Transportation Safety Board Reports

NTSB conducts investigations into all major transportation accidents in the United States. These investigations typically yield accident reports[33] which provide details about the cause of and response to incidents, as well as some information about estimated damages. NTSB does not conduct a detailed investigation for every incident involving the release of crude oil and the accident reports do not provide a comprehensive view of the damages and costs associated with the incident; however, these reports provide some supplementary information about larger and more consequential spills.

For this study, we reviewed four NTSB reports on crude oil releases from pipeline or rail sources.

- NTSB (2012) report on a 2010 release of crude oil from an Enbridge pipeline in Marshall, Michigan (PHMSA incident 20100181). According to the report, the spilled oil saturated a surrounding wetland and contaminated the Talmadge Creek and the Kalamazoo River. As of October 2011, the costs associated with the incident had exceeded $767 million and 320 people reported some adverse health effects consistent with crude oil exposure (including headaches, nausea, and respiratory problems). Included in the $767 million in estimated costs is $42 million in public emergency response costs, which we added to the database. However, we did not add the remaining $725 million to the database under other cost categories, for two reasons: First, it was not clear how much of those costs were for emergency response (the NTSB report has "regulatory support" as part of the costs), and second, more recent data on operator expenditures are available from other sources (see discussion of Enbridge financial reports in Section 3.3.7).
- NTSB (2013) report on a 2010 release of crude oil from an Enbridge pipeline in Romeoville, Illinois (PHMSA incident 20100221). The report states that the release involved 6,430 barrels of Saskatchewan heavy crude oil and $46.6 million in damages and environmental remediation costs plus $550,000 in federal oversight and response costs. As a result of the release, 50 people were evacuated from 11 nearby businesses, and 23 area businesses were closed for up to 9 days. Environmental damages included 15,000 cubic yards of contaminated soil, 32 deceased animals, and

[32] For more information, see U.S. Coast Guard, National Pollution Funds Center. 2016. "About NPFC." http://www.uscg.mil/ccs/NPFC/About_NPFC/default.asp

[33] Available from http://www.ntsb.gov/investigations/AccidentReports/Pages/AccidentReports.aspx.

141 turtles and frogs that were treated by a wildlife response center before being released. We added $550,000 for public emergency response to the database; we did not add other costs from the NTSB report to the database because the costs were not categorized, and more recent data on operator expenditures are available from other sources (see discussion of Enbridge financial reports in Section 3.3.7).

- NTSB (2014) is a preliminary report on a 2013 derailment in Casselton, North Dakota (PHMSA incident X-2014010238). According to the report, more than 400,000 gallons of crude oil were released, 1,400 people were evacuated from the area, and damages were estimated at $6.1 million. However, the report is preliminary and investigations are ongoing; as such we did not enter preliminary estimates from this report into the database, even though we note that the NTSB reported damages are greater than those included in the database ($2.5 million), based on other sources.
- NTSB (2016) is a report on a 2014 derailment in Lynchburg, Virginia (PHMSA incident X-2014050225). According to the report, 29,868 gallons of Bakken crude oil were released from a car that was partially submerged in the James River, resulting in a large fire and the evacuation of 350 nearby residents and 20 businesses. Damages were estimated at $1.2 million, not including environmental remediation. We did not add this estimate to the database, however, since more detailed information on costs were provided in the PHMSA dataset.

3.3.6 Natural Resource Damage Assessments

Section 2.3.3 of this report describes natural resource damages associated with oil spills, as well as some of the methodologies and challenges of placing values on such damages.

Under OPA of 1990, federal and state agencies including the National Oceanic and Atmospheric Administration (NOAA) and the USFWS conduct NRDAs to quantify the damages to natural resources that result from the release of contaminants to the environment, including crude oil. NRDAs conducted under the OPA identify restoration that will "restore, rehabilitate, replace, and acquire the equivalent" of injured natural resources and services.[34]

NRDAs provide detailed information on the magnitude of the damages and long-term impacts of crude oil releases, but are available for selected incidents only and often many years after the incident occurred.

For this study, we reviewed information about selected incidents for which assessments had been completed to identify any damages or costs not reflected in the database, and where possible, supplement the existing estimates. Our review included information from regional USFWS NRDA sites[35] as well as two national NRDA databases, including the U.S. Department of the Interior's Natural Resource Damage

34 See NOAA (n.d.) for more information.

35 For example, see the Midwest Region's NRDAs at http://www.fws.gov/midwest/es/ec/nrda/index.html.

Assessment and Restoration Program (NRDAR Program),[36] and the NOAA Damage Assessment, Remediation, and Restoration Project (DARRP),[37] and some state-level information.[38]

We identified five NRDAs (or equivalent assessments) corresponding to incidents in our database, further described below:

- Enbridge Marshall pipeline spill, 2010 (PHMSA incident 20100181);
- ExxonMobil Yellowstone River pipeline spill, 2011 (PHMSA incident 20110262);
- Marathon Pipeline spill, 2008 (PHMSA incident 20080272);
- Chevron Red Butte Creek pipeline spill, 2010 (PHMSA incident 20100146); and
- Refugio Beach pipeline spill, 2015 (PHMSA Incident 20150224).

USFWS, et al. (2015) is a final NRDA report for the 2010 release of crude oil by Enbridge near Marshall, Michigan (PHMSA incident 20100181). The document includes a detailed summary of the incident and the natural resources impacted by the crude oil, an overview of restoration actions completed and underway, and a discussion of restoration options "to enhance natural resources affected" by the discharge. According to the report, 38 miles of the Kalamazoo River were contaminated with oil, including wetlands, floodplains forests, residential properties, farmland, and commercial properties. From July 2010 to June 2012, 39 miles of the river were closed to public access, and many uses (including fishing and swimming) were further restricted as late as 2014. The NRDA summarizes a variety of damages to natural resources resulting from the release, including 1,560 acres of in-stream habitat oiled, 2,887 acres of floodplain oiled (with 299 acres having residual oil observed), 185 acres of upland habitat impacted by response actions, and wildlife deaths including 52 birds, 40 mammals, 106 reptiles, and 42 fish. Approximately 100,000 recreational user-days were lost (USFWS et al., 2015).

The Trustees responsible for selecting restoration options "selected compensatory restoration projects which they believe will enhance the natural recovery of resources injured…and/or will provide additional resource services to compensate the public for interim losses pending response and remedial actions, restoration required by the State Settlement and natural recovery" (p. 68). The selected projects are expected to cost approximately $1.7 million, in addition to $1.6 million in past Trustee costs and $0.6 million in future Trustee costs. However, it is unclear whether there is overlap between this cost estimate and Enbridge's planned expenditures of $1.1 billion documented in their 2013 annual financial report (see discussion in Section 3.3.7); as such, because we added costs based on Enbridge's financial report and to avoid potentially double-counting the same costs, we did not add the NRDA costs to the environmental remediation costs already in the database for this incident.

---

36 Maps of incidents and exportable databases are available through the NRDAR Program at https://www.cerc.usgs.gov/orda_docs/. We reviewed all incidents involving the key word "crude," which yielded a list of 16 incidents. Only one incident occurred during the period of our analysis (i.e., 2005 or later), which was the release from an ExxonMobil pipeline into the Yellowstone River, incident 20110262.

37 Available at https://darrp.noaa.gov/explore-projects.

38 We identified a small number of state-level NRDA collections, but only California's contributed documents useful for augmenting the existing database. That site is available at https://www.wildlife.ca.gov/OSPR/NRDA.

The Montana Department of Justice (DOJ) prepared a NRDA (Montana DOJ, 2013) for the 2011 ExxonMobil pipeline release of crude oil into the Yellowstone River (PHMSA incident 20110262). According to the report, the spill affected the Yellowstone River and its floodplain for approximately 85 miles downstream. However, the report does not include monetized estimates of damages, and we did not incorporate additional costs for this incident.

In some cases, the responsible party will contribute to or undertake NRDA and restoration planning. A report by Marathon Pipeline LLC (2011) describes the natural resource damages, response actions to date, and restorative and compensatory planning undertaken by Marathon as a result of a 2008 crude oil spill in Wayne County, Illinois (PHMSA incident 20080272). The 5,000-barrel spill of crude oil from a subsurface pipeline affected the floodplain of Elm Creek in Illinois, injuring a freshwater wetland habitat. The settlement included financing and implementation of restoration projects (7.1 acres of palustrine forested wetlands impacted by the spill; 14.2 acres of agricultural fields; two bat houses for endangered Indiana bats; and ten nesting boxes for migratory wood ducks), as well as a $90,629 payment to the Department of Interior for past assessment costs and future oversight and monitoring. However, the document does not include an estimate of the costs associated with past or planned restoration work, and as such we were unable to incorporate additional restoration costs for this incident (and for this incident, no remediation costs were reported to PHMSA). Because of the omissions, the total costs associated with this incident are likely to be underestimated.

On June 12, 2010, a Chevron pipeline in Salt Lake City released approximately 34,000 gallons of crude oil into the Red Butte Creek (PHMSA incident 20100146). After remediation for the incident was completed in August 2011, the Utah Department of Environmental Quality (UDEQ) conducted a screening-level ecological risk assessment (ERA) to determine the potential for residual adverse ecological impacts, and whether additional risk management actions are justified (UDEQ, 2012a). This study differs in purpose from the other NRDAs identified in this section, as it was not done for the purpose of identifying appropriate natural resource remediation projects under the OPA, but instead was undertaken after environmental remediation efforts were concluded to assess long-term ecological impacts from the release. The study entailed the collection and analysis of creek bank soil, sediments, water, and benthic macroinvertebrate samples from both the affected creek and other nearby urban creeks (which were used to establish local background conditions). The results of the study did not indicate the need for additional remediation or risk mitigation, and as such, we did not add any further monetary natural resource costs for this incident beyond those reported to PHMSA.[39]

For some more recent incidents in the database, agencies may not yet have had time to complete the full NRDA process, which can take many years.[40] For the spill that occurred on May 19, 2015 in Refugio Beach, California (PHMSA Incident 20150224), a public meeting to discuss the NRDA process occurred in January of 2016 (DAARP, 2016), and a damage assessment is underway. A November update from the California Department of Fish and Wildlife (California DFW, 2015) reported that 202 birds and 99 mammals were killed, while another 65 and 63, respectively, were oiled and captured by responders

---

[39] UDEQ also conducted a human health risk assessment after remediation for the incident (UDEQ, 2012b) and did not find elevated health hazards in the affected area.

[40] For example, the natural resource damage assessment and restoration plan from a 1994 spill in the Santa Clara River in California was not completed until 2002; see Santa Clara River Trustee Council (2002).

for rehabilitation. However the NRDA has not yet been released so a full accounting of the damages is not yet available.

3.3.7 Company Financial Reports

In addition to federal agency datasets and reports, company financial reports and other sources provide supplementary information about the costs associated with crude oil releases from trains and pipelines. For example, Enbridge's 2014 annual financial report (Enbridge, 2015) includes a discussion of the costs associated with a 2010 release of crude oil from a pipeline in Romeoville, Illinois (PHMSA incident 20100221). The report (p. 90) states that the total estimated costs for the release are approximately $51 million, excluding fines and penalties, for emergency response, environmental remediation, and cleanup activities. However, Enbridge's incident report to PHMSA includes $70,000 in commodity lost, $2.81 million in operator property damages, $12.15 million for emergency response costs, $16.54 million in environmental remediation, and $3.59 million in legal costs, for a total of $35.16 million in operator costs. As such, there are approximately $15.84 million of additional operator costs that are not accounted for in the PHMSA reported operator costs.[41] We added this amount as an "additional operator expense" for this incident.

Similarly, the incident report filed by Enbridge for the Marshall, Michigan incident (PHMSA incident 20100181) includes $840.5 million in reported costs.[42] However, the company's 2014 annual financial report (Enbridge, 2015) reports that the total expected costs for the incident are $1.2 billion. As such, we added $294.5 million in additional operator costs for this incident.

We also included approximately $5.0 million in additional operator costs for a release from an Enbridge pipeline in Wisconsin in 2012 (PHMSA incident 20120248). This amount is based on the difference between Enbridge's expected total costs from the incident as documented in the 2013 annual financial report ($10 million; Enbridge, 2014) and the costs reported to PHMSA in their incident report ($5 million). The $10 million in total costs are "inclusive of approximately US$2 million of lost revenues and excluding any fines and penalties."

Finally, Plains All American Pipeline (2016) reports that the total costs resulting from a pipeline release in 2015 (corresponding to incident 20150224) are approximately $269 million, "which estimate includes actual and projected emergency response and clean-up costs, natural resource damage assessments and certain third-party claim settlements, as well as estimates for fines, penalties, and certain legal fees... Our estimate does not include any lost revenue associated with the shutdown of Line 901 or 903 and does not include any liabilities or costs that are not reasonably estimable at this time." This cost estimate is significantly higher than the operator expenses reported to PHMSA, which include $144,000 in commodity losses, $9.9 million in property damages, $22.4 million in environmental remediation, $90.7 million in emergency response costs, and $19.8 million in other operator costs. As such, the

---

41 Note that Enbridge's incident report itemized $12.2 million in public property damages; we assume this is separate from and in addition to the "emergency response, environmental remediation, and cleanup activities" included in the $51 million in costs to Enbridge.

42 Reported costs include $23 million in public property damages, $126,118 in commodity lost, $3 million in property damage, $177 million in emergency response expenditures, and $635 million in environmental remediation.

operator costs reported in the company's financial report ($269 million) is $126.1 million higher than the operator costs reported to PHMSA ($142.9 million).

However, not all of the costs reported in the company's financial report can be considered true costs to society; some portion of the costs to the company are transfers to other parties.[43] For example, the costs include anticipated fines and penalties that may be imposed on the company from federal or other agencies. We expect fines and penalties to account for a relatively small share of the $269 million in costs. Smith and Mejia (2016) report that the company "could face fines of nearly $3 million." We therefore subtracted $3 million from the additional cost to the company, and added $123.1 million in additional operator costs to those costs reported to PHMSA.[44]

3.4 Spatial Analysis

As noted above, the PHMSA pipeline datasets include variables for the latitude and the longitude of the incident, and the FRA dataset included the same information for 11 of the rail incidents. In a prior spatial analysis of rail incidents using ESRI ArcMap, PHMSA identified the coordinates for three other rail incidents, based on locational information (e.g., road, mile markers, and location descriptions) provided by the operators in incident reports to PHMSA. POLREPs provided coordinate data for an additional rail incident. Finally, the PHMSA rail dataset includes an "Incident Route" field which provides spatial identifiers such as an address or mile marker. We used this field and Google Maps to identify coordinates for an additional 34 rail incidents.

We used these location data to map incidents in the database. The purpose of the spatial analysis was to complement the existing data by characterizing resources in the vicinity of the incidents which could contribute to explaining reported damages or costs (e.g., population residing within given distances of the incident or land cover in the vicinity of the incident).

The rail coordinate data and the pipeline dataset for incidents occurring in 2010 or later have latitude and longitude data formatted in decimal degrees. The pipeline data pre-dating 2010, however, have coordinates that are recorded in a variety of formats, such as degree decimal minute or degree minute second. Several observations also have additional non-standard nomenclature or missing coordinates. To map these incidents, we first standardized all coordinates to decimal degrees based on the original format.[45] [46] There were 22 incidents for which we were unable to standardize the coordinates, and we excluded these incidents from the spatial analysis.

---

43 In other words, if a cost to the company is a gain to another entity in an equal measure, the net impact to society is zero.

44 As noted in the financial report (Plains All American Pipeline, 2016), the $269 million cost estimate also includes "certain third-party claim settlements." These settlements may also be reasonably interpreted to be transfers rather than social costs; however, we expect that any settlements arising from the incident will be compensatory for damages inflicted on the public. While the settlements are themselves transfers, we assume that the amount transferred is equivalent to real damages to the public resulting from the incident and as such, the transfer simply reallocates that cost from the public to the operator. Since the operator did not report any public costs in its report to PHMSA, we expect these costs to be additive.

45 We employed a method that involved classifying each record by the coordinate format, editing the coordinates to remove text, spaces, and unnecessary characters, parsing each into the components needed for the

The study is primarily interested in incidents affecting onshore resources and we excluded from further analysis those incidents that were flagged as occurring offshore in the PHMSA pipeline datasets. Based on GIS mapping of the latitude and longitude data (as reported in the two PHMSA pipeline datasets), we identified an additional 19 incidents that occurred offshore (primarily in the Gulf of Mexico). We excluded these incidents from the remainder of the analysis.

We conducted a GIS analysis on the remaining incidents to obtain values for several variables described below, based on proximity or overlap.

3.4.1 Population

After mapping each incident based on latitude and longitude, we used 2010 Census Block population data[47] to estimate the population within 300 and 800 meters[48] of the incident. First, we overlaid each incident buffer with the Block polygons using a GIS Intersect tool, which splits an incident feature by the underlying Block features. Using the proportion of the Block area within the incident buffer, we area-weighted the population of the portion of the Block within the incident buffer, using the following equation:

$$Population\ in\ Buffer = \left(\frac{Buffer\ Area}{Block\ Area}\right) X\ Block\ Population$$

where:

Buffer Area	=	Area of census block within the buffer
Block Area	=	Total area of the census block
Block Population	=	Population of the census block

When portions of multiple Blocks are present within a single incident buffer, we summed the area-weighted population from each Block to find the total area-weighed population within the buffer. This calculation yields the population within the 300- and 800-meter buffer zones surrounding the incident location.

conversion, and finally applying a conversion formula (e.g., Degrees + (decimal minutes/60) or Degrees + (minutes/60) + (seconds/3600)).

46 Classifying coordinate format required professional judgment for some pairs. For example, it is not immediately clear if a longitude reported in the database as "102.64.909" intends to be read as degrees minutes seconds or as degree decimal minutes. In this case, it was determined that this represented degrees decimal minutes format because seconds cannot be greater than sixty. Therefore, we assumed that this represented "102° 64.909'." In other cases where this format was repeated, but the clear violation of the degree minute second format was not violated (minutes and seconds must each be less than sixty), we assumed that the example above was indicative of the intended format.

47 Obtained from https://catalog.data.gov.

48 These distances are based on Emergency Response Guidebook (ERG) recommended actions for incidents involving releases of hazardous materials from rail cars or tank trucks (PHMSA, 2012). In ERG Guide 128, which covers flammable liquids (non-polar/water-miscible) including crude oils, 300 meters (1,000 feet) is the initial downwind evacuation distance in the event of a large spill and 800 meters (1/2 mile) is the recommended evacuation distance (in all directions) in the event of a fire.

3.4.2 Land Use

For each incident with latitude and longitude data available, we summarized the land use or land cover in the immediate vicinity of the incident and calculated the percentage of each buffer zone that consists of water or wetland, developed land, or other land uses.

The National Land Cover Database[49] provides spatial information on land use, including the following categories:

- water (open, perennial ice/snow);
- developed (open space, low intensity, medium intensity, high intensity);[50]
- barren;
- forest (deciduous, evergreen, mixed);
- shrubland (dwarf, shrub/scrub);
- herbaceous (grassland/herbaceous, sedge/herbaceous, lichens, moss);
- planted/cultivated (pasture/hay, cultivated crops); and
- wetland (woody, emergent herbaceous)

Greater fractions of land cover consisting of water or wetland may make it more likely for an incident to affect aquatic resources, or estuarine, riverine, lacustrine or palustrine ecosystems. In total, we identified 81 incidents for which the area within an 800-meter radius of the release consists of a majority of water or wetland.

3.4.3 Endangered and Threatened Species

The USFWS identifies the extent of habitats considered critical for the conservation of species listed as endangered or threatened under the Endangered Species Act (ESA).We obtained a geospatial database of the extent of these critical habitats.[51] For each incident with latitude and longitude data available, we identified whether the 300-meter or 800-meter buffer area around the incident overlaps a critical habitat. We flagged incidents with any such overlap by setting the corresponding binary indicator variable in the database to "true."

---

49 See http://www.mrlc.gov/nlcd11_leg.php

50 *Developed, open space* is defined as "areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes." *Low intensity* is "areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% percent of total cover. These areas most commonly include single-family housing units." *Medium intensity* is "areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units." *High intensity* is "highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the total cover."

51 Data obtained from http://ecos.fws.gov/ecp/report/table/critical-habitat.html.

In total, there are 26 incidents (25 from pipelines) with critical habitats within 300 meters of the incident location. Forty-one incidents (39 from pipelines) have critical habitats within 800 meters.

3.5 Incident Data Cleaning and Adjustments

This section describes adjustments and quality checks that we made to the data, and discusses associated limitations and uncertainties.

3.5.1 Data Consistency

First, we adjusted the data so that all incidents report quantities and damages in consistent terms. For example, some datasets report the volume released in gallons, while others report in barrels. We converted all quantity estimates to gallons (1 barrel = 42 gallons) for comparison across incidents and data sources. We also updated monetary estimates of damages to 2015 dollars based on the national consumer price index (CPI) or Gross Domestic Product (GDP) deflator, depending on the cost category.

3.5.2 Data Consolidation

Next, we consolidated and reduced redundancy for some variables. There may be some overlap between the public emergency response and environmental remediation costs reported to PHMSA and those reimbursed to federal agencies in NPFC settlements. Since the NPFC dataset does not include details about the nature of the expenditures included (i.e., whether they were for response, remediation, or both), we did not include both the NPFC settlement and reported public response and remediation costs for incidents that had data in both categories. Instead, we consolidated the public emergency response costs and public environmental remediation costs reported in the PHMSA datasets, then created a new variable for *Public_Response_Remediation*. For each incident for which we had both (a) an NPFC settlement, and (b) reported public emergency response and/or remediation costs, we assigned the *Public_Response_Remediation* cost as the maximum of either (a) the NPFC settlement, or (b) the sum of public emergency response costs and public environmental remediation costs, as reported in NTSB reports, POLREPs, or PHMSA datasets.[52,53] There were eight incidents (five pipeline and three rail incidents) that included both (a) and (b); in three cases, we used the NPFC expenditure as the *Public_Response_Remediation* value, and in the other five we used the reported values.

Additionally, as discussed in Section 2.4, the cause of the incident may have some explanatory significance in the cost analysis. The incident reports for pipeline spills indicate one of eight different primary cause categories, in both the pre-2010 and 2010-present datasets (CAUSE field). Exhibit 3-2 shows the primary cause categories in both datasets; the categories use slightly different terms but are otherwise consistent. The pipeline incident reports also include data on sub-categories for incident causes

---

52 Choosing the maximum is appropriate, since public response and remediation costs may be understated in PHMSA reports, and NPFC settlements may not represent the entirety of federal expenditures in those categories; neither source is likely to overestimate such expenses.

53 In cases where multiple data sources reported public emergency response costs for a single incident, we prioritized data sources in the following order: NTSB reports, POLREPs, then PHMSA datasets.

(e.g., lightening under the natural hazard cause category) in the "CAUSE_DETAIL" field. For a full listing of cause sub-categories under each of the primary cause categories, see Appendix B.

Exhibit 3-2: Cause Categories In PHMSA Pipeline Datasets

2010-Present Categories	2005-2009 Categories
CORROSION FAILURE	CORROSION
EQUIPMENT FAILURE	EQUIPMENT
EXCAVATION DAMAGE	EXCAVATION DAMAGE
INCORRECT OPERATION	INCORRECT OPERATION
MATERIAL FAILURE OF PIPE OR WELD	MATERIAL AND/OR WELD FAILURES
NATURAL FORCE DAMAGE	NATURAL FORCES
OTHER INCIDENT CAUSE	OTHER
OTHER OUTSIDE FORCE DAMAGE	OTHER OUTSIDE FORCE DAMAGE

In the rail data, operators assigned one of 17 primary failure codes to each incident. In order to facilitate analysis, we aggregated these 17 failure codes into 4 categories of failures based on their reported primary failure causes:[54] (1) Derailment, (2) Loose, Missing, or Broken Component, (3) Deterioration or Aging, and (4) Other Human Error. For a crosswalk from the 17 failure cause descriptions to the 4 aggregated categories, see Appendix B.

3.5.3 Additional Cost Categories and Data Imputation

A small number of incidents report fatalities or injuries. Each of the three PHMSA datasets includes different reporting requirements for injuries: the pipeline data from 2010 and later solicits only information about injuries requiring hospitalization, the earlier pipeline data includes information about injuries generally (without information about the severity), and the rail data includes information about hospitalizations and non-hospitalization injuries. We used U.S. DOT guidance on the value of a statistical life (VSL) to place a value on each fatality and injury associated with incidents in the database. Based on U.S. DOT (2016), we valued each fatality at $9.6 million. For injuries requiring hospitalization, we assigned a value of $3.08 million, which represents the average of values for serious to critical injuries. For injuries that do not require hospitalization, we used the average of minor and moderate injuries, or $240,000. Finally, for unspecified injuries, we used the average of all survivable injury categories (minor, moderate, serious, severe, and critical), $1.95 million.

The database includes information about the volume of crude oil released as well as the volume recovered in each incident. A majority of incident records provide the dollar value of the lost product but this value is missing for approximately 50 incidents for which there was a net loss of product (i.e., amount recovered was less than the amount released). For these incidents, we estimated the value of the product based on net loss times the average price of crude oil for the year the release occurred (U.S. EIA, 2016b), adjusted to 2015 dollars using a GDP deflator.

---

[54] There were four incidents that were exceptions to this grouping. Derailment-related spills are characterized by substantially different impact characteristics, and two incidents were reported with derailment as the secondary cause. These two incidents were reassigned to the derailment category. Two other incidents did not include any reported failure causes.

As noted in Section 2.3.7, incidents may result in travel delays and disruptions for passengers and freight carriers, both on the roadways and railways. We do not have estimates on the number of rail passengers or highway travelers delayed by incidents, nor the duration of such delays. We also lack the detailed site- and incident-specific information needed to conduct detailed freight delay modeling for rail or highway disruptions. However, the rail dataset includes a field to indicate whether a major artery (including roads and transportation facilities) was closed as a result of the incident, and if so, the duration of the closure. 14 rail incidents in the database reported more than zero hours closed, with 12 of them being derailments (accounting for 63% of all derailments in the database). The closures lasted between 4 and 236 hours. We estimated the economic impacts associated with these closures based on a per-hour estimate of $2,000 in costs to passengers and freight carriers, based on data from Hallenbeck, et al. (2014).

We note that adjustments to impute values for fatalities, injuries, product loss, and travel delays assume that these values are not already included in other cost categories reported by the operators or other sources. If these costs are already counted in other categories (e.g., as part of the costs obtained from company financial reports and classified as other operator costs), then adding them to the total costs in the database results in double counting. However, we generally expect these imputed costs to be additive.

3.5.4 Data Quality Checks

We conducted several checks to verify the reasonableness of the data and identify outliers. For example, the PHMSA rail data presents a separate observation (record) for each rail car involved in the incident, with the quantity released reported separately for each observation but damages duplicated across the observations. In compiling the analytic incident database, we summed quantities across all observations for a given incident to calculate the total quantity released (but did not sum damages). One incident report (E-2013120116), however, appeared to report the same total quantity released for each of the 26 observations for individual railcars. We manually changed the quantity for that incident to reflect the actual total amount released rather than the calculated sum.[55]

Additionally, for those incidents with latitude and longitude data available, we checked the reasonableness of the coordinates based on the state location reported to PHMSA compared with the state location of the coordinates. There were 8 incidents where the reported state was inconsistent with the coordinate location. For these 8 cases, we looked at the relationship between the coordinates and the rest of the reported information for the incident; if the coordinate point was a great distance from the reported state, or in a non-adjacent state, we assumed that the coordinates were wrong and excluded the coordinate data from the spatial analysis described in Section 3.4. In this way, we eliminated the coordinate data for 3 pipeline incidents.

55 Specifically, this incident was a November 2013 train derailment in Aliceville, Alabama that released a total of approximately 455,520 gallons of crude oil.

4 Database and Crude Oil Incidents Description

This section provides key information about the analytic database and the information it contains about pipeline and rail incidents involving crude oil. These data are used to develop cost models in Section 5.

4.1 Database Description

As described in Section 3, the analytic database is comprised of information compiled from multiple datasets and incident-specific reports, cleaned and consolidated to variables that summarize incident characteristics and costs. Because this study seeks to assess the costs associated with crude oil released from pipelines and railcars, we exclude from the remainder of the analysis 22 onshore incidents that reported zero gallons released. As such, the final database includes 2,487 incidents.

Exhibit 4-1 summarizes the variables included in the database, the sources of the data for each variable, and the number of observations from each data source. Most of the basic information about incidents (i.e., date, location, and transportation mode) was obtained directly from the PHMSA datasets described in Sections 3.1 and 3.2. We supplemented data for some variables with other sources, as indicated in the exhibit. In some cases, multiple datasets may provide information for a single variable in the database. For these variables, we assigned decision rules for prioritizing sources based on our understanding of the reliability of different datasets. These decision rules include:

- **Quantity released** is based on NTSB reports where available, followed by PHMSA datasets;
- **Latitude/longitude** are based first on FRA data when available, followed by PHMSA datasets and then POLREPs, then by the coordinates derived in PHMSA's previous spatial analysis, finally by coordinates derived from additional incident details provided in PHMSA reports[56];
- **Surface water impact** indicator is based on information in POLREPs, followed by indicators in PHMSA datasets;
- **Soil contamination** indicator is based on information in POLREPs, followed by indicators in PHMSA datasets;
- **Public response and remediation costs** are based on a combination of NPFC settlements or NTSB reports followed by information in POLREPs then by PHMSA datasets; and
- **Additional operator costs** are based on information provided in company financial reports and reflect costs not reported elsewhere.

Note that the counts of observations shown in the exhibit reflect the number of observations actually present in the final analytic database for a particular variable – additional data may be available from other sources but were not needed to construct the analytic database because they were duplicative and we already had information deemed of higher quality. For example, the NRC dataset includes the

56 Specifically, the "Incident_Route" field in the rail database provided an address or other identifiable location for some incidents, from which we derived coordinates using Google Maps.

quantity released for 554 incidents and POLREPs provide this information for 17 incidents. However, we prioritized NTSB reports followed by PHMSA datasets as the best sources of data on quantity released. Together, these two sources provided coverage for all observations and the quantity data provided in the NRC and POLREP data sources were therefore not needed or used in the final database.

Exhibit 4-1: Data Dictionary and Sources[1]

Field	Description	Data Source	Number of Observations		
			Pipeline (2010-2016)	Pipeline (2005-2009)	Rail
PHMSA_ID	All incidents based on report number in PHMSA datasets	PHMSA datasets	1,256	779	452
Onshore_Offshore	Indicator for whether the incident occurred onshore or offshore	PHMSA datasets	1,256	779	452
		Spatial analysis	1,256	706	49
Transportation_Mode	Rail or pipe	PHMSA datasets	1,256	779	452
Incident_Year	The year in which the incident occurred	PHMSA datasets	1,256	779	452
Owner	The owner of the pipeline or train involved in the release	PHMSA datasets	1,256	779	452
Incident_Date	The date of the incident	PHMSA datasets	1,256	779	452
State	The state where the incident occurred	PHMSA datasets	1,256	779	452
County	The county where the incident occurred	PHMSA datasets	1,256	779	452
Zip	The zip code where the incident occurred	PHMSA datasets	1,256	779	452
Underground_Flag	Indicator for whether the incident originated underground	PHMSA datasets	1,256	779	452
Cause[2]	The cause of the incident	PHMSA datasets	1,256	779	452
Pipe_Pressure	The pressure of the pipeline at the time and location of the incident	PHMSA datasets	1,256	779	0
Pipe_Diameter	The diameter of the pipe at the location of the incident	PHMSA datasets	1,256	779	0
Train_Speed	The speed of the train at the time of the release	PHMSA datasets	0	0	452
People_Evacuated[3]	The number of people evacuated as a result of the incident	PHMSA datasets	1,256	779	452
Fatalities	The number of fatalities resulting from the incident	PHMSA datasets	1,256	779	452
Hospitalizations	The number of injuries requiring hospitalization as a result of the incident	PHMSA datasets	1,256	0	452

Exhibit 4-1: Data Dictionary and Sources[1]

Field	Description	Data Source	Number of Observations		
			Pipeline (2010-2016)	Pipeline (2005-2009)	Rail
NonHosp_Injuries	The number of injuries not requiring hospitalizations as a result of the incident	PHMSA datasets	0	0	452
Unspec_Injuries	The number of injuries (unspecified severity) resulting from the incident	PHMSA datasets	0	779	0
Artery_Closure_Hours	The number of hours that a major artery was closed as a result of the incident	PHMSA datasets	0	0	452
Fire	Indicator for whether there was a fire associated with the incident	PHMSA datasets	1,256	779	452
HCA	Indicator for whether the incident occurred within an HCA	PHMSA datasets	1,256	779	0
Public_Property_Damage	Value of public (non-operator) property damages	PHMSA datasets	1,256	779	452
Public_Other_Cost	Other public (non-operator) costs resulting from incident	PHMSA datasets	0	779	0
Cost_Commodity_Lost	The value of the crude oil released during the incident	PHMSA datasets	1,256	779	452
Operator_Property_Damage	Operator's property damage resulting from the incident	PHMSA datasets	1,256	779	452
Operator_Env_Remed	Operator's expenditures on environmental remediation	PHMSA datasets	1,256	0	452
Operator_Other_Cost	Other costs reported by the operator	PHMSA datasets	1,256	779	452
Operator_Emerg_Resp_Cost	Operator's expenditures on emergency response as a result of the incident	PHMSA datasets	1,256	0	452
Additional_Operator_Costs	Other unreported operator expenses	Financial reports	4	0	0

Exhibit 4-1: Data Dictionary and Sources[1]

Field	Description	Data Source	Number of Observations: Pipeline (2010-2016)	Number of Observations: Pipeline (2005-2009)	Number of Observations: Rail
Public_Resp_Remed[4]	Maximum of either 1) expenditures for response and remediation, or 2) NPFC expenditure.	NTSB report: emer. resp.	1	0	0
		POLREP: emer. resp.	3	0	2
		PHMSA datasets: emer. resp.	0	779	0
		PHMSA datasets: env. remediation	0	779	0
		NPFC expenditures	14	0	5
Surface_Water_Affected	Indicator for whether surface waters were affected	POLREP	7	0	5
		PHMSA datasets	1,256	779	452
Soil_Contamination	Indicator for soil contamination	POLREP	1	0	4
		PHMSA datasets	1,256	779	0
Quantity_Released_Gallons	The quantity released in the incident, measured in gallons	NTSB reports	2	0	1
		PHMSA datasets	1,256	779	452
Quantity_Recovered_Gallons	The quantity recovered from the incident, measured in gallons	PHMSA datasets	1,256	779	0
Quantity_Net_Gallons	The difference between the quantity released and the quantity recovered, measured in gallons [Calculated]		1,256	779	0
Response_Time_Hours	Difference between the time of the incident and the time response personnel arrived onsite	PHMSA datasets	1,160	0	0
Latitude[5]	The latitude coordinate where the incident occurred	PHMSA datasets	1,255	704	0
		FRA	0	0	11
		POLREPs	0	0	1
		prior PHMSA spatial analysis	0	0	3
		derived based on Incident Route field	0	0	34
Longitude[5]	The longitude coordinate where the incident occurred	PHMSA datasets	1,255	704	0
		FRA	0	0	11
		POLREPs	0	0	1
		prior PHMSA spatial analysis	0	0	3
		derived based on Incident Route field	0	0	34
BLK_POP10_800	The population within 800 meters of the incident	Spatial analysis	1,255	704	49

Exhibit 4-1: Data Dictionary and Sources[1]

Field	Description	Data Source	Number of Observations		
			Pipeline (2010-2016)	Pipeline (2005-2009)	Rail
LU_WATWET_800	The percent of area within the 800-meter buffer from the incident that consists of water or wetland	Spatial analysis	1,255	711	46
LU_HDDEv_800	The percent of area within the 800-meter buffer from the incident that consists of high intensity development	Spatial analysis	1,255	711	46
LU_ODEv_800	The percent of area within the 800-meter buffer from the incident that consists of other developed land uses	Spatial analysis	1,255	711	46
ET_800	Indication that the 800-meter buffer overlaps with a critical habitat	Spatial analysis	1,255	704	45
Value_Fatalities[6]	Number of fatalities times the value of a statistical life [Calculated]		1,256	779	452
Value_Hospitalizations[6]	Number of hospitalizations times the value of a hospitalization [Calculated]		1,256	0	452
Value_NonHosp_Injuries[6]	Number of non-hospitalization injuries times the value of a non-hospitalization injury [Calculated]		0	0	452
Value_Unspec_Injuries[6]	Number of unspecified injuries times the value of an unspecified injury [Calculated]		0	779	0
Value_Fatalities_Injuries_2015[6]	Sum of values for all fatalities and injuries [Calculated]		1,256	779	452
Value_ProductLost_2015	Estimated value of crude oil lost (net) if not reported by operator, based on crude oil prices and updated to 2015$ using the GDP deflator [Calculated]		1,256	779	452
Value_Travel_Delays_2015[7]	Artery_closure_hours times value of travel delays per hour [Calculated]		0	0	452
Total_Costs_Monetary	Sum of all monetary cost categories (not including values of fatalities and injuries, value of product lost, or value of travel delays) [Calculated]		1,256	779	452
Total_Costs_Monetary_2015	Total_costs_monetary updated to 2015$ using the CPI [Calculated]		1,256	779	452
Total_Costs_2015	Total_costs_monetary_2015 plus Value_fatalities_injuries_2015 plus Value_product_lost_2015 plus Value_travel_delays_2015 [Calculated]		1,256	779	452
Unit_Costs_2015	Total_costs_2015 divided by Quantity_released_gallons [Calculated]		1,256	779	452
Unit_Costs_Monetary_2015	Total_costs_monetary_2015 divided by Quantity_released_gallons [Calculated]		1,256	779	452

Exhibit 4-1: Data Dictionary and Sources[1]

Field	Description	Data Source	Number of Observations		
			Pipeline (2010-2016)	Pipeline (2005-2009)	Rail
Public_Property_Damage_2015	Reported public property damages updated to 2015$ using the CPI [Calculated]		1,256	779	452
Public_Other_Cost_2015	Reported public other costs updated to 2015$ using the CPI [Calculated]		0	779	0
Cost_Commodity_Lost_2015	Reported public other costs updated to 2015$ using the CPI [Calculated]		1,256	779	452
Operator_Property_Damage_2015	Reported operator property damages updated to 2015$ using the CPI [Calculated]		1,256	779	452
Operator_Env_Remed_2015	Reported operator environmental remediation expenditures updated to 2015$ using the CPI [Calculated]		1,256	0	452
Operator_Other_Cost_2015	Reported operator other costs updated to 2015$ using the CPI [Calculated]		1,256	779	452
Operator_Emerg_Resp_Cost_2015	Reported operator emergency response expenditures updated to 2015$ using the CPI [Calculated]		1,256	0	452
Additional_Operator_Costs_2015	Additional operator costs (based on company financial reports) updated to 2015$ using the CPI [Calculated]		4	0	0
Public_Resp_Remed_2015	Reported public other costs updated to 2015$ using the CPI [Calculated]		18	779	6
PHMSA_Only_Costs	Sum of all costs reported to PHMSA [Calculated]		1,256	779	452
PHMSA_Only_Costs_2015	Sum of all costs reported to PHMSA, updated to 2015$ using the CPI [Calculated]		1,256	779	452

1. Includes 2,487 onshore incidents involving the nonzero release of crude oil. Note that the data in this exhibit reflects the final database with retained observations, and does not reflect additional sources of data that were not used.
2. The causes reported to PHMSA were consolidated into cause categories, as described in Section 3.5.
3. Additional data sources (such as NTSB reports and POLREPs) include information about the number of people evacuated; however, these data sources do not supplement the data already reported to PHMSA.
4. Note that the numbers of observations from different data sources for this variable are not additive, since reported emergency response and environmental remediation costs are summed in populating the variable. There are 803 total observations with data for this variable, including 19 based on NPFC expenditures and 784 based on the sum of public emergency response and environmental remediation expenses.
5. Number of observations for latitude and longitude include those incidents for which the coordinates were used in the spatial analysis. As described in Section 3.4, some coordinate data reported to PHMSA were not usable or were dropped due to data quality concerns.
6. Fatalities and injuries valued based on U.S. DOT (2016). Fatalities valued at $9.6 million; injuries requiring hospitalization valued at $3.08 million (average of the value of serious, severe, and critical); non-hospitalization injuries valued at $240,000 (average of the value of minor and moderate), and unspecified injuries valued at $1.95 million (average of the value of minor, moderate, serious, severe, and critical). See Section 3.5.
7. Travel delays valued at $2,000 per hour based on Hallenbeck, et al. (2014). See Section 3.5.

Exhibit 4-2 summarizes the year and transportation mode for the 2,487 onshore incidents that reported release of crude oil from pipelines or railcars between January 2005 and June 2016.

Exhibit 4-2: Number of Incidents by Year and Transportation Mode

Incident Year	Pipeline Releases	Rail Releases	Total
2005	163	2	165
2006	157	1	158
2007	157	1	158
2008	151	8	159
2009	151	1	152
2010	150	9	159
2011	138	34	172
2012	184	88	272
2013	202	118	320
2014	226	144	370
2015	251	42	293
2016[1]	105	4	109
Total	**2,035**	**452**	**2,487**

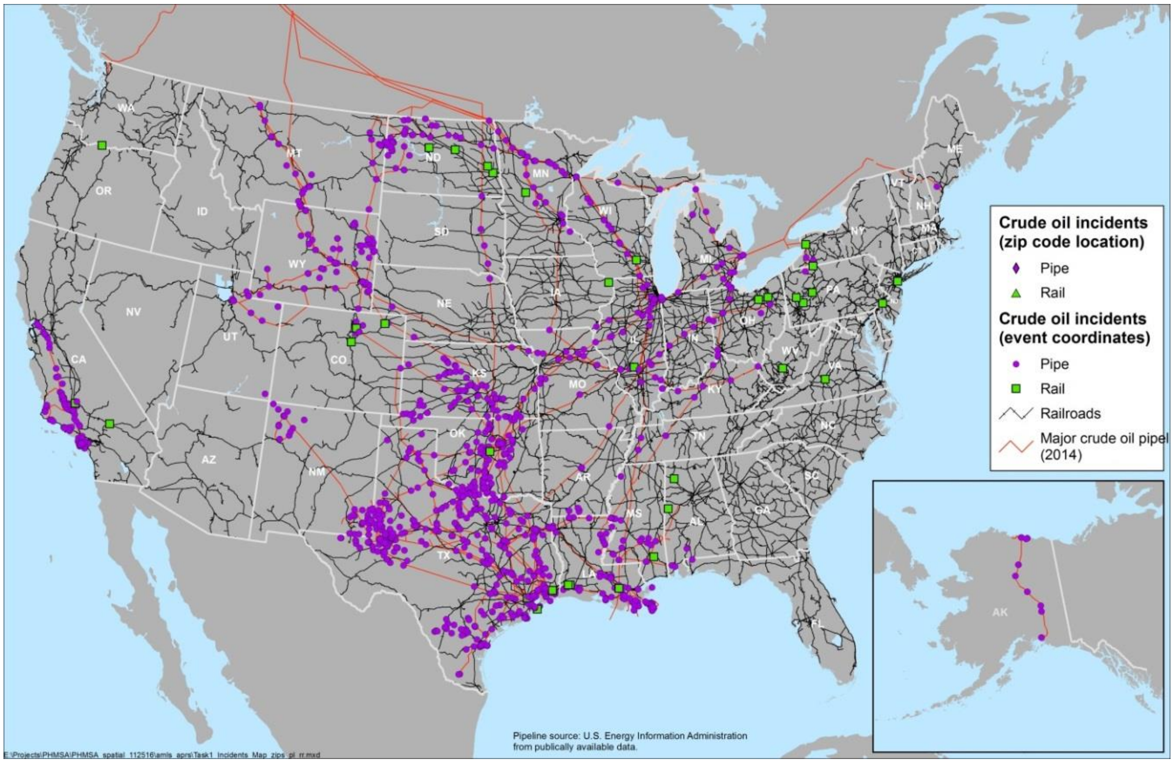
Note: Includes all onshore pipeline and rail incidents between January 2005 and June 2016 that involved the release of crude oil, as identified in PHMSA datasets (see Section 3 for a description of the methods for identifying such incidents).
1. Data for 2016 includes incidents that occurred between January 1 and June 30.

4.2 Characteristics of the Universe of Incidents

This section presents initial summary statistics that explore the characteristics of the 2,487 onshore incidents involving the release of crude oil. The map of Exhibit 4-3 shows the geographic distribution of pipeline and rail incidents with latitude and longitude information, relative to major crude oil pipelines and railroads. As discussed above, the rail dataset does not include latitude and longitude coordinates of each incident, and we were able to identify coordinates for 49 out of the 452 rail incidents (see Section 3.4).

In addition to incidents with latitude and longitude coordinates, the map also shows the approximate locations of the remaining pipeline and rail incidents in the analytic database based on the centroid of the zip code area in which they occurred (as reported to PHMSA).

Exhibit 4-3: Incident Locations



Note: Based on latitude and longitude when available; based on center of zip code otherwise.

Of the 2,487 incidents in the database with non-zero quantify released, 2,341 incidents (including 315 rail incidents and 2,026 pipeline incidents) reported some costs associated with the release. The remaining 146 incident reports did not include any costs associated with the release, even though they did report non-zero quantity released. We exclude the 146 zero-reported-cost incidents from the remainder of this analysis.

Exhibit 4-4 summarizes statistics for the key quantitative variables in the analytic database, including those associated with incident characteristics, non-monetary incident damages, and monetary incident damage categories.

Exhibit 4-4: Incident Summary Data[1]

Variable Name	# Non-Zero	# Non-Missing	Mean	St. Dev	Median	Min	Max
Incident Characteristics and Location							
blk_pop10_800	1,760	1,976	212.16	886.30	7.44	0	14,471
lu_watwet_800	1,243	1,980	0.08	0.17	0.01	0	1
lu_hddev_800	1,121	1,980	0.05	0.11	0	0	0.74
lu_odev_800	1,850	1,980	0.24	0.24	0.14	0	0.98
et800_bi	41	1,972	0.02	0.14	0	0	1
train_speed	20	315	1.95	8.11	0	0	55
quantity_released_gallons	2,341	2,341	9,349.77	72,322.13	84.00	0	2,058,000
quantity_recovered_gallons	1,676	2,026	7,356.87	63,074.03	84.00	0	2,053,800
quantity_net_gallons	1,038	2,026	2,586.22	34,341.58	2.00	0	1,301,622
hca_bi	528	2,026	0.26	0.44	0	0	1
fire_bi	37	1,947	0.02	0.14	0	0	1
surfwater_bi	137	2,341	0.06	0.23	0	0	1

Exhibit 4-4: Incident Summary Data[1]

Variable Name	# Non-Zero	# Non-Missing	Mean	St. Dev	Median	Min	Max
soilcont_bi	1,130	2,030	0.56	0.50	1	0	1
underground_bi	643	1,756	0.37	0.48	0	0	1
winter	594	2,341	0.25	0.44	0	0	1
response_time_hours	633	1,154	20.03	370.42	0	0	8,785
Non-Monetary Incident Consequences							
fatalities	2	2,341	0.00	0.06	0	0	2
hospitalizations	3	1,565	0.00	0.11	0	0	3
non-hospitalization injuries	2	315	0.01	0.13	0	0	2
unspecified injuries	1	776	0.00	0.04	0	0	1
people_evacuated	29	2,341	1.44	34.16	0	0	1,500
Monetary Incident Damages							
public_other_cost_2015	30	776	$11,353	$240,226	$0	$0	$6,603,223
public_prop_dam_15	218	2,341	$31,228	$624,542	$0	$0	$25,000,000
public_resp_remed_2015	358	800	$221,801	$2,712,050	$0	$0	$70,600,000
operator_other_cost_2015	707	2,341	$58,842	$767,512	$0	$0	$23,600,000
operator_env_remed_2015	928	1,565	$586,454	$17,500,000	$1,017	$0	$690,000,000
operator_prop_dam_15	1,225	2,341	$144,424	$3,835,781	$100	$0	$182,000,000
operator_emerg_resp_15	1,275	1,565	$412,482	$6,278,448	$2,634	$0	$192,000,000
cost_comm_lost_15	1,775	2,341	$7,883	$87,498	$83	$0	$3,206,026
addl_operator_costs_15	4	4	$116,000,000	$146,000,000	$70,100,000	$5,146,781	$320,000,000
value_fatal_injuries_15	7	2,341	$26,754	$738,083	$0	$0	$28,400,000
value_productlost_2015	49	328	$223	$1,896	$0	$0	$24,654
value_travel_delays_2015	14	315	$4,590	$32,106	$0	$0	$472,000
total_costs_2015	2,341	2,341	$1,216,021	$28,100,000	$14,017	$1.14	$1,300,000,000
unit_costs_2015	2,341	2,341	$6,103	$58,818	$124	$0.06	$1,692,531

1. This table is restricted to observations with more than zero gallons released and more than zero total costs. Mean, standard deviation, median, minimum, and maximum values are for all observations with data for the variables (i.e., including reported zeros).

4.2.1 Incident Cause

Exhibit 4-5 summarizes incident characteristics by cause. Values for pipeline releases are based on the spill cause indicated in the incident report. Values for rail releases reflect primary cause only, organized into four general categories (see Section 3.5.2). For the rail data, derailments have, by far, the largest impacts in terms of gallons spilled and total costs.

Exhibit 4-5: Median Impacts by Cause Categories for Pipeline and Rail Spills[a]

Cause Type	Count	Median Gallons Spilled	Median Total Cost	Median Unit Cost per Gallon
Pipeline Releases				
Corrosion	652	210	$30,454	$97
Equipment Failure	746	63	$8,076	$91
Excavation Damage	80	5,250	$63,285	$25
Incorrect Operation	239	126	$14,790	$71
Material Failure	130	189	$93,136	$218
Natural Forces	85	210	$55,043	$165
Other Outside Forces	36	405	$41,019	$91
Other Incident Cause	58	252	$20,830	$62
Total Pipeline	**2,026**	**126**	**$19,486**	**$95**

Exhibit 4-5: Median Impacts by Cause Categories for Pipeline and Rail Spills[a]

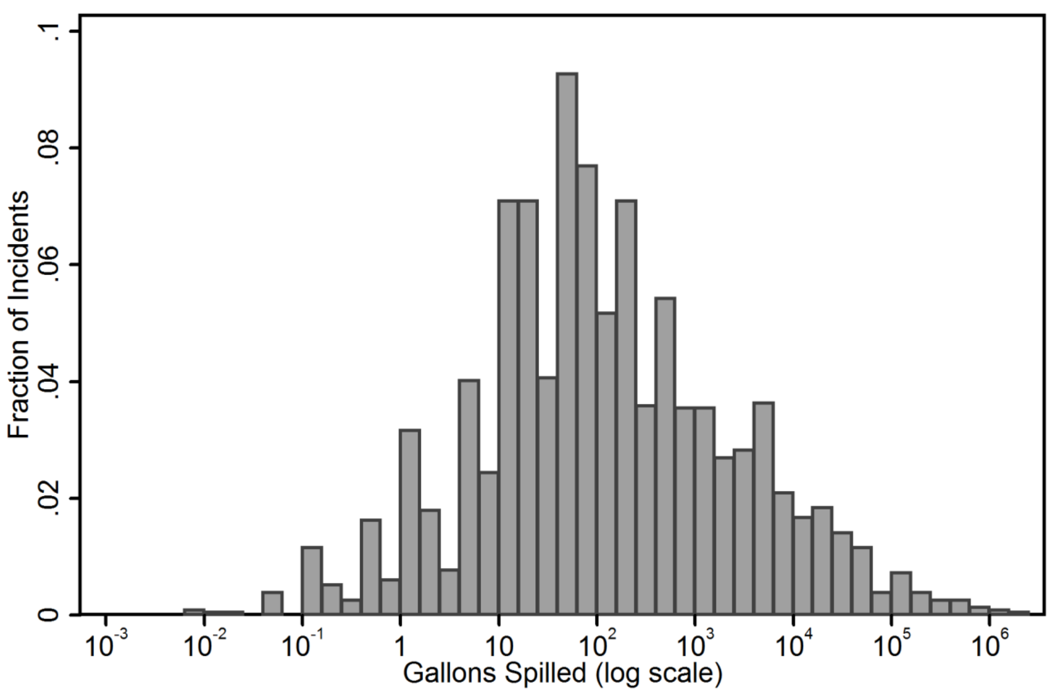
Cause Type	Count	Median Gallons Spilled	Median Total Cost	Median Unit Cost per Gallon
Rail Releases[a]				
Derailment	19	26,449	$805,284	$53
Loose, Missing, or Broken Component	174	1	$1,945	$1,503
Deterioration or Aging	21	1	$1,752	$2,002
Other Human Error	101	2	$2,544	$1,201
Total Rail	**315**	**1**	**$2,180**	**$1,296**

a. For details on the cause counts by category, see Section 3.5.2 and Appendix B.
b. Data for rail spills include incidents by cause for only primary causes, except in the case of two incidents which reported derailment as a secondary cause. These two incidents were reassigned to the derailment category.

4.2.2 Quantity Released

Exhibit 4-6 shows how the quantity released varies across the set of incidents. The histogram shows that most incidents are between 0.1 and 100,000 gallons. Spills of 10 gallons to 1,000 gallons are most common, and the median incident released 84 gallons of crude oil. The smallest incident in the dataset released only 0.0003 gallons, while the largest incident in the dataset released over 2 million gallons.

Exhibit 4-6: Distribution of Quantity Released[1]



Note: Includes all incidents with greater than zero gallons spilled and total costs.

4.2.3 Non-Monetized Damages

As described in Section 3.3.6, NRDAs include quantified impacts where possible, including counts of wildlife killed or oiled, acres of habitat affected, and even number of recreational days lost. In cases where restoration projects are identified and monetized in NRDAs, these impacts may be included in the monetized damages from incidents shown in Exhibit 4-4. In addition, the database contains information about non-monetized impacts for some incidents.

Appendix G summarizes available quantitative information about damages to natural habitats and wildlife from crude oil releases. The data provide insight into some of the impacts of crude oil spills on natural resources and ecosystem services. We did not develop separate estimates of the monetary value of the ecosystem services affected by the incident. Although damages to natural resources may be included in costs compiled from NRDAs and other sources (see Section 3.3), it is likely that incident costs in the database understate the impacts of the crude oil spills since we were able to get estimates of ecosystem service impacts and damage estimates for relatively few incidents. See Section 7 for further discussion of omitted or underreported costs.

5 Analysis of Crude Oil Incident Costs

In this section, we analyze the factors that determine the costs associated with crude oil releases from pipelines and trains. The key outcome variable for our analysis is total incident costs.

This section is organized as follows. Section 5.1 uses summary statistics and graphical techniques to explore how incident costs are related to a variety of incident characteristics. These characteristics include spill quantity (in gallons), transportation mode (pipeline versus rail), and spill location (e.g., above- or below-ground, inside or outside HCA). Next, in Section 5.2, we develop a regression model of the predictors of total cost. The dependent variable in these regressions is total costs; however, the results can also be used to estimate unit costs.[57] Finally, Section 5.3 describes how to use the regression to predict the costs of an incident, and presents an example of calculations utilizing the regression model.

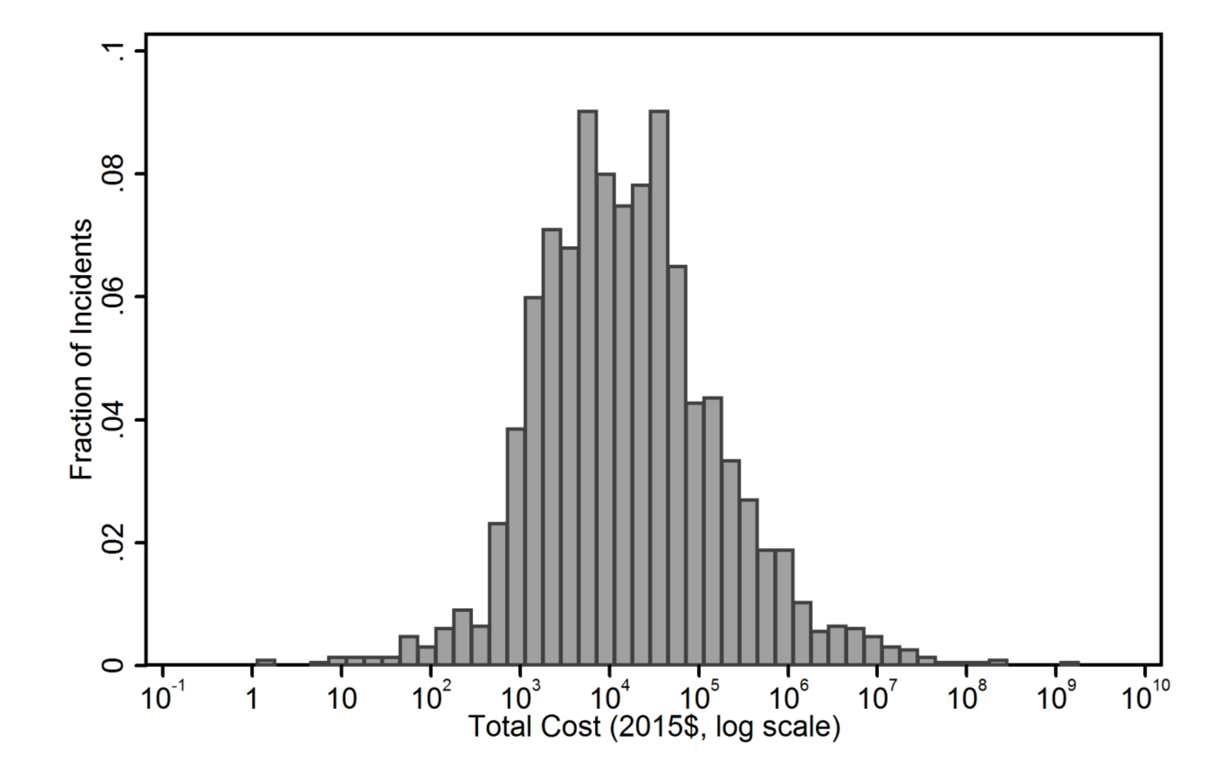
5.1 Summary Statistics and Graphical Analysis

The purpose of this section is to present costs graphically in relation to incident characteristics and provide insight into the effect of each characteristic on the incident cost. We start by presenting basic information about the distribution of costs across incidents. We then present a series of figures that explore how the distribution varies for different subsets of incidents (e.g., train and pipeline incidents).

Exhibit 5-1 presents a histogram of total costs, which shows how the total costs of spills vary across all incidents in the database. In the figure, the horizontal axis represents the total cost of spills (in a log scale), and each vertical bar represent the number of incidents that have a particular level of costs. As shown in the histogram, total costs are somewhat right skewed, with a few incidents that have very high total costs.

[57] See Appendix D.5.

Exhibit 5-1: Distribution of Total Costs



Note: Includes incidents with quantity released and total costs greater than zero.

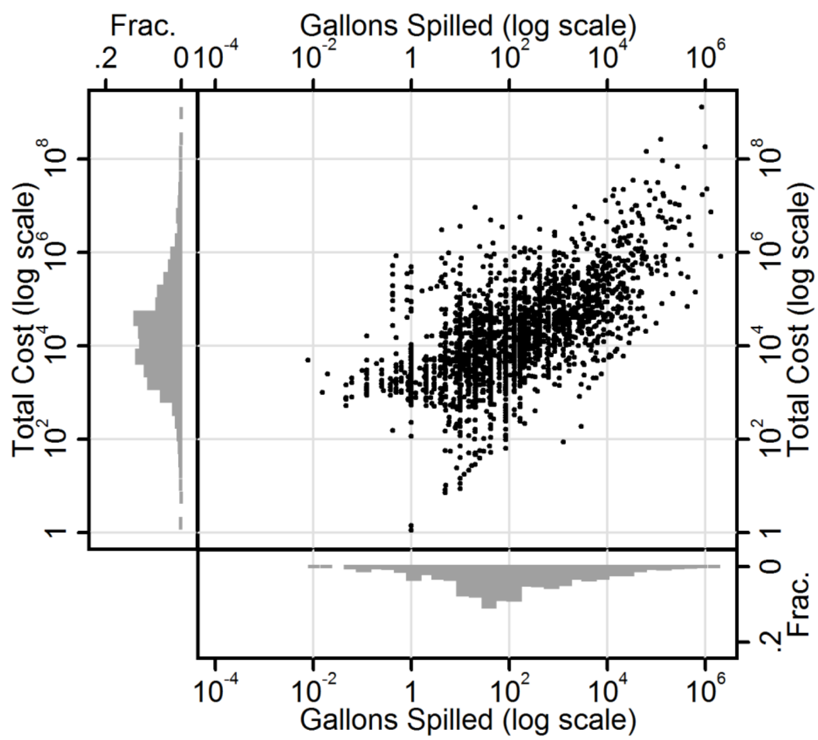
Total costs are reasonably expected to increase with the quantity of oil released, because of the potential for larger extent of the contamination, greater response costs, greater amount of waste generated and needing to be disposed of, etc. Appendix C provides a narrative summary of the top ten incidents in the database by total costs and quantity released.

Exhibit 5-2 shows a scatterplot of the relationship between quantity released and total costs. Each point in the figure represents an incident. As expected, the figure shows a strong positive correlation between total costs and quantity released. Total costs appear to increase slightly less than proportionally with quantity released, although the relationship is quite noisy. Some incidents with very low volume of oil released have relatively high costs; this includes several spills of one gallon (or less) having total costs upward of $10,000.[58] See Appendix C for information about the 10 incidents with the highest unit costs. To provide additional information on the distributions of quantity released and total costs, the figure also presents histograms of these two variables (also shown in Exhibit 4-6 and Exhibit 5-1).

---

58 For example, two 2014 incidents (X-2014050147 and X-2014050348) from Kansas City Southern Railway Company (KCS) railcars involved the release of small quantities of crude oil, discovered during routine inspections by FRA. When the leaks were detected, hazmat responders were dispatched by KCS to the scene resulting in emergency response costs of approximately $5,000. Ultimately, responders determined that both leaks resulted from bolts or nuts that were not sufficiently tightened and involved less than 0.1 gallons.

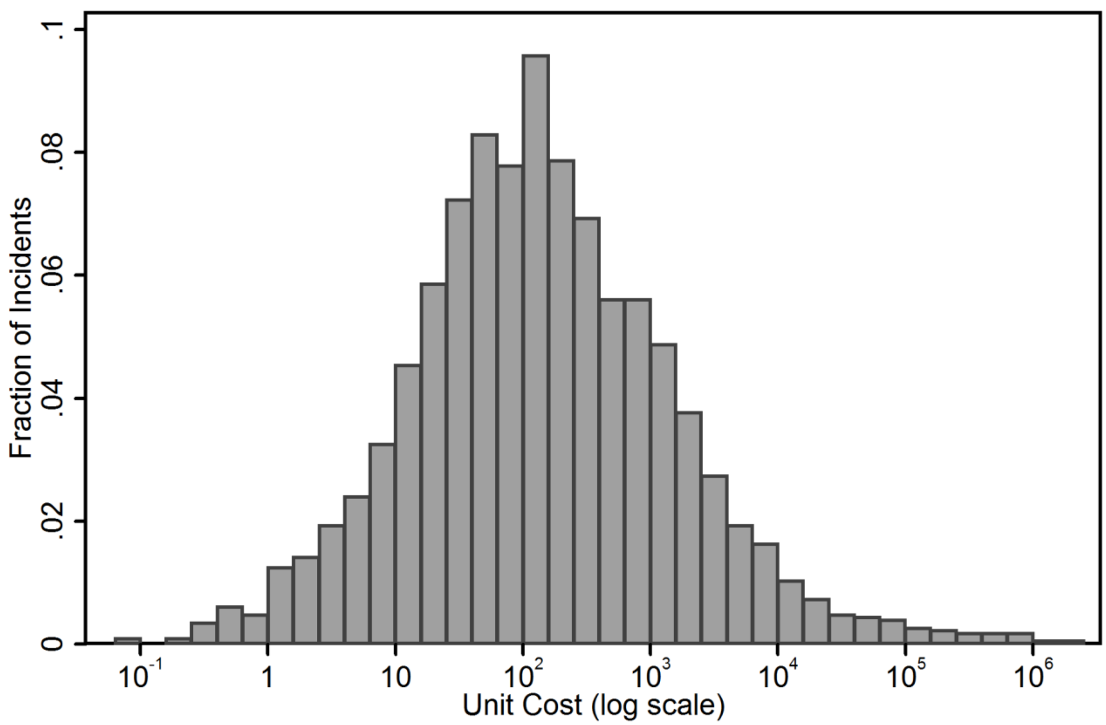
Exhibit 5-2: Relationship between Quantity Released and Total Costs



Note: Includes 2,341 incidents with more than zero gallons released and more than zero total costs.

Exhibit 5-3 presents a histogram of unit costs, with the same interpretation as Exhibit 5-1, except it shows the frequency of incidents by *unit* costs (in a log scale) rather than total costs. The distribution in the graph appears to follow a normal (bell-shaped) distribution, suggesting that unit costs are approximately log-normally distributed. This implies that in absolute terms, most unit costs are relatively small, with a few incidents that have very large values (or in more technical terms, that unit costs are heavily skewed to the right). Summary statistics confirm this implication: the median unit cost is only $124 per gallon, whereas the average unit cost is $6,103. Overall, 95 percent of incidents have unit costs between $1.46 per gallon and $4,485 per gallon.

Exhibit 5-3: Distribution of Unit Costs



Note: Includes incidents with quantity released and costs greater than zero.

One way of investigating the relationship between unit costs and quantity released is to calculate average unit costs for different ranges of gallons spilled. Exhibit 5-4 presents the results of such an analysis. The table presents the average of incident unit costs for all spills, spills of 10 or more gallons, spills of 42 or more gallons (i.e., one or more barrels), spills of 1,000 or more gallons, and very large spills of 50,000 or more gallons. In this table as for other analyses discussed in this section, each incident (and its associated unit cost) represents one observation.[59] The results are shown for all types of incidents (rail and pipeline together), as well as for rail and pipeline spills separately. The table shows that the average unit costs of incidents decline as the quantity released increases. The average spill has a unit cost of $6,103, based on all incidents, whereas for spills greater than 1,000 gallons, the average unit cost is only $102. This pattern of lower unit costs for spills involving larger volumes generally holds for both pipeline and rail incidents. However, for pipeline incidents, the largest spills of more than 50,000 gallons have higher unit costs.

---

[59] Note that an alternative approach for calculating average unit cost would be to sum the costs across all the incidents and divide these costs by the total volume of oil released across all the incidents. See Appendix A for a discussion of unit costs calculated using this alternative approach.

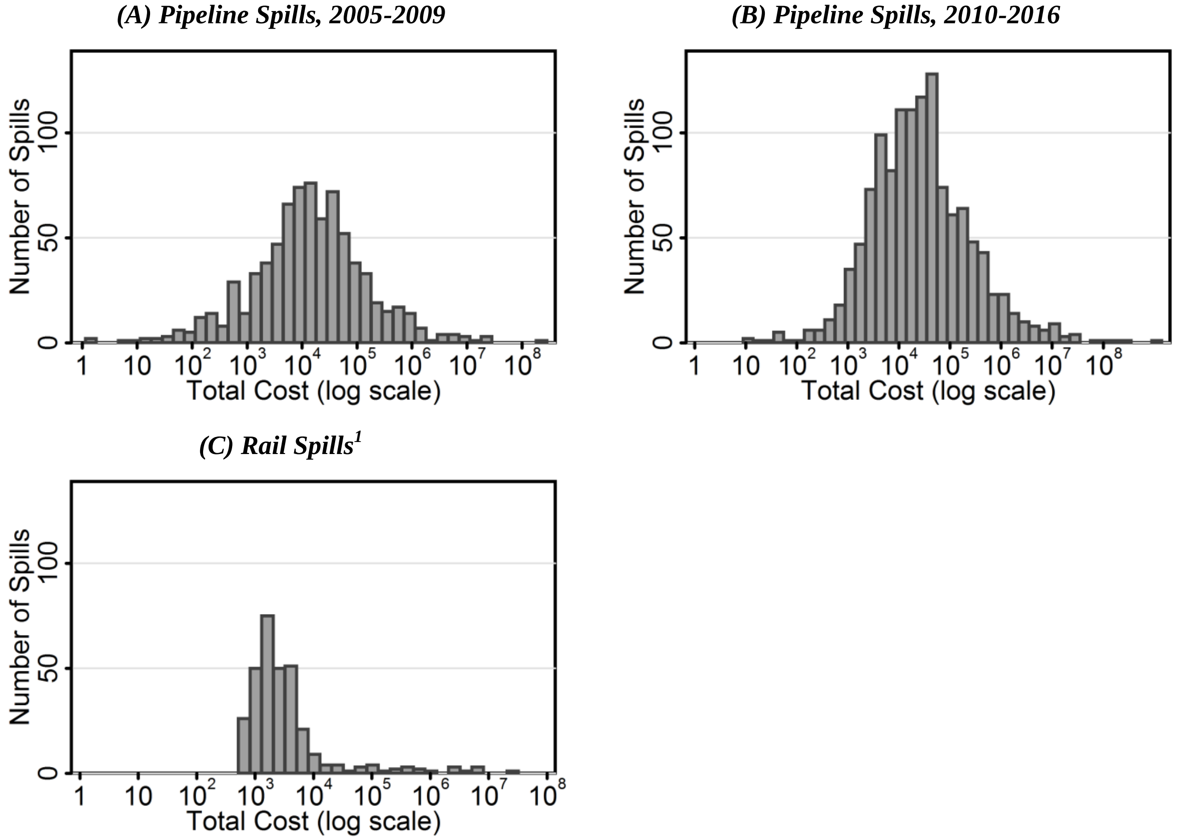
Exhibit 5-4: Average Unit Costs for Incidents Based on Quantity Released

Release Quantity	Number of Incidents	Average Unit Costs	Median Unit Costs
All Incidents			
Any release quantity	2,341	$6,103	$124
10 or more gallons	1,946	$1,109	$84
42 or more gallons	1,509	$444	$59
1,000 or more gallons	541	$102	$21
50,000 or more gallons	68	$143	$14
Pipeline Incidents			
Any release quantity	2,026	$4,551	$95
10 or more gallons	1,886	$1,137	$83
42 or more gallons	1,476	$451	$59
1,000 or more gallons	520	$104	$20
50,000 or more gallons	61	$156	$15
Rail Incidents			
Any release quantity	315	$16,084	$1,296
10 or more gallons	60	$204	$119
42 or more gallons	33	$131	$45
1,000 or more gallons	21	$63	$25
50,000 or more gallons	7	$28	$13

While it is useful to examine the distribution of unit costs across incidents in the database, the primary variable of interest in this analysis is total incident costs; as such, the remainder of this section focuses on the factors contributing to total incident costs, rather than the costs per gallon spilled.

Exhibit 5-5 presents histograms of total costs by transportation mode. Total costs are generally lowest in the 2005-2009 pipeline dataset and highest in the 2010-2016 pipeline dataset. The figure shows that the rail costs are relatively more skewed compared with the pipeline incident costs, with fewer incidents below $1,000 in total costs and more higher-cost incidents.

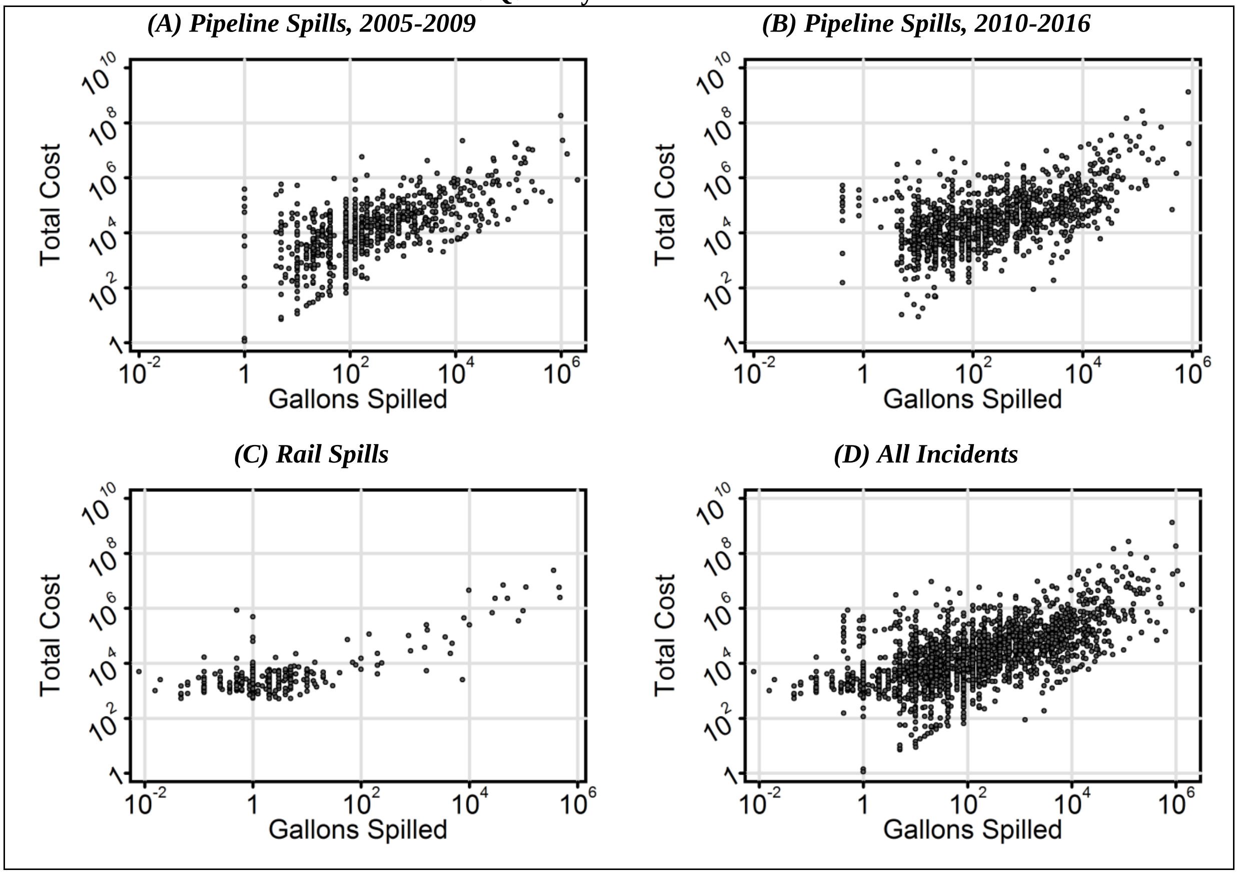
Exhibit 5-5: Distribution of Total Costs, by Transportation Mode



1. We did not separate rail incidents into two distinct time periods. Unlike the pipeline data, the format of the reported information did not change during the period. Further, there were relatively few rail incidents involving crude oil releases reported before 2010. Of the 315 incidents in the database, only 10 occurred before 2010.

Exhibit 5-6 presents a series of scatterplots of cost versus quantity of oil released, which provides additional information about the influence of quantity released and transportation mode on total costs. The four panels in the figure show the relationship separately for pipeline incidents in 2005-2009, pipeline incidents in 2010-2016, rail incidents, and all incidents combined. All four panels show total costs increasing with quantity released. Comparing panels (A), (B), and (C) shows that the slope of this relationship appears relatively consistent across the datasets. However, for smaller volume spills (below about 100 gallons) rail incidents tend to have higher total costs, while pipeline incident costs are more variable in that range. Additionally, there are more rail incidents with very small quantities released, while there are few spills from pipes that involve less than a gallon.

Exhibit 5-6: Quantity Released and Total Costs



Quantity released and transportation mode are not the only variables that may influence total incident costs. There are many other potentially relevant factors, ranging from the population living in the vicinity of the spill to the speed of a train at the time of release. Exhibit 5-7 shows how the relationship between costs and spill quantity varies across three categorical variables: (1) whether the incident occurred within or outside an HCA, (2) whether the release occurred above or below the surface, and (3) whether the release affected surface water. To control for potential differences in outcomes and dataset completeness, the exhibit shows results only for pipeline incidents from 2010 to June 2016.

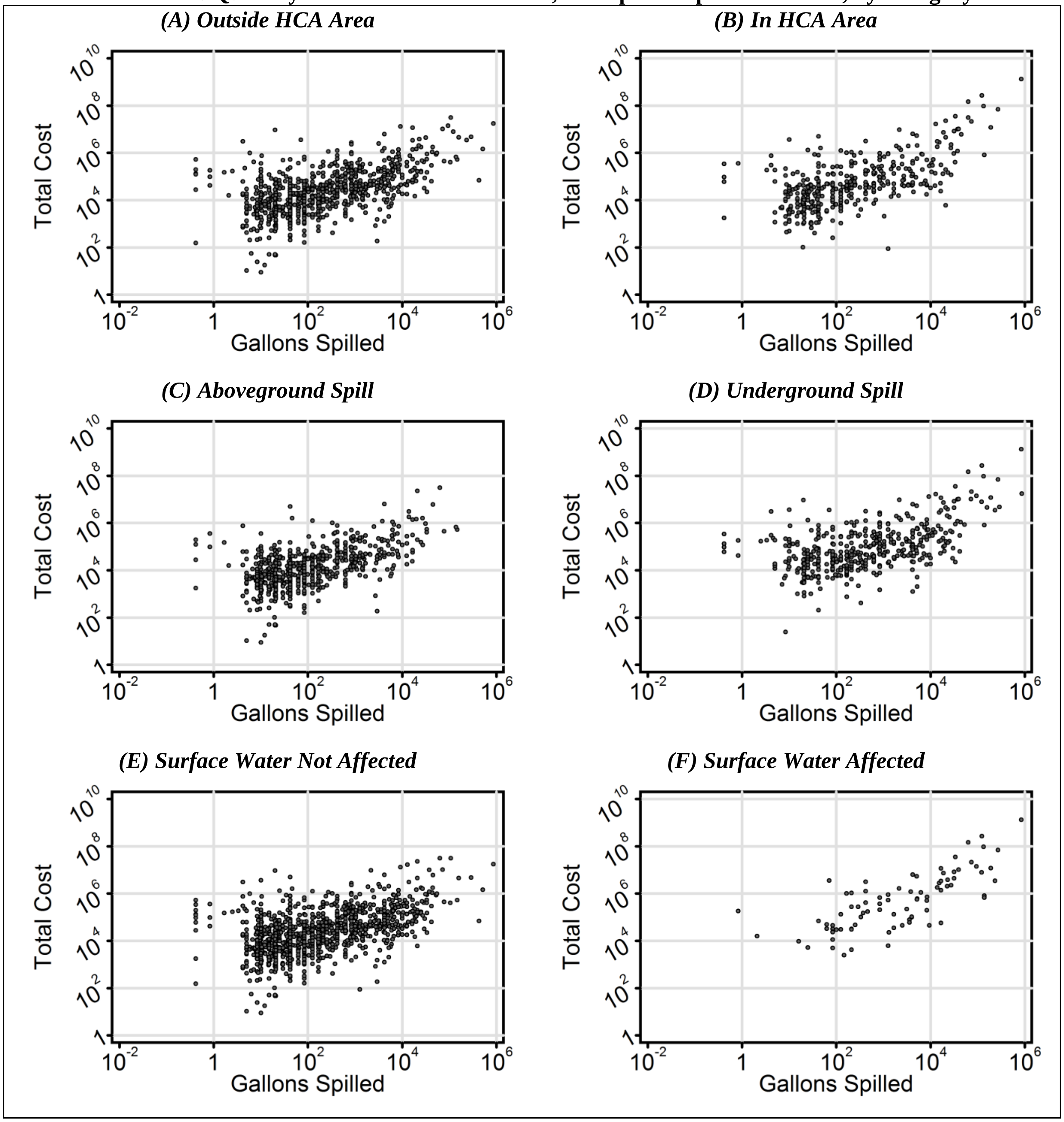
Panels (A) and (B) present the relationship between unit costs and quantity, for incidents that are located outside and inside HCAs, respectively.[60] The figure suggests that HCA and non-HCA incidents have generally similar costs, although costs do appear higher for HCA incidents when spill quantities are large. Panels (C) and (D) present a similar comparison for incidents that involve aboveground and underground releases. The figure shows that underground releases generally have higher costs than aboveground releases. Finally, Panels (E) and (F) present a comparison of releases that did and did not affect surface

[60] Identification of HCAs for hazardous liquid pipelines consider populated areas, drinking water sources, and unusually sensitive ecological resources (such as locations where critically imperiled species can be found, areas where multiple examples of federally listed threatened and endangered species are found, and areas where migratory waterbirds concentrate).

water. Incidents that affect surface water appear to have higher costs, compared to incidents that do not involve surface water contamination.

Overall, the purpose of Exhibit 5-7 is not to demonstrate that any particular variable is a robust predictor of costs. Instead, the exhibit highlights that there are many variables that could be related in some way to incident cost. To evaluate the relative influence of different incident attributes on the costs, we use a regression-based analysis, discussed in Section 5.2.

Exhibit 5-7: Quantity Released and Total Costs, for Pipeline Spills 2010-2016, by Category



5.2 Regression-based Analysis of Predictors of Incident Costs

The summary statistics and graphics in the previous section suggest that incident cost is influenced by a multitude of variables, including the transportation mode, volume of crude oil released, and incident location. In this section, we develop a regression model that formalizes those relationships. First, in Section 5.2.1, we describe the functional form for the model. In Section 5.2.2, we then present the estimated regression results. Section 5.2.3 discusses alternative model specifications.

5.2.1 Regression Model

The validity of using historical cost estimates to predict the costs of future incidents depends on the ability to account for presence of systematic, identifiable variation in the underlying data. If costs are shown to vary systematically according to attributes distinguishing previous and future incidents, the justification for using a single per unit cost value in predicting future costs of incidents becomes more tenuous. Nonetheless, average incident costs are often estimated without assessment of whether underlying data display systematic variation associated with relevant spill attributes. Regression analyses provide a statistical means to assess systematic variation in economic values. Following best practices in value transfer, we develop a regression-based model that allows us to identify systematic influences of incident attributes on costs (as described in Johnston et al., 2015; Johnston et al., 2006).

To assess systematic variation in costs based on various spill attributes, we developed a meta-regression model that predicts per incident cost as a function of the incident characteristics. The model has the following general form:

$$LogCost_i = f(SpillChar_i, SiteChar_i, CostMethod_i)$$

In this equation:

- $LogCost_i$ is the log of total cost of incident i.
- $SpillChar_i$ is a vector that describes the characteristics of the spill for incident i. This includes variables such as the quantity released, the spill source (pipeline vs. rail), the spill mechanism (derailment vs. seepage), and other spill characteristics.
- $SiteChar_i$ is a vector that describes the characteristics of the spill site for incident i. This includes variables such as the time of year, the population within the block group or within 800 m of where the spill occurred, proximity to water, whether the spill occurred in an HCA, etc.
- $CostMethod_i$ is a vector that describes the methodology used to estimate spill costs for incident i. For example, this might include whether or not particular categories of costs (e.g., commodity lost, operator damages) were accounted for in the cost estimate.

Our regression uses an ordinary least squares (OLS) regression approach. This means that the function $f()$ has a linear functional form. Using an OLS approach is standard in the literature and provides considerable flexibility in model specifications. For example, based on the relationship apparent in Exhibit 5-1, the logarithm of the cost appears to be the most appropriate dependent variable for the model. This approach can also accommodate the use of interactions between different explanatory variables—e.g., an interaction between the quantity of oil released and the transportation mode or categories of costs with non-zero values.

5.2.2 Regression Results

Exhibit 5-8 shows results for two regression models: one model specific to rail incidents and one specific to pipeline incidents. For the pipeline model, we based the regression on the pipeline incidents since 2010 only, given better data availability for those incidents.[61] Each column in the table represents a different

[61] See Appendix D.1 for the results of a model that uses the full pipeline dataset, including the 2005-2009 incidents. Since the data provided for the older pipeline incidents is systematically different from the data

regression specification, and each row in the body of the table represents the coefficient and standard error for the corresponding variable on the left-hand side.

Regression results reveal several statistically significant and intuitive patterns that influence incident costs. In general, the statistical fit of the estimated models is good; the model results suggest a considerable systematic component of total cost variation associated with transportation mode and the spill and location characteristics. The rail model has an R^2 value of 0.761 while for the pipeline model R^2 is 0.497. These R^2 values indicate that the models are predicting 76 and 50 percent of the variation in total costs, respectively.[62]

The coefficient value of the intercept terms can be interpreted, roughly, as the expected natural log of total cost, when quantity released is equal to one gallon (i.e., the natural log of quantity released is equal to zero) for incidents within the transportation mode. For example, the coefficient of 8.113 on the pipeline model intercept indicates that the natural log of total costs is expected to be 8.113 (approximately $3,338) for a spill of one gallon from a pipeline.

Each model also includes coefficients on the variables *ln_qreleased* and *ln_qreleased_sq*, which represent the natural log of the quantity released and the natural log of the quantity released squared (respectively). Together, these variables capture the effect of spill size on incident costs. In both models, the quadratic term (the natural log of quantity released squared) is positive and significant, indicating that while costs increase with quantity, they do so at a decreasing rate.

The pipeline model shows that costs are statistically higher in HCA areas, in areas with surface waters present within 800 meters, in incidents that involve fires, and in incidents that originate underground. The model also shows incident costs increasing where there are more people living within 800 meters of the incident, and when the cause of the release is categorized as “natural forces.” The rail model shows that incidents involving derailments have significantly and substantially higher costs than other types of releases from railcars, while the component and aging causes have relatively lower costs.

---

provided in the more recent pipeline datasets, that model has separate intercept and slope terms for the two pipeline datasets.

62 For OLS regressions, R^2 is calculated as [1 – (sum of squared errors/sum of squares total)]. Alternative R^2 calculations exist and are employed in situations where the standard R^2 calculation produces misleading or incorrect values. One such example is the McFadden R^2, which modifies the calculation for logit/probit (logistic) models which employ categorical values. However, in this case the standard R^2 calculation is appropriate given the continuous dependent variable of the log of total costs.

Exhibit 5-8: Regression-based Analysis of Total Costs

Variable	Regression Specification	
	Rail	Pipeline[1]
intercept	7.687*** (0.674)	8.113*** (0.264)
ln_qreleased	0.141*** (0.0279)	-0.0769 (0.0899)
ln_qreleased_sq	0.0171*** (0.00503)	0.0464*** (0.00723)
pop800	0.000195 (0.000333)	0.000166*** (0.0000460)
watwet800_bi	1.009 (0.641)	0.298** (0.0922)
spatial_missing	-0.859 (0.470)	-1.195*** (0.143)
et800_bi		0.708 (0.361)
lu_hddev_800		0.839 (0.576)
hca_bi		0.478*** (0.107)
fire_bi		2.857*** (0.459)
underground_bi		1.143*** (0.124)
rcause_derail	2.147*** (0.537)	
rcause_component	-0.277* (0.118)	
rcause_aging	-0.469** (0.153)	
pcause_corrosion		0.0337 (0.122)
pcause_natlforces		0.653* (0.279)
Observations	315	1,250
R^2	0.761	0.497
Root MSE	0.857	1.604

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*
The R^2 above is calculated based on the log-transformed costs. Calculating the R^2 for total costs shows lower R^2 of 0.561 for rail and 0.311 for pipeline.
1. Based on data from pipeline incidents between 2010 and June 2016.

In developing the regression models, we made adjustments to account for the presence of collinearity among the variables. In cases where two or more variables are highly correlated, the coefficients on those variables may be reflecting the relationship between the variables, rather than isolating the relationship between each independent variable and the dependent variables. We tested for collinearity by systematically iterating over alternative model specifications. The presented model is the result of extensive testing, which validated the inclusion of attribute main effects and selected interactions. Model specifications adhere to economic theory and intuition with model coefficients that are significant, match the expected sign, and are economically meaningful. Results suggest that collinearly does not affect coefficient estimates or interpretation. See Appendix D.2 for more information about a related sensitivity analysis, and Appendix E for a matrix of correlations across variables.

Where tests suggested high collinearity concerns, we removed some variables from the regression. For example, we removed a binary variable indicating whether surface waters were affected by the crude oil spill due to concerns about collinearity with the *watwet800_bi* variable, which indicates whether there are open waters or wetlands within 800 meters of the release. In other cases, the high number of observations available for a variable overcomes collinearity concerns. For example, in the pipeline model, we include indicators for endangered and threatened species habitat, land use, and population in the vicinity of the release in addition to a binary variable for HCA. Since the HCA definition is based in part on such site characteristics, these variables may be related to each other. However, based on our collinearity tests, we do not expect that such relationships have a significant impact on the coefficients in the model. Rather, the coefficients on the spatial variables can be interpreted as the effect of that characteristic beyond the effect of the HCA designation.

Exhibit 5-9 presents graphs that test the robustness of the regression results from Exhibit 5-8.

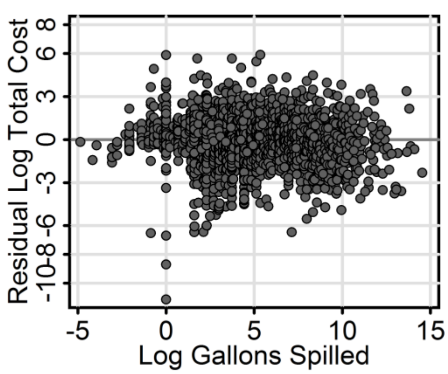
Panels (A) and (C) on the left-hand side present graphs of the model residuals versus quantity released. The purpose of these graphs is to assess whether the functional form for quantity released is appropriate, and to look for any signs of heteroskedasticity (i.e., whether the variability of the residuals is related to spill size). The panels confirm that the models fit the data well, since the residuals are generally balanced above and below zero for low and high values of gallons spilled. Both panels also show some modest signs of heteroskedasticity, particularly at the lower range of gallons spilled. To control for this issue, all regressions in this report use heteroskedasticity-robust standard errors.

As additional robustness test, Panels (B) and (D) on the right-hand side present the distribution of residuals from each model. Standard OLS regressions assume that the model residuals are normally distributed. The purpose of these graphs is to test visually whether this assumption is reasonable. In both panels, the residuals have a bell shape. The data are somewhat noisy, but there are no signs of skew or bunching at discrete values. Overall, these graphs confirm that our choice to use a log-normal OLS specification is reasonable.

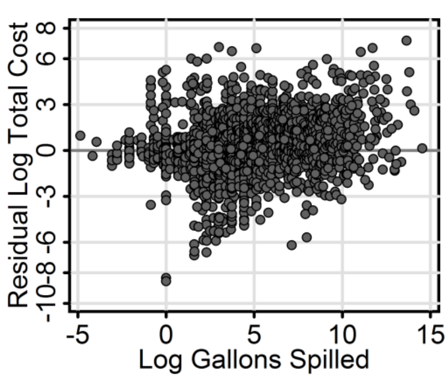
Exhibit 5-9: Regression Diagnostics

Residuals vs Quantity Released

(A) Pipeline Model

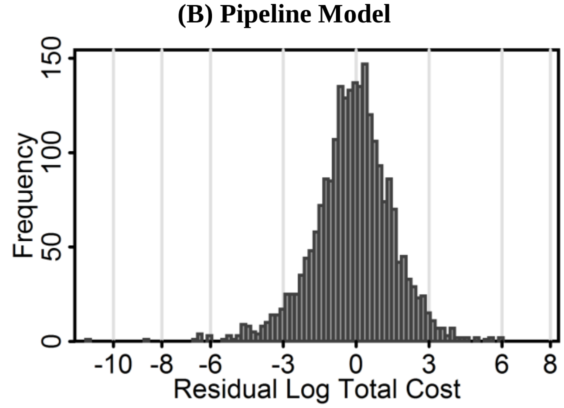


(C) Rail Model

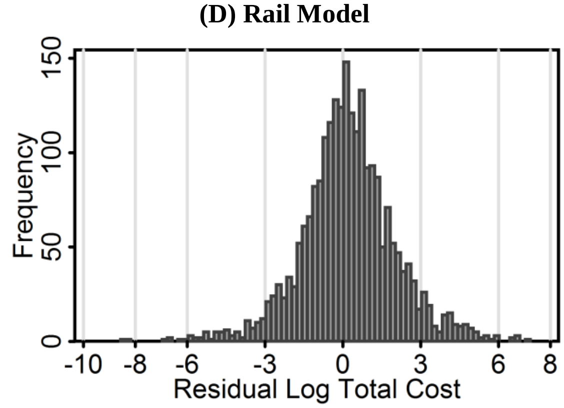


Distribution of Residuals

(B) Pipeline Model



(D) Rail Model



5.2.3 Additional Model Specifications

To further test the robustness of the primary regression specifications described in Section 5.2.2, and to address key uncertainties in the analysis (see Section 7), we also developed several additional model specifications. This section briefly characterizes each of the additional specifications, which are described more fully in other sections, as indicated.

- **Cost category models** (Section 7.1): use each individual cost category instead of total cost as the dependent variable; summing the predicted costs for each category may provide an upper-bound on the expected total costs of a release.
- **Instrumental variable analysis** (Section 7.2): addresses concerns about reverse-causality in the model.
- **Full pipeline dataset model** (Appendix D.1): includes the full set of pipeline data, including for those incidents that occurred between 2005 and 2009.
- **Kitchen sink models** (Appendix D.2): include additional variables beyond what is in the primary regressions as a test for collinearity among variables (see discussion in Section 5.2.2 as well).
- **Incident cause models** (Appendix D.3): uses incident causes as the independent variables (with variable slopes by cause) to explore the impact of each cause on total costs.
- **Combined quadratic log-log model** (Appendix D.4): combines pipeline and rail dataset into a single regression model using the same functional form as the primary regression specifications.
- **Unit cost model** (Appendix D.5): presents the combined quadratic log-log model, except using unit costs as the dependent variable rather than total costs.
- **Linear models** (Appendix D.6): presents the main regression models for rail and pipeline incidents, except with a linear form (excluding the variable for the natural log of quantity squared).
- **Binned models** (Appendix D.6): presents the main regression models for rail and pipeline incidents, except allowing for a different slope based on specific “bins” of quantity released.

5.3 Predicting Costs of Incidents

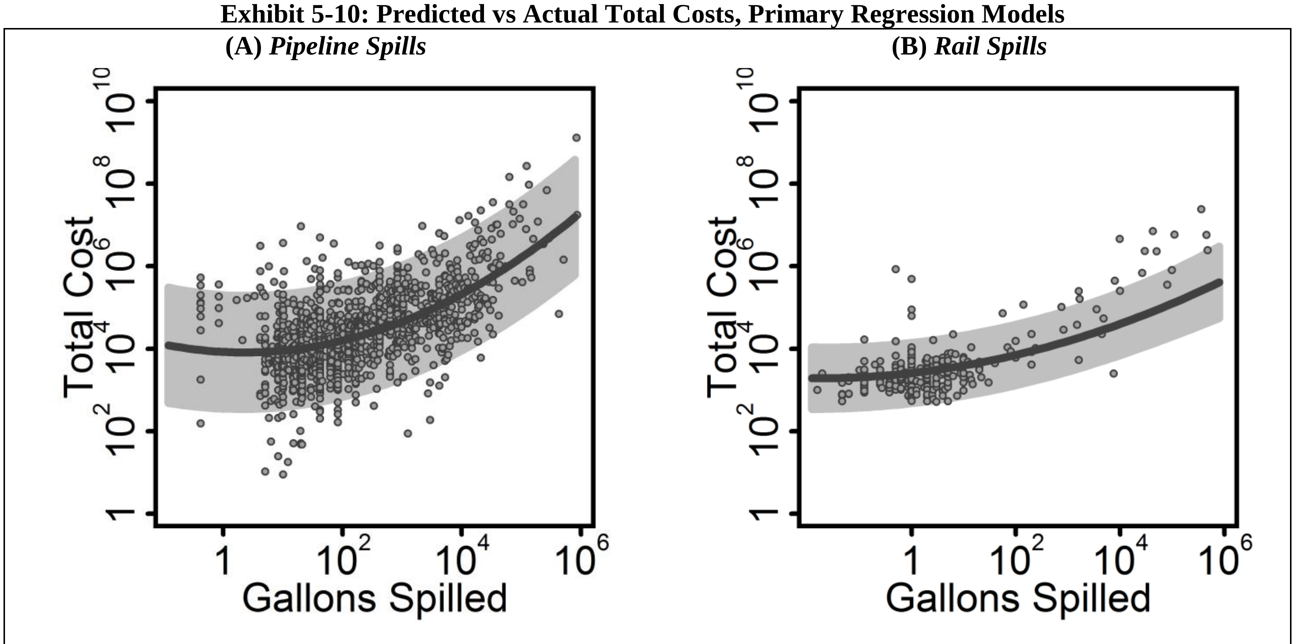
In this section, we demonstrate how well the regression models predict the costs of an incident, given information about the incident’s characteristics and location. Such information may be known from past incidents or assumed for analyses that consider subsets of policy-relevant incidents (e.g., pipelines in HCAs). For additional discussion about application of the model in a policy analysis context, see Section 6.2.

Exhibit 5-10 presents graphs of the predicted and actual costs of incidents for the regression models described in Section 5.2.2. In each panel, the dots represent actual total costs for incidents in our dataset. The dark gray line shows average predicted total costs as a function of quantity released. For the purposes of projecting the costs for these graphs, we set all other variables in the dataset equal to their average values for the specified transportation mode. Finally, the light gray shaded area shows a 95 percent confidence interval for the range of likely cost observations in the dataset.[63]

Panel (A) shows that for pipeline spills, the quantity released has little effect on costs until it exceeds approximately 10 gallons. In other words, these small spills have a fixed cost of cleanup, but little variable costs. Above 10 gallons, total costs increase substantially as quantity released increases, which is consistent with intuition. The model appears to predict the costs of pipeline incidents reasonably well.

Panel (B) shows a similar pattern for rail incidents. The predicted value line for rail incidents appears to under-predict the costs for higher-quantity incidents, on the right side of the graph. However, note that the predicted values here are for a “typical” incident in all variables—allowing variation only in the quantity released. Since the “typical” incident in the rail dataset is not a derailment, that variable is set to its average which is close to zero. If the graph were to show predicted values with the derailment variable set to one instead, the graph would show a better predictive value for those incidents (based on the positive coefficient on the derailment variable in Exhibit 5-8). As such, since the regression models allow for specification of values for all variables, they do a better job predicting the costs than these graphs suggest.

Exhibit 5-10: Predicted vs Actual Total Costs, Primary Regression Models



Note: The central line in each graph is based on the regression equation from Exhibit 5-8 assuming average values for all variables except quantity released.

To illustrate how to use the model to calculate oil spill costs for an incident with specific characteristics, Exhibit 5-11 presents examples of modeled results for 4 selected spill scenarios – two using the rail model and two using the pipeline model.

63 Note that this confidence interval represents the sampling distribution for single observations—not the sampling distribution for the mean.

- **Scenario 1**: A 100-gallon spill from a railcar, resulting from human error. This incident occurs in an area where, within a 800-meter radius, 400 people reside and there is a surface water present.
- **Scenario 2**: A 10,000-gallon spill from a railcar resulting from a derailment. There are 250 people living within an 800-meter radius but there is no surface water in the vicinity.
- **Scenario 3**: A 100-gallon spill from a pipeline inside an HCA, originating underground. The area within 800 meters of this incident has 50 residents, critical habitats for endangered or threatened species, wetlands present, and 7 percent high density development.
- **Scenario 4**: A 5,000-gallon spill from a pipeline inside an HCA that resulted in a fire, and was caused by corrosion. The area within 800 meters of this incident includes 500 residents, and is 40 percent high density development. There are no open waters or endangered and threatened species habitats in the vicinity of the release.

For each scenario, we specify the characteristics of the incident in terms of the variables in the applicable (rail or pipeline) model (e.g., spill volume, population density, whether the incident occurs in an HCA, whether the spill is accompanied by a fire). We transform the variables, as needed, depending on their model specifications. For example, binary variables are specified as 1 or 0 (for true or false), and the spill volume is converted into a natural logarithm.

The rows at the bottom of the table illustrate the steps in using the model to calculate total costs for each incident, and also present predicted costs.

The differences in costs across the scenarios are due not only to the quantity released and transportation mode, but also to the site characteristics. If we assume an incident similar to that in Scenario 4 but that involves a pipe outside of an HCA, the modeled cost would be $4.97 million, or $995 per gallon, or 1.6 times less than the value shown in the table for Scenario 4.

Exhibit 5-11: Example Application of Regression Models

Variable	Rail Incident Model			Pipeline Incident Model		
	Model Coefficient[a]	**Scenario 1 Values**	**Scenario 2 Values**	**Model Coefficient[a]**	**Scenario 3 Values**	**Scenario 4 Values**
intercept	7.687	1	1	8.113	1	1
ln_qreleased	0.141	4.61	9.21	-0.0769	4.61	8.52
ln_qreleased_sq	0.0171	21.21	84.83	0.0464	21.21	72.54
pop800	0.000195	400	250	0.000166	50	500
watwet800_bi	1.009	1	0	0.298	1	0
spatial_missing	-0.859	0	0	-1.195	0	0
et800_bi	-	-	-	0.708	1	0
lu_hddev_800	-	-	-	0.839	0.07	0.40
hca_bi	-	-	-	0.478	1	1
fire_bi	-	-	-	2.857	0	1
underground_bi	-	-	-	1.143	1	0
rcause_derail	2.147	0	1	-	-	-
rcause_component	-0.277	0	0	-	-	-
rcause_aging	-0.469	0	0	-	-	-
pcause_corrosion	-	-	-	0.0337	0	1
pcause_natlforces	-	-	-	0.653	0	0
Root MSE	0.857	-	-	1.604	-	-
A. Sum coefficient times value		9.79	12.63	-	11.44	14.61
B. (Root MSE)^2/2		0.37	0.37	-	1.29	1.29
C. Total Cost [e^{A+B}]		$25,673	$442,074	-	$335,485	$8,022,106
D. Unit Cost [C/spill volume]		$257	$44	-	$3,355	$1,604

a. See Exhibit 5-8.

6 Model Application

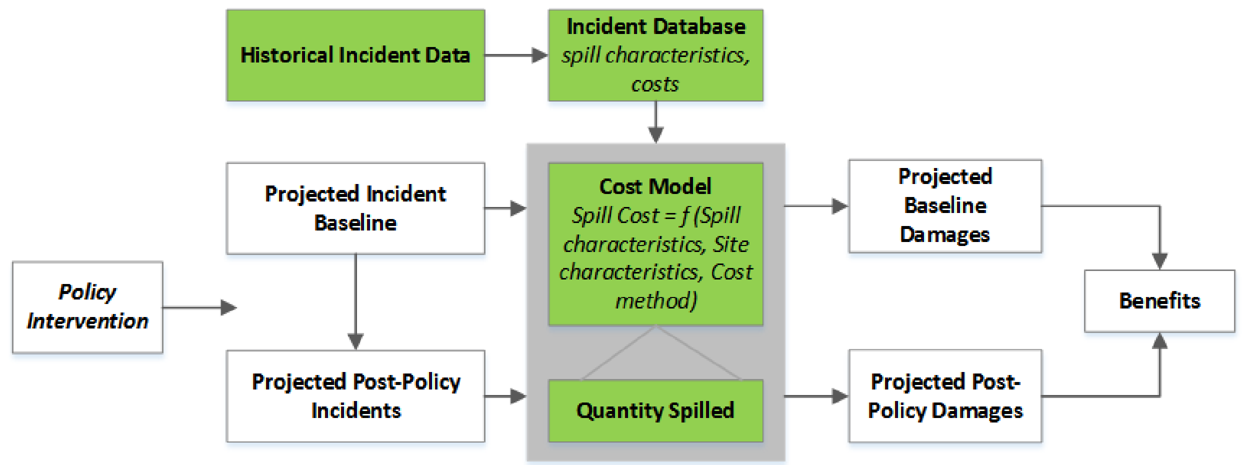
This section discusses the role of the cost models in analysis of regulations or other policy interventions. It also describes example applications of the regression models from Section 5.2 to predict expected social costs of inland crude oil spills from pipeline or rail transportation.

6.1 Generic Approach for Estimating the Social Costs of Oil Spills

Safety regulations or other policy interventions may help prevent, change the character of, or mitigate the harm caused by, crude oil spills. Regulatory analyses of these interventions are concerned with how to estimate, and monetize, the benefits of these changes.

Following the general approach illustrated in Exhibit 6-1 (which reproduces Exhibit 1-1 previously presented), PHMSA can use the cost models to calculate the expected costs of incidents that are projected to occur under future baseline and post-policy conditions. The difference in the projected damages, monetized using expected spill costs, provides a measure of the benefits.

Exhibit 6-1: Simplified Outline of Benefits Analysis Approach



Source: Abt Associates

As detailed in Section 5.2, the expected costs will vary according to the transportation mode (rail or pipeline), crude oil quantity released, incident location (e.g., HCA, densely populated area), and other incident characteristics. The difference between the sum of total baseline incident costs and total post-regulation incident costs represents the avoided social costs (i.e., the benefits) of the policy. We illustrate this general approach in Section 6.2 through two examples.

The cost models provide a way to value damages, conditional on an incident occurring. To use the models, therefore, one needs to have information on incidents that are expected to occur in the baseline, and how these incidents would change, with respect to the cost model variables, as a result of the policy intervention. Development of the baseline and post-policy incident populations is beyond the scope of this study; in general, however, the incident population can be developed based on historical data (in which case the analysis implicitly assumes that history is a reasonable predictor of future conditions), by making assumptions regarding the likelihood of various representative incidents, using outputs from a fault-tree analysis, or other approaches. As context for this separate effort, Appendix F discusses some considerations for analyzing incident frequencies.

6.2 Illustrative Application in Policy Analysis

We illustrate the use of the cost models from Section 5.2 in two example analyses of hypothetical policy interventions designed to reduce the risk of pipeline and rail crude oil spills. For each example below, we introduce the hypothetical scenario; describe the set of baseline incidents and projected changes in the incident characteristics resulting from the policy; and calculate the change in expected damages and benefits of the policy.

6.2.1 Pipeline Scenario

Hypothetical Policy Scenario

Anecdotal information and historical data suggest that there are often delays between when a pipeline rupture occurs and when the operator detects it (for example following a noticeable drop in pressure), and when the operator isolates the pipeline segment and stops the flow of oil. Reducing the length of this delay reduces the quantity of oil released and the resultant damages, all else being equal.

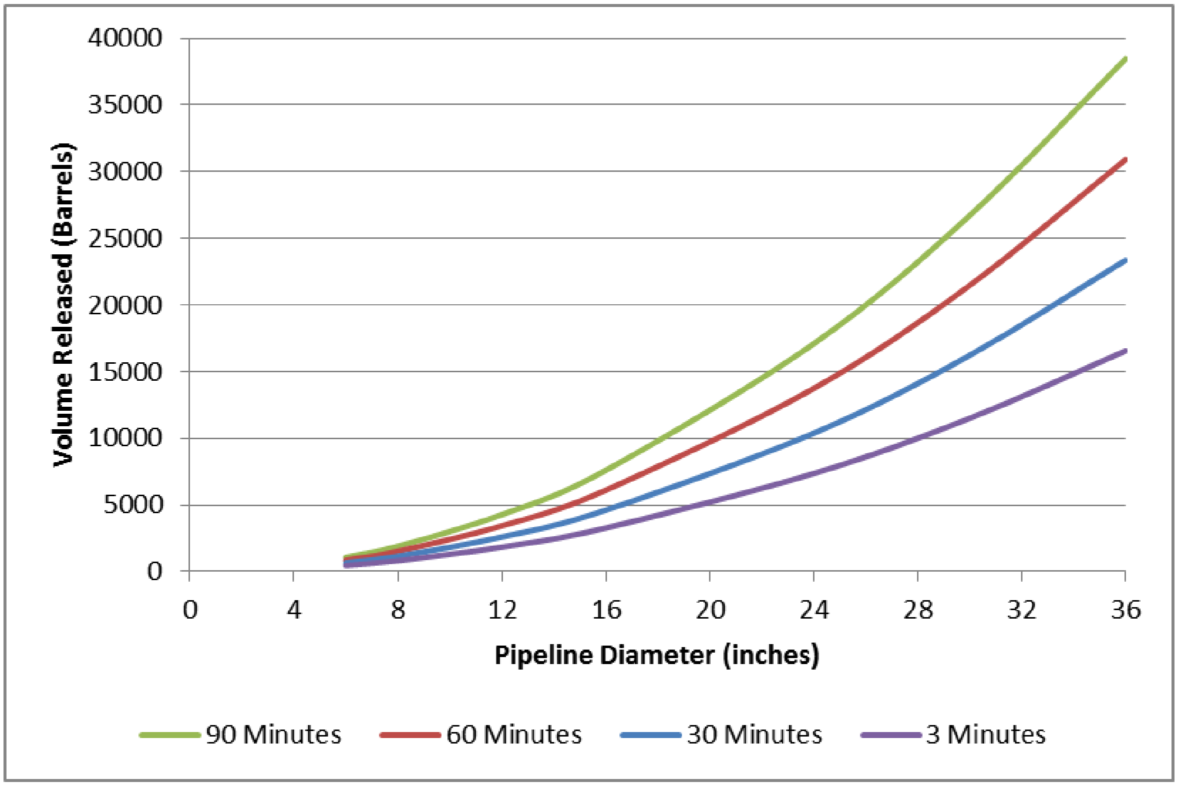
For this first illustrative application, we assumed a policy intervention that sets a maximum of 30 minutes delay between a pipeline rupture and the moment when the operator stops the flow of oil in the pipeline. We further assumed that this policy affects only pipelines larger than 6 inches in diameter and that could affect HCAs.

Baseline and Post-Policy Incidents

The first step is to obtain the baseline population of incidents and determine how the policy would affect these incidents. For the purpose of this illustrative application, we assumed that past incidents provide an appropriate baseline. The historical data described in Section 4 provides a starting point for this analysis.

To the degree that operators took more than 30 minutes to detect and stop the flow of oil, the policy would reduce the volume released in similar future incidents. Unfortunately, the PHMSA pipeline incident data do not provide information about the duration of the release. A study developed by Oak Ridge National Laboratory (ORNL) used engineering principles to estimate the volume of hazardous material released in various pipeline rupture scenarios (ORNL, 2012). ORNL provides spill volumes as a function of detection and block valve closure time, pipe diameters, velocity, operating pressure, elevation change, and segment length. To identify incidents that would be affected by the policy, we generalized ORNL scenario results into the relationship shown in Exhibit 6-2 between quantity released, pipeline diameter, and delay.

Exhibit 6-2: Estimated Discharge as a Function of Pipeline Diameter and Detection and Shut-off Delay



Source: Abt Associates, based on scenario results in ORNL (2012; Tables: A-1; A-37; A-73; A-111; A-145; A-181).Abt Associates adjusted volumes estimated by ORNL by a factor of 1.5 after comparing the estimates to actual releases for which the delay was longer and the spilled volume less than obtained by ORNL.

The volume calculated using the relationship in Exhibit 6-2 for a 30-minute delay represents the maximum volume we would expect for the rupture of a pipeline of a given diameter (V_{30}). For each incidents to which the policy would apply (defined above as incidents from pipes with diameter larger than 6 inches and inside HCA), we calculated V_{30} based on the reported pipe diameter.[64] We then assumed that incidents for which the volume reported by the operator ($V_{baseline}$) was less than V_{30} would not change as a result of the policy intervention, whereas the policy would cap at V_{30} the volume for incidents where $V_{baseline}$ exceeded V_{30}. In other words, the policy volume is set to the minimum between $V_{baseline}$ and V_{30} for a given historical incident.

We identified 13 incidents between 2005 and 2016 with $V_{baseline}$ greater than V_{30}. The corresponding volume reduction is 46,857 barrels, or approximately 2.0 million gallons.

Incident Costs and Benefits

The next step involves estimating expected costs under baseline and post-policy conditions. For each of the 2,184 incidents in the historical dataset, we applied the pipeline model in Exhibit 5-8 to estimate

[64] For incidents for which operators did not report a diameter, we assumed 6 inches.

expected costs based on the reported characteristics of each incident (HCA, fire, etc.) and reported or modeled spill volume as described above for the baseline and post-policy scenarios. Exhibit 6-3 shows the quantity released and total costs for the baseline and post-policy. We estimated the benefits of avoiding the release of 46,857 barrels of oil at $268.7 million, which is an average of $137 per gallon. The benefits are associated with changes to the volume released in 13 incidents, which is an average of $20.7 million per incident.

Exhibit 6-3: Summary of Results for Illustrative Pipeline Example

	Quantity Released (Barrels)	Incident Costs (Million 2015$)
Baseline	478,600	$2,164.7
Post-Policy	431,743	$1,896.0
Change	46,857	$268.7
Source: Abt Associates		

Note that the benefits are based on all pipeline incidents reported over the period of 2005 and through mid-2016 and assume that future incidents would be similar in terms of the types of pipelines involved, geographical distribution, probability of occurrence, response actions, and other factors.

6.2.2 Rail Scenario

Hypothetical Policy Scenario

Nineteen derailments reported between 2005 and mid-2016 resulted in the release of 1.8 million gallons of crude oil from 114 railcars.

For this second illustrative application, we assumed an intervention that helps prevent the release of oil in the event of a derailment by reducing the likelihood that the railcar would rupture, for example by strengthening railcar shells to better resist an impact. We assumed that the measures would not affect the probability that a train carrying crude oil would derail or the number of cars affected by the derailment, but would reduce by 49 percent the probability that the affected railcars would release their content, relative to the baseline.

Baseline and Post-Policy Incidents

The first step is to determine the universe of incidents in the baseline, and how the policy intervention would affect the characteristics of these incidents. As noted above, there were 19 derailments reported in 2005-2016 involving 114 cars. PHMSA data provide the volume released from each of the individual rail cars. Railcar-specific spill volumes range from approximately 1 gallon to more than 30,000 gallons.

To derive the post-policy incident set, we randomly assigned outcomes (rupture/no rupture) to the set of railcars in the baseline data set for derailment incidents, assuming that each railcar that released oil under the baseline would have a 49 percent probability of avoiding any release under the policy. We calculated the post-policy spill volume by summing the adjusted quantities across all railcars corresponding to each incident and left unchanged other incident characteristics that depend on the incident location.

Incident Costs and Benefits

The next step involves estimating expected costs under baseline and post-policy conditions. We applied the rail model from Exhibit 5-8 to estimate expected costs for the 452 rail incidents.[65] Exhibit 6-4 summarizes the result of the analysis. Of the 114 railcars that failed in the baseline, 56 are expected to remain intact with the policy intervention (as far as releasing their content), reducing the volume of oil spilled by 21,696 barrels. The total avoided costs are $19.3 million, which is an average of $1.0 million per incident or $21 per gallon.

Exhibit 6-4: Summary of Results for Illustrative Rail Example

	Derailed Railcars Releasing Crude Oil[1]	**Quantity Released (Barrels)**	**Incident Costs (Million 2015$)**
Baseline	114	42,037	$45.8
Post-Policy	58	20,341	$26.5
Change	56	21,696	$19.3

1. Based on 19 derailments between 2005 and mid-2016, out of a total of 452 rail incidents. The results reflect randomly assigned outcomes for each railcar (with a 49% probability of avoiding the release) and therefore vary across iterations of the calculations.
Source: Abt Associates

Similar to the first illustrative example, the benefits of this scenario are based on rail incidents reported over the period of 2005-2016 and predicated on future rail incidents being similar in terms of geographical distribution, probability of a derailment, railcar fleet, response actions, and other factors.

6.3 Discussion and Additional Considerations

The two examples above highlight some of the limitations of using uniform unit cost assumptions to project expected costs for different types of incidents. Whereas PHMSA (2015) used an average unit cost of $211 per gallon in prior analyses, the subset of pipeline and rail incidents affected in the two illustrative examples suggest significantly lower unit values ($137 per gallon for pipeline example, $21 per gallon for rail example). The lower values reflect the focus of the examples on fairly large incidents and are consistent with our finding in Section 5.1 of declining unit costs as spill volumes increase. Conversely, analyses of policies that preferentially affect smaller spills could have much higher unit costs.

The two examples above show how one can use the cost models to estimate the benefits of specified changes in the characteristics of crude oil incidents prompted by a regulation or other policy intervention. This is only one of many potential uses of the cost models. The models may also be used to inform the understanding of the influence of different factors on incident costs and focus attention on pipelines and rail scenarios with higher potential impacts and benefits.

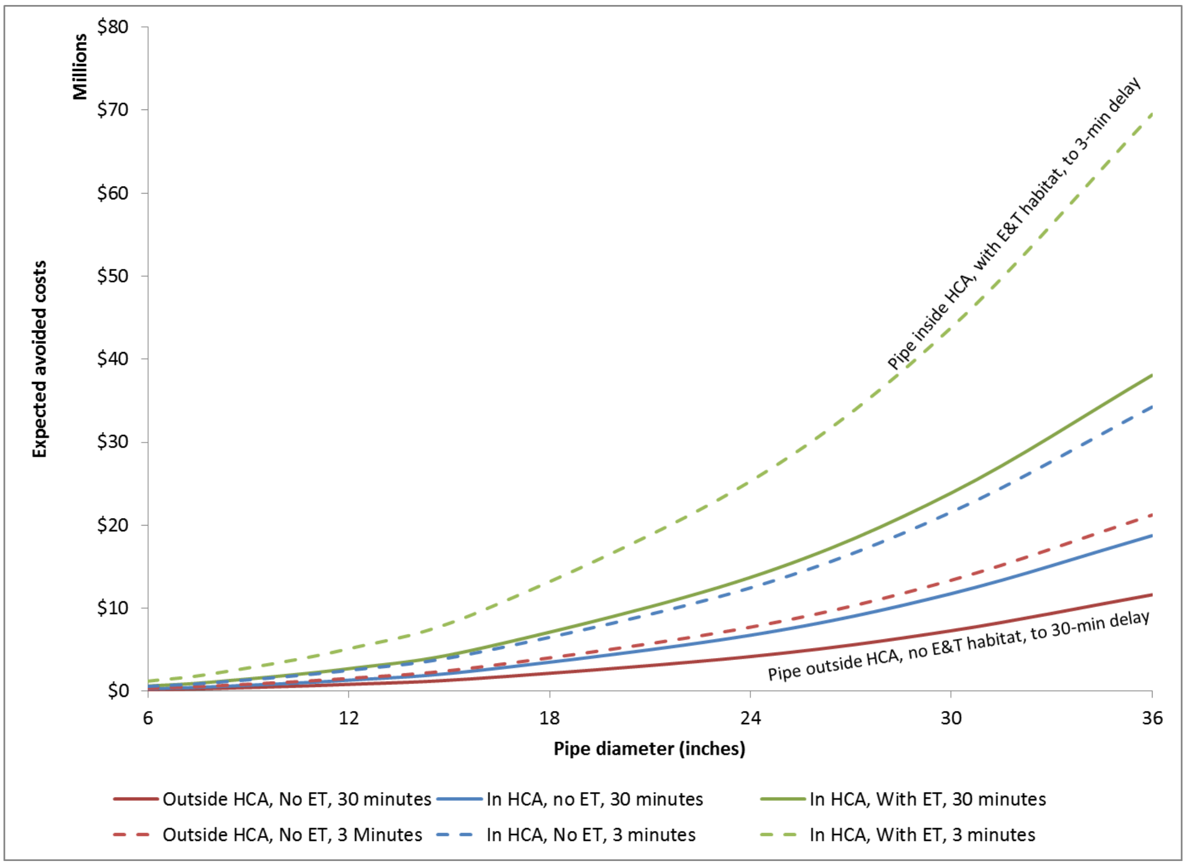
We illustrate this through a simple example that builds on the pipeline example above. Say we are interested in questions such as: If operators typically respond to pipeline ruptures within 60 minutes, what would be the benefits of shortening this period to 30 minutes, or to 3 minutes? How do these benefits compare for different types of pipelines?

65 For this analysis, we used the full set of 452 incidents in the database since we are using the historical data as the basis for modeling expected costs for both the baseline and policy scenarios.

We considered three different types of lines: (1) Pipelines of various diameters (6, 12, 18, 24, and 36 inches) outside of HCA; (2) Pipelines of various diameters in HCA; and (3) Pipelines of various diameters in HCA and in proximity to E&T habitats. We then constructed incident scenarios for the various pipeline types above by also varying the spill volumes to correspond to delays of 60, 30, and 3 minutes, based on the relationship presented earlier in Exhibit 6-2 between pipeline diameter, delay, and spill volume.

Exhibit 6-5 shows the benefits curves derived by comparing expected costs avoided when reducing the delay from 60 to 30 minutes, and from 60 to 3 minutes for the various scenarios. As expected, the benefits of faster response increase with the degree of improvement in response delays and with the pipe diameter. This trend derives from the influence of the two factors on modeled spill volume reductions. The benefits also vary depending on the type of pipeline in accordance with the effects of locational variables on spill cost. Pipelines inside HCAs yield higher benefits than those outside HCAs. Of the scenarios analyzed, the highest relative benefits are achieved for lines in HCAs and in proximity to E&T habitats.

Exhibit 6-5: Expected benefits of reducing 60-minute delay to 30 minutes (solid lines) or 3 minutes (dashed lines) for different types of pipelines



The spill volume is based on the relationships shown in Exhibit 6-2.
Source: Abt Associates.

7 Sensitivity Analyses and Uncertainties

The analysis presented in this report includes several limitations and uncertainties that may overstate or understate the social costs of crude oil releases. This section discusses some of the key limitations of the model described in Section 5.2 and describes the methods and results of sensitivity analyses that explore the impact of the limitations on the results and applicability of the model.

Section 7.1 provides a discussion of the impact of missing cost data in the analysis, and Section 7.2 explores the impact of uncertainties in the causality of quantity released and incident costs. In Section 7.3, we discuss additional sources of uncertainty in the database and analysis, and note the expected direction of known biases.

7.1 Issues Related to Missing Cost Data

One challenge in estimating a regression model of incident costs is distinguishing between incidents for which particular cost categories are zero versus simply missing. For example, if a particular data source does not quantify all categories of spill costs (e.g., public emergency response costs), and instead treats those cost categories implicitly as $0, then our regressions could underestimate actual costs.

Additionally, the availability of incident characteristic data – the explanatory variables used in a regression analysis – also varies by data source, which may introduce further bias in the model. While both pipeline datasets include an indicator for whether surface waters were affected by the crude oil release, the rail dataset has no such indicator. In some cases, as described in Section 3.3, we supplemented these fields with data from other sources. POLREPs indicated that surface waters were affected for 6 rail incidents. Since larger incidents may receive more attention, the availability of supplemental data is not uniform over the data set and may introduce bias.

Exhibit 7-1 summarizes the cost categories that are present in our three primary datasets. Total costs are comprised of operator costs (including property damage, emergency response costs, environmental remediation costs, the costs of the lost commodity (crude oil), and other reported costs), costs accruing to the public (including property damages, emergency response costs, environmental remediation costs, and other costs), and additional costs as described in Section 3.5.3. The figure shows substantial variation in the completeness of reporting across datasets.

Exhibit 7-1: Percentage of Cost Categories with Non-Zero Values, by Dataset

Variable	Rail		Pipeline 2005-2009		Pipeline 2010-2016	
	Present	% Non-Zero	Present	% Non-Zero	Present	% Non-Zero
Public property damages	✓	3%	✓	11%	✓	10%
Public environmental remediation costs	~	2%	✓	43%	~	1%
Other public costs	×	0%	✓	4%	×	0%
Costs of commodity lost	✓	31%	✓	81%	✓	84%
Operator property damages	✓	2%	✓	27%	✓	81%
Operator emergency response costs	✓	98%	×	0%	✓	77%
Operator environmental remediation costs	✓	30%	×	0%	✓	67%
Other operator costs	✓	0%	✓	73%	✓	11%
Additional operator costs	×	0%	×	0%	~	0.3%

Note: Values represent the percent of observations that have a non-zero value for each cost category, across all observations from that dataset. Excludes imputed values for fatalities, injuries, value of product lost, and value of travel delays.
✓ = *data reported*
× = *data not reported in underlying dataset*
~ = *data not reported in underlying dataset, but some supplementary data is used for a subset of incidents.*

Exhibit 7-2 shows the shares of per-gallon costs in each different cost category, based on those incidents reporting non-zero damages in the category.[66] In the figure, the per-gallon costs represent only costs from the specific cost category being considered. As indicated in the table, a small share (less than 10 percent) of incidents have positive public property damages or other public costs, whereas a majority of incidents have positive costs for commodity lost and operator emergency response.

Exhibit 7-2: Unit Cost by Cost Category

Cost Category	Percentage of Incidents, by Range of Cost per Gallon							
	$0 or missing	$0 to $1	$1 to $10	$10 to $100	$100 to $1,000	$1,000 to $10,000	$10,000 to $100,000	>$100,000
Public property damages	90.7%	2.3%	3.6%	2.1%	0.9%	0.3%	0.1%	0.0%
Public environmental remediation costs	84.7%	1.2%	3.7%	6.8%	3.0%	0.5%	0.1%	0.0%
Other public costs	98.7%	0.3%	0.5%	0.3%	0.2%	0.0%	0.0%	0.0%
Costs of commodity lost	24.2%	23.2%	51.1%	1.4%	0.0%	0.1%	0.0%	0.0%
Operator property damages	47.7%	7.5%	12.2%	17.7%	10.7%	2.9%	0.8%	0.6%
Operator emergency response costs	45.5%	2.7%	9.7%	16.4%	15.5%	8.0%	2.0%	0.2%
Operator environmental remed. costs	60.4%	1.3%	7.5%	16.6%	11.5%	2.3%	0.3%	0.0%
Other operator costs	69.8%	3.2%	5.4%	12.1%	7.4%	1.6%	0.2%	0.1%
Additional operator costs	99.8%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%

Notes: Data excludes imputed values for fatalities, injuries, value of product lost, and value of travel delays. Observations are classified into cost ranges after adjustment to 2015 dollar values.

Exhibit 7-3 presents scatterplots of costs versus quantity released, by cost category. Each panel in the figure represents a different category of costs. The figure shows that within almost all cost categories,

[66] We use per-gallon costs for this comparison to normalize across potentially large differences in the magnitude of total costs across incidents.

costs are much more likely to be reported for larger releases. Within each category, costs increase with quantity released.

Exhibit 7-3: Costs versus Quantity Released, by Cost Category

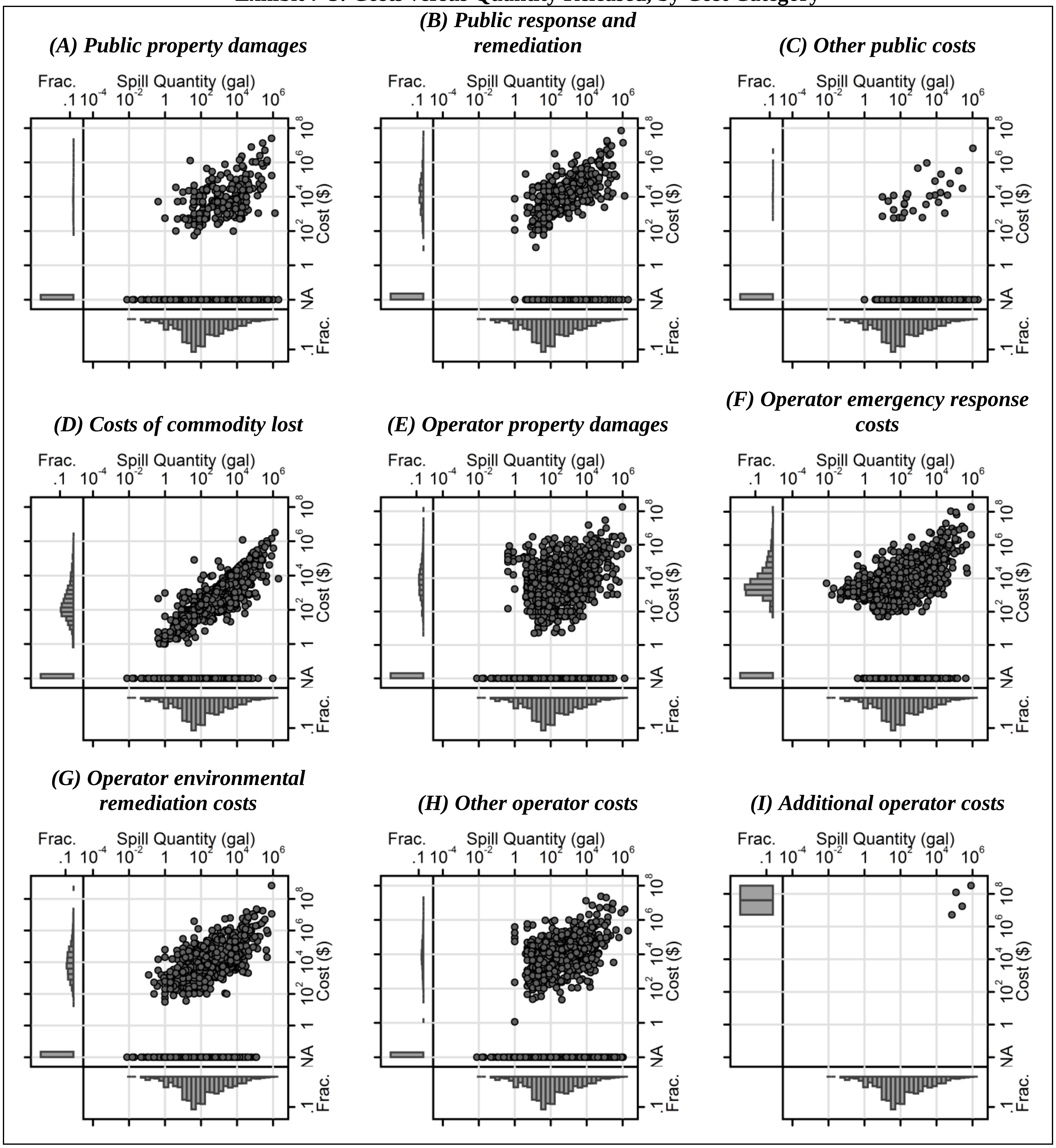
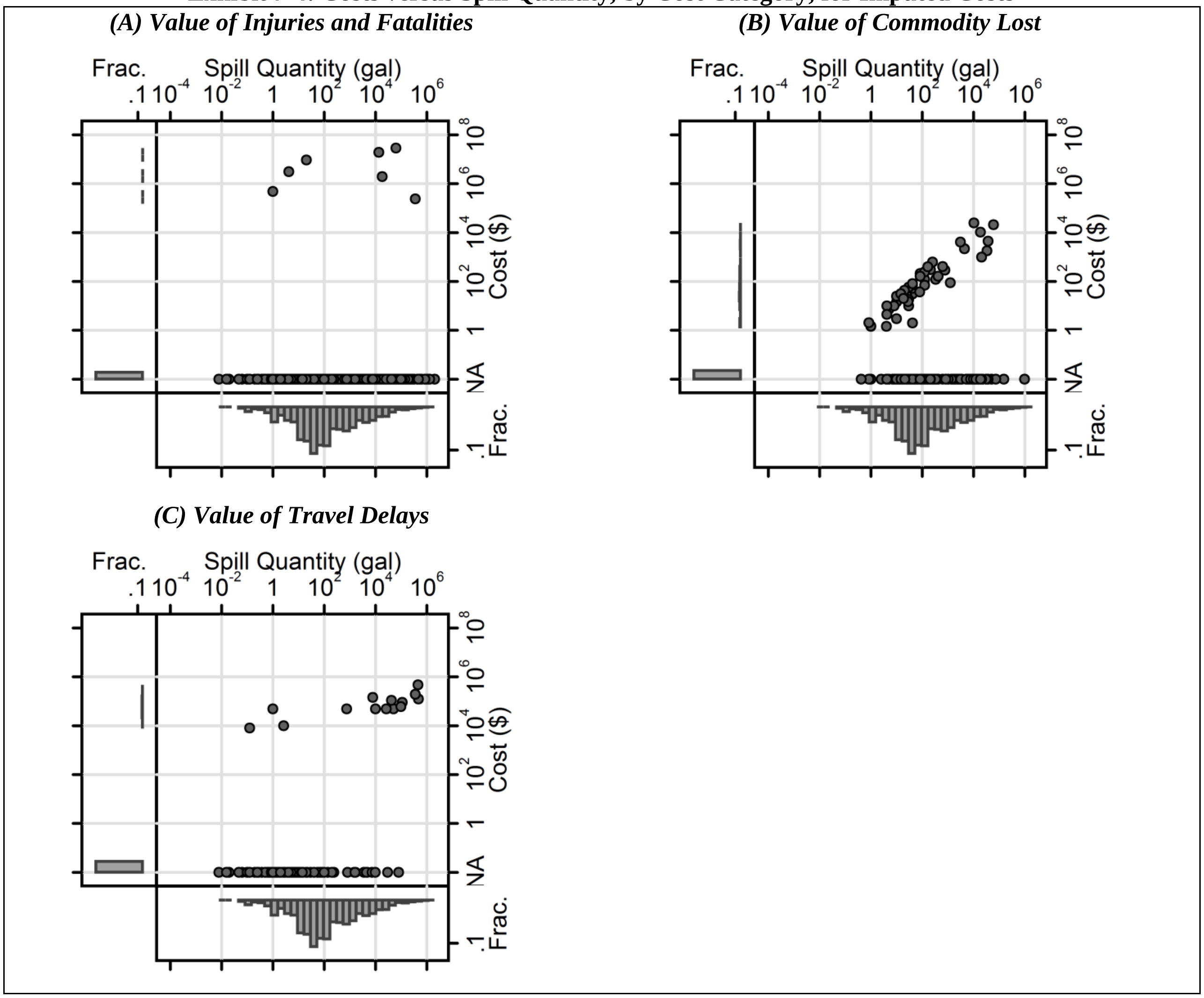


Exhibit 7-4 presents scatterplots of imputed costs versus quantity released, by cost category. Each panel in the figure represents a different imputed category of costs.

Exhibit 7-4: Costs versus Spill Quantity, by Cost Category, for Imputed Costs



To mitigate concerns about under-reporting of cost categories in the database underlying the analysis, we also ran regression models that use each cost category as the dependent variable. Exhibit 7-5 and Exhibit 7-6 show these model specifications for rail and pipeline incidents, respectively, for the cost categories that had a sufficient number of observations to generate a robust model. Running each of these cost category models for an incident and summing across them provides an upper-bound estimate of the costs for the incident, assuming that all cost categories would be incurred for the incident.

Exhibit 7-5: Regression-based Analysis of Costs Categories for Rail

Variable	Regression Specification				
	Public Property Damage	Commodity Loss Cost	Op. Property Damage	Op. Emergency Response Cost	Op. Env. Remediation
Intercept	3.707**	0.835***	13.79***	7.414***	6.333***
ln_qreleased	0.543	1.044***	0.192	0.0326	0.542***
ln_qreleased_sq	0.0220	-0.00846	-0.0161	0.0347***	0.00383
Observations	9	98	6	307	94
R^2	0.852	0.929	0.090	0.494	0.730
Root MSE	1.571	0.965	1.156	1.023	1.370

*Note: Each column represents a different regression model, where the dependent variable is the log of cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Exhibit 7-6: Regression-based Analysis of Costs Categories for Pipeline

Variable	Regression Specification						
	Public Property Damage	Public Response/ Remediation	Commodity Loss Cost	Op. Property Damage	Op. Emergency Resp.	Op. Env. Remed.	Op. Other Costs
intercept	7.662***	8.913	1.558***	8.214***	7.919***	7.210***	9.130***
ln_qreleased	-0.216	-0.595	0.732***	-0.232	-0.201	0.108	-0.541*
ln_qreleased_sq	0.0473**	0.0697	0.00306	0.0439***	0.0545***	0.0357***	0.0680***
hca_bi	1.037*	1.327	0.0326	0.568***	0.838***	0.472***	1.360**
fire_bi	0.884**	-3.275***	0.682*	3.055***	0.587	0.361	1.326***
Observations	126	18	1,044	1,013	968	834	142
R^2	0.338	0.460	0.782	0.173	0.354	0.420	0.371
Root MSE	2.062	2.092	1.019	2.141	1.769	1.718	1.957

*Note: Each column represents a different regression model, where the dependent variable is the log of cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

7.2 Issues Related to Spill Causality

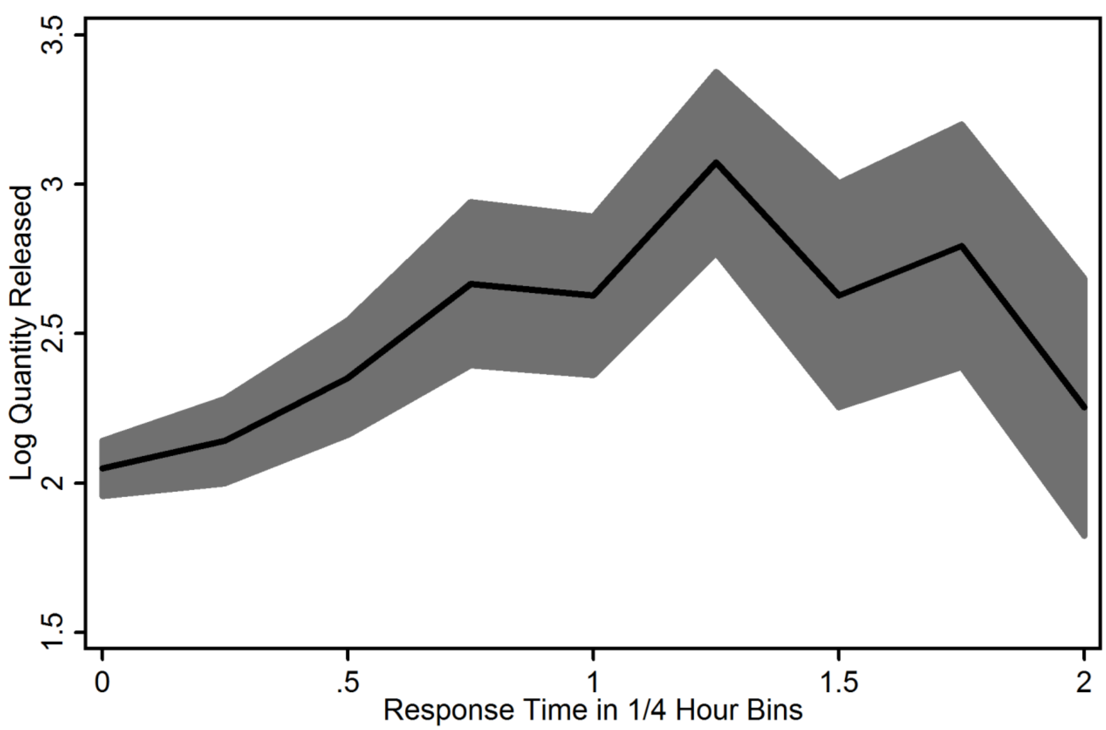
As discussed in Section 2.2, the causal chain that connects incident characteristics to incident costs is complicated. When more oil is released, incident response costs and remediation costs both increase. However, the causation can also go in the other direction: implementing a rapid, resource-intensive response can reduce the quantity of oil released, contain its spread in the environment, and reduce eventual remediation costs. This possibility of reverse causation creates an analytical challenge in interpreting the regression results.

To explore this issue, we used a common statistical procedure known as "instrumental variables" (IV) regression (sometimes referred to as "two-stage least-squares"). An IV regression is a way of measuring a causal relationship while controlling for the possibility of reverse causation or omitted variable bias. In this report, IV regression could be used to measure the effect of quantity released on costs.

The first step in an IV regression is to identify an exogenous source of variation in the independent variable; in other words, an "instrumental variable" that influences the quantity released, but does not directly affect response or remediation costs (except via its effects on quantity released). One such variable is response time; small differences in response time can have large effects on quantity released.

Exhibit 7-7 graphs average quantity released as a function of response time. The figure shows that on average, quantity released increases strongly with response time, particularly for response times less than one hour. Moreover, response time seems unrelated to at least some other incident characteristics. Exhibit 7-8 tabulates average incident characteristics, by response time. The exhibit shows that being in a HCA is correlated with response time (since the percentage of incidents in an HCA decreases as the quantity released increases), but that most other characteristics are not. Together, these two exhibits suggest that although response time is not a perfect instrumental variable, it may be sufficient to provide a reasonable robustness check.

Exhibit 7-7: Log Quantity Released vs Response Time in 15-minute Bins



Note: All response time values are rounded up to the nearest ¼ hour.

Exhibit 7-8: Spill Characteristics by Binned Response Time

Variable	Average Value of Variable, by Response Time (hours)								
	0	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**
blk_pop10_800	11,724	35,171	43,326	14,196	23,274	1,035	9,691	32,121	35,066
hca_bi	37	32	34	31	31	18	29	24	22
soilcont_bi	85	89	86	98	93	97	86	89	90
pcause_corrosion	26	37	39	40	39	38	37	10	13
pcause_equipment	41	44	29	37	33	35	46	52	52
pcause_excavation	2	2	5	7	6	10	0	10	4
pcause_incorrectopp	19	12	7	3	6	10	3	10	22
pcause_mwfail	5	2	5	5	6	0	6	5	9
pcause_natlforces	4	2	7	2	6	0	3	5	0
pcause_outforce	1	0	5	5	1	5	3	10	0
pcause_other	2	1	3	2	4	3	3	0	0
Count of Incidents	521	170	121	62	70	40	35	21	23

Notes: All variables represent percentage non-zero, except for the mean value of population block within 800 meters. All values are rounded up to nearest whole percentage.

After identifying an appropriate instrumental variable, the next step is to run a "first-stage" regression in which we regress the independent variable we wish to instrument (quantity released) on the full set of independent variables, replacing the target variable with the instrumental variable. In the equation below, z_1 is the instrument for x_d. In this particular case we define z_1 as the response time and x_d as the quantity released.[67]

$$\widehat{x_d} = \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_{d-1} x_{d-1} + \theta_1 z_1 + \theta_2 z_2 + \cdots + \theta_n z_n + \nu$$

Once we have our fitted or projected x_d values, we used this predicted variable in a "second-stage" regression where we replace x_d with $\widehat{x_d}$.

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{d-1} x_{d-1} + \beta_d \widehat{x_d} + \epsilon$$

Here, we regress the log of incident costs on the *predicted* (rather than actual) quantity released.[68] As shown in Exhibit 7-9, the coefficients on the "second-stage" regression resulting from this IV analysis are similar to the main results presented in Exhibit 5-8, suggesting that reverse-causality between costs and quantity released has little impact on the overall estimation of the coefficients in the pipeline model.

---

67 For additional information on the instrumental variables analysis, please see section 5.1.2 in Wooldridge (2010).

68 The intuition for this procedure is that by using response time to predict quantity released, we are isolating variation in quantity released that is truly unrelated to other factors that influence incident costs (including response costs). We can then use this exogenous variation to calculate an unbiased estimate of the effect of changes in quantity released on costs. (Of course, as we note above, response time is an imperfect instrumental variable, since it could be related to some characteristics—such as the spread of oil—that also affect costs.)

Exhibit 7-9: Regression-based Analysis of Total Costs, Instrumental Variable Analysis Comparison

Variable	Regression Specification	
	Primary Pipeline Regression[1]	Second-Stage IV Regression
intercept	8.113	7.97
ln_qreleased	-0.0769	N/A[2]
ln_qreleased_sq	0.0464	0.0393
pop800	0.000166	0.000133
watwet800_bi	0.298	0.333
spatial_missing	-1.195	0
et800_bi	0.708	0.456
lu_hddev_800	0.839	0.751
hca_bi	0.478	0.447
fire_bi	2.857	2.896
underground_bi	1.143	1.106
pcause_corrosion	0.0337	0.058
pcause_natlforces	0.653	0.698
Observations[2]	1,250	1,051

1. See Exhibit 5-8.
2. The instrumental variable regression uses the first-stage to predict a single independent variable value, in this case ln_qreleased_sq.
3. We excluded incidents with response time exceeding two hours due to high variance and apparent inconsistencies in reporting.

7.3 General Uncertainties in Analysis

The database consists of data assembled from different datasets, with varying data collection objectives and levels of detail. As such, there may be systematic differences in data availability and quality across incident types. For example, the PHMSA pipeline dataset from 2009 and earlier solicits data on public emergency response and environmental remediation expenditures whereas the later pipeline dataset and the rail dataset are both limited to operator expenditures in those categories. As noted in Section 7.1, these inconsistencies may introduce systematic bias in the data that does not correspond to actual variation in total costs but rather simply reflects reporting requirements. Furthermore, even when the data entry forms request the same information, operators may or may not account for all cost categories in their reporting, and they may not report costs consistently.

The rail data set includes substantially less data on potentially important variables such as location (e.g., latitude and longitude), surface water impact indicators, and other information. In some cases, we supplemented the data based on other datasets such as the POLREP indicator for surface water impacts and FRA data on latitude and longitude. However, our ability to fill in the missing information was limited by the availability of matching records and data in these other datasets.

Another limitation is that we do not distinguish between types of crude oil that may behave differently when released in the environment. For example, Bakken crude oil is more volatile and has more damage potential than other types of crude oil (Frittelli et al., 2014). The PHMSA datasets do not include an indicator for the type of crude oil, and data do not exist to supplement this information for a sufficient number of incidents in the database and analysis.

7.3.1 Uncertainty Leading to Potential Overstatement

The analysis is based on total reported damages or costs reported for all rail or pipeline incidents in which crude oil was released. However, not all damages may be directly attributable to the crude oil release. Some damages may have been caused by other precipitating events (e.g., physical impact, fire, natural catastrophe) that are independent of the crude oil release. It is not possible to disentangle damages and costs directly attributable to the crude oil release versus those associated with other concurrent events. For example, a vehicular accident or derailment may cause damages to a railcar or other property that are independent of whether crude oil was also released during the incident. Including all costs may overstate the costs associated directly with a crude oil release by itself, but is appropriate if future incidents will be of a similar character as those used in deriving the cost model, i.e., where PHMSA would apply the cost model to incidents where crude oil release is one of several precipitating factors, or the main event. One possible approach to address the uncertainty would be to limit the total costs for deriving the regression models to only those costs that are most likely to be directly attributable to the crude oil release – e.g., cleanup and remediation, environmental damages. Alternatively, one could also use the models for individual cost categories in Exhibit 7-5 and Exhibit 7-6 to estimate expected costs for cleanup, remediation and environmental damages, and exclude other cost categories from the predictions.

7.3.2 Uncertainty Leading to Potential Understatement

As described in Section 3, operator reports to PHMSA constitute the primary source of data used in this analysis. However, the data entry forms that are used to collect the information from the operators are not designed to collect a comprehensive accounting of the costs associated with an incident. We supplemented the costs for incidents where possible based on supplemental data sources; however, it is likely that some costs associated with the crude oil releases in the database are under-represented. This may be particularly significant for more recent incidents, since it can often take months or even years for a full accounting of available cost estimates.

There is some indication that incident consequences may be underreported or underestimated. For example, Enbridge Energy reported that a 2010 release from a pipeline in Marshall, Michigan (PHMSA incident 20100181) resulted in 61 evacuations and no injuries or fatalities. However, an NTSB report on the incident (NTSB, 2012) stated that 331 people (including 11 worksite employees) experienced adverse health effects from the incident, including headaches, nausea, and respiratory effects. The impacts to these individuals were not included in the information reported to PHMSA, since the pipeline reports solicit information about injuries requiring hospitalization only.

As another example, the Refugio oil spill (PHMSA incident 20150224) has estimated operator costs of $142.9 million, but these costs do not include legal claims and potential settlements estimated at $257 million (Pacific Coast Business Times, 2015a) or the $74 million in estimated impacts to Santa Barbara County's economy from foregone taxes, employment income, and federal royalties if the pipeline remains closed while repairs are done (Pacific Coast Business Times, 2015b).

Additionally, operators may be inconsistent in how they report incident damages and consequences. For example, for the Marshall incident, Enbridge reported 61 evacuations, while NTSB (2012) states that 61 *houses* were evacuated, which indicates that the number of *individuals* affected may be higher. For that same incident, Enbridge reported that 843,444 gallons were released; however, USFWS, et al. (2015) reports that by April 2012, 1.2 million gallons of oil had been recovered with additional oil remaining in river sediments.

8 Conclusions and Recommendations

This section highlights conclusions to be drawn from this study and outlines some recommendations for PHMSA to consider as it continues to enhance its ability to monetize the benefits of safety regulations or other policy interventions that reduce the risk of hazardous liquid releases in transportation.

8.1 Conclusions

The study achieves PHMSA's stated purpose, which was to develop rigorous, defensible estimates of the social costs of onshore releases of crude oil from pipelines and railcars. The regression models presented in Section 5.2 allow PHMSA to estimate expected costs from crude oil releases, accounting for incident characteristics that include the transportation mode, quantity released, incident location and affected environments, and cause, among other factors.

As discussed throughout the report and demonstrated by the regression analysis, incident costs can vary significantly across incidents, so the ability to account for the differences should markedly improve the validity of the estimates as compared to prior analyses that relied on uniform unit costs. Additional cost models provide further insight on the sensitivity of the cost estimates to modeling assumption or to data limitations.

As part of the study, we made a concerted effort to develop a comprehensive database of the social costs of onshore crude oil releases, imputing costs where we could reasonably do so. This database provides a rich set of information about historical incidents that could support a variety of future analyses.

8.2 Recommendations

Develop Companion Methodology to Project Incident Occurrence

The regression models provide the expected costs of an incident, conditional on its occurrence. Application of the model will therefore rest in large part on the ability to project incident probabilities and characteristics, and to determine how incidents change as a result of a regulation or other policy intervention. As presented in the illustrative analyses of Section 6.2, one simple approach for applying the model entails using the historical record as the basis for estimating baseline conditions, modeling the characteristics for these incidents, and identifying which incidents in the dataset may have been prevented or changed by the policy.

A possible companion research task, however, could entail developing more sophisticated models to estimate the probability of an incident for various scenarios (e.g., pipeline within HCA) and the amount of oil released during the incident. Empirical models could be developed that relate the probability of an incident and the incident characteristics to the characteristics of the pipeline or rail network, geographical location, product throughput, and other factors. This could be another way of capturing differences in incident characteristics for different types of crude oils or hazardous liquids (e.g., where an incident involving Bakken Crude may be relatively more likely to also result in a fire).

The two types of models could then be combined into a two-step analysis whereby PHMSA would first predict the probability of a given incident occurring for various scenarios, and then use the models described in Section 5.2 to estimate the corresponding costs for each scenario. The policy or regulation could be modeled as having an effect on the probability of an incident occurring, the expected magnitude of the incident, and/or the resulting costs. This approach may enable PHMSA to account for more factors that determine the likelihood of different types of incidents occurring, for example as the volume of product transported changes over time or across geographical areas.

Continue to Maintain the Database and Cost Models

Accompanying this report is a comprehensive database of incidents involving onshore crude oil releases from rail and pipeline, as well as all the programs we used to develop the cost models. Data gaps such as the geographic coordinates of most rail incidents, limited our ability to account for some potentially important predictors of incident costs. PHMSA may want to compile these data – which may require changes to the incident data requested from operators – and update the models as additional data become available. Similarly, PHMSA may want to continue to maintain the database and update the models to incorporate lagging cost data or information for more recent incidents.

Develop Models for Other Hazardous Liquids

The study provides a generalizable methodology for relating the cost of incidents to their characteristics. PHMSA can use the same approach for estimating models for other hazardous liquids. Doing this will require a similar effort to assemble data on other hazardous liquid incidents, complement these data as needed, and conduct analyses to understand and quantify the effect of various factors on the total incident costs.

9 References

Abt Associates, Inc. and Bear Peak Economics. 2014. Potential Fishing Impacts and Natural Resource Damages from Worst-Case Discharges of Oil on the Columbia River. Report in the Matter of Application No. 2013-01 Vancouver Energy Distribution Terminal EFSEC Case Number 15-001. Submitted to Matthew Kernutt, Assistant Attorney General, Washington Attorney General's Office.

Association of American Railroads (AAR). 2016a. Railroads and the Economy, North East Association of Rail Shippers. April 27, 2016.

Association of American Railroads (AAR). 2016b. What We Haul: Crude Oil by Rail. Available from: https://www.aar.org/todays-railroads/what-we-haul/crude-oil-by-rail

Association of American Railroads (AAR). 2015. U.S. Rail Crude Oil Traffic.

Association of Oil Pipe Lines and American Petroleum Institute. 2014. U.S. Liquids Pipeline Usage & Mileage Report.

Atkin, E. 2015. "Data: Oil Trains Spill More Often, But Pipelines Spill Bigger." ThinkProgress.

Bureau of Transportation Statistics, United States Department of Transportation. 2015. Table 1-10: U.S. Oil and Gas Pipeline Mileage. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_01_10.html

California Department of Fish and Wildlife (California DFW). 2015. Refugio Beach Oil Spill Natural Resource Damage Assessment Update: November 2015.

Caleyo, F., Velázquez, J. C., Valor, A., & Hallen, J. M. 2009. Probability distribution of pitting corrosion depth and rate in underground pipelines: A Monte Carlo study. Corrosion Science, 51(9), 1925-1934.

Chang, S.E., J. Stone, K. Demes, and M. Piscitelli. 2014. Consequences of Oil Spills: A Review and Framework for Informing Planning. Ecology and Society (19): 26.

CNN Wire Service. 2015. Investigators Reveal Cause of Huge West Virginia Train Derailment and Fireball. Fox 6 Now, October 9, 2015. Available from: http://fox6now.com/2015/10/09/investigators-reveal-cause-of-huge-west-virginia-train-derailment-and-fireball/

Damage Assessment, Remediation, and Restoration Project (DARRP), National Oceanic and Atmospheric Administration. 2016. Case Documents. Case: Refugio Beach, CA. https://casedocuments.darrp.noaa.gov/southwest/refugio-beach/admin.html

Dawotola, A. W., Trafalis, T. B., Mustaffa, Z., Van Gelder, P. H. A. J. M., & Vrijling, J. K. 2012. Risk-based maintenance of a cross-country petroleum pipeline system. Journal of Pipeline Systems Engineering and Practice, 4(3), 141-148.

Dawotola, A. W., Van Gelder, P. H. A. J. M., & Vrijling, J. K. 2009. Risk Assessment of Petroleum Pipelines using a com-bined Analytical Hierarchy Process-Fault Tree Analysis (AHP-FTA).

Eifling, S. 2013. The Forgotten in Mayflower: People Who Live Within Sight of the Ruptured Pegasus Pipeline Say They've Been Ignored by ExxonMobil and Government Officials. Arkansas Times, August 08, 2013. Available from: http://www.arktimes.com/arkansas/the-forgotten-mayflower-residents/Content?oid=3007639&showFullText=true

Enbridge, Inc. 2015. 2014 Annual Report.

Enbridge, Inc. 2014. Annual Report: 2013.

Environmental Research Consulting (ERC). 2016. Crude-by-Rail Spill Risk Analysis for Proposed Shell Puget Sound Refinery Anacortes Rail Unloading Facility: Rail Spill Probability and Volume Analysis Draft Report.

Etkin, D.S. 2004. Modeling Oil Spill Response and Damage Costs. Proceedings of the 5th Biennial Freshwater Spills Symposium.

Fehling, D. 2012. Moving Crude Oil Relies on Aging Pipeline System. State Impact: A Reporting Project of NPR Member Stations. Available from: https://stateimpact.npr.org/texas/2012/09/05/moving-crude-relies-on-aging-pipeline-system/

Francingues, M. 2015. Officials Issue Fish-Consumption Advisory for Yellowstone River. Sidney Herald. Retrieved from: http://www.sidneyherald.com/news/officials-issue-fish-consumption-advisory-for-yellowstone-river/article_f00804bc-a358-11e4-a7e8-43b15d3051ef.html

Freshwater Fishing News. 2015a. Yellowstone River Oil Spill 2015. Retrieved from: http://www.freshwater-fishing-news.com/2015/01/yellowstone-river-oil-spill-2015/

Freshwater Fishing News. 2015b. Kanawha River Oil Spill. Retrieved from: http://www.freshwater-fishing-news.com/2015/02/kanawha-river-oil-spill/

Frittelli, J. A. Andrews, P.W. Parfomak, R. Pirog, J.L. Ramseur, and M. Ratner. 2014. U.S. Rail Transportation of Crude Oil: Background and Issues for Congress. Congressional Research Service. 7-5700.

Frosch, D. 2013. Oil Spill In North Dakota Raises Detection Concerns, New York Times, October 23, 2013. Available from: http://www.nytimes.com/2013/10/24/us/oil-spill-in-north-dakota-raises-detection-concerns.html

Girgin, S., & Krausmann, E. 2016. Historical analysis of US onshore hazardous liquid pipeline accidents triggered by natural hazards. Journal of Loss Prevention in the Process Industries, 40, 578-590.

Hallenbeck, M.E., A. Goodchild, and J. Drescher. 2014. Travel Costs Associated with Flood Closures of State Highways Near Centralia/Chehalis, Washington. Washington Transportation Center (TRAC-UW). Prepared for the State of Washington Department of Transportation.

Hansen, J., E.D. Benson, and D.A. Hagen. No date. Environmental Hazards and Residential Property Values: Evidence from a Major Pipeline Event.

Helton, D. and T. Penn. 1999. Putting Response and Natural Resource Damage Costs in Perspective. International Oil Spill Conference Proceedings: March 1999: 577-583.

Hill, R. T. and A.D. Director. 1993. Pipeline risk analysis. In Institution of Chemical Engineers Symposium Series (Vol. 130, pp. 657-657). HEMSPHERE PUBLISHING CORPORATION.

Holland, M. 2010. Trans-Alaska Pipeline Spill Toll 5,000 Barrels. AND.Com, May 28, 2010; Updated September 29, 2010. Available from: https://www.adn.com/economy/article/trans-alaska-pipeline-spill-toll-5000-barrels/2010/05/28/

Hovey, Diane J., Edward J. Farmer. 1993. PIPELINE ACCIDENT, FAILURE PROBABILITY DETERMINED FROM HISTORICAL DATA. Oil and Gas Journal.

Jepson, Michael and Lisa L. Colburn 2013. Development of Social Indicators of Fishing Community Vulnerability and Resilience in the U.S. Southeast and Northeast Regions. U.S. Dept. of Commerce., NOAA Technical Memorandum NMFS-F/SPO-129, 64 p.

Johnston, Robert J., et al. (eds). 2015. Benefit Transfer of Environmental and Resource Values. Springer Science + Business Media.

Johnston, R.J., E. Besedin, and M. Ranson. 2006. "Characterizing the Effects of Valuation Methodology in Function-Based Benefits Transfer." Ecological Economics 60(2): 407-419.

Kacik, A. 2016. Plaintiffs File for Class Certification in Refugio Oil Spill Lawsuit. Pacific Coast Business Times. Retrieved from: http://www.pacbiztimes.com/2016/08/25/plaintiffs-file-for-class-certification-in-refugio-oil-spill-lawsuit/

Klemann, D. 2015. Official: 35,000 gallons of oil spilled in eastern Montana derailment. KPAX. Retrieved from http://www.kpax.com/story/29576456/official-35000-oil-gallons-spilled-in-eastern-montana-derailment

Little, A.D. 1999. "Risks from Gasoline Pipelines in the UK", Health and Safety Executive, UK.

Lovett, A.H., C.T. Dick, and C.P.L. Barkan. 2015. Determining Freight Train Delay Costs on Railroad Lines in North America.

Marathon Pipe Line LLC. 2011. Natural Resource Damages Assessment and Restoration Planning (DARP) Report. Mt. Erie Pipe Line Release Near Albion, Wayne County, Illinois. Prepared by Arcadis. November 2010, Revised March 2011.

Merten, A.A. and C.B. Henry. 2008. Decision-Making Process to Use In-Situ Burning to Restore and Oiled Intermediate Marsh Following Hurricanes Katrina and Rita. International Oil Spill Conference: 545-550.

Mesa-Arango, R., X. Zhan, S.V. Ukkusuri, A. Mitra, and F. Mannering. 2013. Estimating the Economic Impacts of Disruptions to Intermodal Freight Systems Traffic. NEXTRANS Project No. 053PY03: Final Report.

Montana Department of Justice (Montana DOJ). 2013. Yellowstone River Oil Spill (July 2011): Natural Resource Damage Assessment and Restoration.

Montana Fish, Wildlife, and Parks (Montana FWP). 2015. FWP Issues Fish-Consumption Advisory Below Oil Spill: Fishing – Region 7. Retrieved from: http://fwp.mt.gov/news/newsReleases/fishing/nr_0882.html

Nalder, E. 2010. Old Pipelines Are a Hidden Menace. Houston Chronicle, November 15, 2010. Available from: http://www.chron.com/business/energy/article/Old-pipelines-are-a-hidden-menace-1605649.php

National Oceanic and Atmospheric Administration (NOAA). 2015. 10 Years After Being Hit by Hurricane Katrina, Seeing an Oiled Marsh at the Center of an Experiment in Oil Cleanup. Available from: http://response.restoration.noaa.gov/about/media/10-years-after-being-hit-hurricane-katrina-seeing-oiled-marsh-center-experiment-oil-clea

National Oceanic and Atmospheric Administration (NOAA). No date. Introduction to Natural Resource Damage Assessment. Damage Assessment, Remediation, and Restoration Program (DAARP).

National Transportation Safety Board (NTSB). 2016. CSXT Petroleum Crude Oil Train Derailment and Hazardous Materials Release: DCA14FR008. NTSB/RAB-16/01.

National Transportation Safety Board (NTSB). 2014. Preliminary Report: Railroad: DCA14MR004.

National Transportation Safety Board (NTSB). 2013. Pipeline Accident Brief: DCA-10-FP-009. NTSB/PAB-13/03.

National Transportation Safety Board (NTSB). 2012. Accident Report: Enbridge Incorporated Hazardous Liquid Pipeline Rupture and Release, Marshall Michigan, July 25, 2010. NTSB/PAR-12/01. PB2012-916501.

Oak Ridge National Laboratory (ORNL). 2012. Studies for the Requirements of Automatic and Remotely Controlled Shutoff Valves on Hazardous Liquids and Natural Gas Pipelines with Respect to Public and Environmental Safety. ORNL/TM-2012/411.

O'Brien, E. 2015. FWP Lifts Fish Advisory After Yellowstone River Oil Spill. Montana Public Radio. Retrieved from: http://mtpr.org/post/fwp-lifts-fish-advisory-after-yellowstone-river-oil-spill#stream/0

Pacific Coast Business Times. 2015a. Refugio oil spill pipeline might take five years to get back online. November 5, 2015. http://www.pacbiztimes.com/2015/11/05/refugio-oil-spill-pipeline-might-take-five-years-to-get-back-online/

Pacific Coast Business Times. 2015b. 3 months after oil spill, Santa Barbara region recovering. August 8, 2015. http://www.pacbiztimes.com/2015/08/14/3-months-after-oil-spill-santa-barbara-region-recovering/

Pholphiboun, A. 2015. Montana train derailment strands passengers in Whitefish. KPAX. Retrieved from http://www.kpax.com/story/29578633/montana-train-derailment-strands-passengers-in-whit

Pipeline and Hazardous Materials Safety Administration (PHMSA), United States Department of Transportation. 2016a. Preliminary Regulatory Impact Analysis for Hazardous Materials: Oil Spill Response Plans and Information Sharing for High-Hazard Flammable Trains – Proposed Rule. PHMSA-2014-0105 (HM-251B).

Pipeline and Hazardous Materials Safety Administration (PHMSA), United States Department of Transportation. 2016b. Hazardous Liquid Pipeline Miles and Tanks by Commodity. Retrieved from: https://hip.phmsa.dot.gov/analyticsSOAP/saw.dll?Portalpages

Pipeline and Hazardous Materials Safety Administration (PHMSA), United States Department of Transportation. 2015. Final Regulatory Impact Analysis for Hazardous Materials: Enhanced Tank Car Standards and Operational Controls for High-Hazard Flammable Trains. PHMSA-2012-0082.

Pipeline and Hazardous Materials Safety Administration (PHMSA), United States Department of Transportation. 2012. Emergency Response Guidebook.

Plains All American Pipeline LP. 2016. Form 10-K (Annual Report) Filed 02/25/16 for the Period Ending 12/31/15.

Puckett, K. 2015. Wildlife, Fish Being Tested at Oil Spill. Great Falls Tribute, January 22, 2015. Retrieved from: http://www.greatfallstribune.com/story/news/local/2015/01/21/wildlife-fish-tested-oil-spill/22116321/

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(1), 38-50.

Santa Clara River Trustee Council. 2002. Final Restoration Plan and Environmental Assessment for the Santa Clara River ARCO Oil Spill. United States Fish and Wildlife Service and California Department of Fish and Game, Office of Spill Prevention and Response.

Schleifstein, M. 2010. Extent of Oil Spills from Hurricanes Katrina and Rita is Still Being Assessed. The Times-Picayune, August 19, 2010. Available from:

http://www.nola.com/katrina/index.ssf/2010/08/extent_of_oil_spills_from_2005_hurricanes_is_still_being_assessed.html

Schreiner, B. 2005. Pipeline Spills 63,000 Gallons of Oil into Kentucky River. St. Augustine Record, January 27, 2005. Available from: http://staugustine.com/stories/012705/nat_2850874.shtml#.WFMLk1UrLb0

Shahriar, A., Sadiq, R., & Tesfamariam, S. 2012. Risk analysis for oil & gas pipelines: A sustainability assessment approach using fuzzy based bow-tie analysis. *Journal of Loss Prevention in the Process Industries*, *25*(3), 505-523.

Simons, R. A. 1999. Settlement of an Oil Pipeline Leak with Contaminated Residential Property: A Case Study. Real Estate Issues, 24(2), 46-52.

Simons, R. A., Bowen, W. M., & Sementelli, A. 1996. The Effect of Underground Storage Tanks on Residential Property Values in Cuyahoga County Ohio. Journal of Real Estate Research, 14, 29-42.

Simons, R. A., Mikelbank, B. A., & Winson, K. 2001. The Effects of An Oil Pipeline Rupture on Single Family House Prices. The Appraisal Journal, 37165.

Smith, D. and B. Mejia. 2016. Pipeline Company Indicted in 2015 Santa Barbara County Oil Spill. Los Angeles Times. http://www.latimes.com/local/lanow/la-me-ln-santa-barbara-county-oil-spill-20160517-snap-story.html

Steen, A., D.E. Fritz, W. Stubblefield, and J. Giddings. 1999. Environmental Effects of Freshwater Oil Spills. 1999 International Oil Spill Conference: 607-612.

United States Department of Transportation (U.S. DOT). 2016. Guidance on Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analyses – 2016 Adjustment. Memorandum to: Secretarial Officers, Modal Administrators.

United States Department of Transportation (U.S. DOT). 2015a. The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations. Revision 2 (2015 Update).

United States Energy Information Administration (U.S. EIA). 2016a. Total Crude Oil by Rail (Thousand Barrels). Release date 10/31/2016. (available from: http://tonto.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=ESM_EPC0_RAIL_ZAMN-ZAMN_MBBL&f=M)

United States Energy Information Administration (U.S. EIA). 2016b. U.S. Crude Oil First Purchase Price. Release date 4/1/2016. (available from: https://www.eia.gov/dnav/pet/hist_xls/F000000__3a.xls)

United States Energy Information Administration (U.S. EIA). 2015. Petroleum & Other Liquids: Refinery Receipts of Crude Oil by Method of Transportation. Release date 6/19/2015. (available from: http://www.eia.gov/dnav/pet/pet_pnp_caprec_dcu_nus_a.htm)

United States Environmental Protection Agency (U.S. EPA). 2016. Sensitivity of Freshwater Habitats. Retrieved from: https://archive.epa.gov/emergencies/content/learning/web/html/freshwat.html

United States Environmental Protection Agency (U.S. EPA). April 2015. News Release. "U.S. Settles with ExxonMobil over Violations Stemming from 2013 Oil Spill in Mayflower, Arkansas" https://yosemite.epa.gov/opa/admpress.nsf/0/9330c87e8d8e843e85257e2f0047fcd7

United States Environmental Protection Agency (U.S. EPA) Science Advisory Board (SAB). 2009. Valuing the Protection of Ecological Systems and Services. EPA-SAB-09-012.

United States Environmental Protection Agency (U.S. EPA) Region V. 2015a. Pollution/Situation Report. Plains Pipeline Oil Spill -- Polrep #8 -- Final.

United States Environmental Protection Agency (U.S. EPA) Region V. 2015b. Pollution/Situation Report. BNSF Galena Derailment -- Polrep #15 -- Final.

United States Environmental Protection Agency (U.S. EPA) Region V. 2014. Pollution/Situation Report. Mid-Valley Pipeline Crude Oil Release -- Removal Polrep: Final Removal Polrep.

United States Environmental Protection Agency (U.S. EPA) Region VI. 2014a. Pollution/Situation Report. Delek Pipeline -- Polrep #6.

United States Environmental Protection Agency (U.S. EPA) Region VI. 2014b. Pollution/Situation Report. Mid Valley Pipeline Oil Spill -- Polrep #3.

United States Environmental Protection Agency (U.S. EPA) Region VI. 2013. Pollution/Situation Report. Lion Oil Magnolia Spill -- Polrep #7.

United States Environmental Protection Agency (U.S. EPA) Region VI. 2011. Pollution/Situation Report. Centurion Pipeline Spill -- Polrep #2: Final Removal Polrep.

United States Environmental Protection Agency (U.S. EPA) Region VIII. 2015a. Pollution/Situation Report. Bridger Pipeline Release -- Polrep #12 -- Final.

United States Environmental Protection Agency (U.S. EPA) Region VIII. 2015b. Pollution/Situation Report. Heimdal Crude Oil Train Derailment-- Polrep #2 – Final.

United States Fish and Wildlife Service (USFWS), Nottawaseppi Huron Band of the Potawatomi Tribe, and Match-E-Be-Nash-She-Wish Band of the Pottawatomi Indians. 2015. Final Damage Assessment and Restoration Plan/Environmental Assessment for the July 25-26, 2010 Enbridge Line 6B Oil Discharges near Marshall, MI. Prepared in Cooperation with: National Oceanic and Atmospheric Administration, Michigan Department of Environmental Quality, Michigan Department of Natural Resources, and Michigan Department of Attorney General.

United States Fish and Wildlife Service (USFWS) and United States Census Bureau. 2014. 2011 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation. FHW/11-NAT (RV).

Utah Department of Environmental Quality (UDEQ). 2012a. Screening Level Ecological Risk Assessment – Final: Chevron Pipeline June 2010 Crude Oil Spill, Lower Red Butte Creek, Salt Lake City, Utah. Project No. 0145323.

Utah Department of Environmental Quality (UDEQ). 2012b. Human Health Risk Assessment: Red Butte Creek, Salt Lake City, Utah. Prepared by McDaniel Lambert, Inc.

Utah Water Quality Board. 2011. Settlement Agreement in the Matter of Chevron Pipeline Company, Salt Lake City, Utah. Docket Number 110-1.

Weinstein, B. L., & Clower, T. L. 1998. The Impacts of the Union Pacific Service Disruptions on the Texas and National Economies: An Unfinished Story.

Woolridge, J.M. 2010. Econometric Analysis of Cross Section and Panel Data. 2nd ed., The MIT Press.

Young, L. 2014. "Crude oil spills are bigger from trains than pipelines." Global News. http://globalnews.ca/news/1069624/how-do-crude-spills-compare-by-rail-truck-pipeline-you-may-be-surprised/

Zhang, Y., M.K. Lindell, and C.S. Prater. 2009. Vulnerability of Community Businesses to Environmental Disasters. Disasters 33: 38-57.

Appendix A Comparison with Prior Oil Unit Costs Developed by PHMSA

This study builds on prior efforts by PHMSA to estimate a per-gallon cost of crude oil released during transportation incidents. As discussed in Section 1.1, PHMSA previously used historical incident data to estimate the unit costs from crude oil spill discharges and applied the unit costs to estimate the benefits of preventing crude oil releases though enhanced rail safety requirements for the final HM-251 rule (Hazardous Materials: Enhanced Tank Car Standards and Operational Controls for High-Hazard Flammable Trains; PHMSA, 2015). In this appendix, we compare the values previously developed by PHMSA, to those that would be obtained using the database developed for this study, as described in Section 2.

We note that there are two possible approaches to calculating the "cost per gallon." In this Appendix, we sum the total costs across all incidents then divide these total costs by the total quantity released across all incidents. This approach values each gallon of crude oil released equally, implicitly assuming a linear relationship between the quantity spilled and the total costs of the incident. An alternative unit cost calculation places equal value on each incident (rather than on each gallon released), by calculating a unit cost for each incident then averaging across all incidents.

Exhibit A-1 summarizes the data that PHMSA used to derive unit cost estimates for the HM-251 analysis. The data consist of 973 onshore and offshore releases of crude oil from pipelines between 2010 and 2015.[69]

Exhibit A-1: Pipeline Spill Size and Cost of Spill from PHMSA (2015)

Spill Size	Onshore and Offshore Pipeline Incidents through 3/16/2015)			
	Count of Incidents	Total Quantity Released (Gal)	Total Cost[1] (Million $)	Cost per Gallon[2] ($/Gal)
All	973	7,140,432	$1,508	$211
>100,000	15	4,670,484	$1,050	$225
50,000 to 99,999	7	474,348	$174	$368
10,000 to 49,999	59	1,309,300	$152	$116
1,000 to 9,999	158	576,685	$55	$96
500 to 999	70	51,484	$14	$276
100 to 499	187	42,394	$20	$465
50 to 99	97	7,547	$8	$1,050
5 to 49	350	8,143	$20	$2,481
<5	30	47	$14	$294,496

Source: PHMSA (2015); Table EB4.
1 Based on the sum of costs reported in nominal dollars for each incident; Total costs are not adjusted to constant dollars.
2 Cost per gallon calculated by dividing the total costs by the total quantity released in each size category.

69 PHMSA provided the underlying data in an email dated March 29, 2016. The data seem to be based on the Hazardous Liquid Accident Data Form F 7000-1 (rev 7-2014), January 2010 to Present available at the time the regulatory impact analysis was conducted. The operator-reported costs do not seem to have been adjusted to constant dollars.

Exhibit A-2 summarizes the same calculations but this time based on the raw PHMSA data used in this study. These data reflect updated data available in the fall of 2015. In the table, we show the costs as reported by operators, as well as the same costs restated in 2015 constant dollars. As compared to data used for Exhibit A-1, this data set includes revisions operators made to their incident reports between March 2015 and November 2016, including additional incidents. Thus, Exhibit A-2 includes a larger set of incidents (1,282 compared with 973). The overall cost per gallon based on the more comprehensive data set is higher at $217 per gallon as compared to $211, before adjusting to constant dollars, and $230 per gallon after the adjustment.

Exhibit A-2: Pipeline Crude Oil Spill Quantities and Costs Based on Raw PHMSA Pipeline Data[1]

Spill Size (Gallons)	Count	Total Quantity Released (Gallon)	Total Cost[2] (Million $)	Cost[2] per Gallon ($/gal)	Total Cost[3] (Million 2015$)	Cost[3] per Gallon (2015$/gal)
All	1,282	8,032,896	$1,745	$217	$1,849	$230
>100,000	17	4,809,756	$1,203	$250	$1,286	$267
50,000 to 99,999	8	552,048	$178	$323	$188	$340
10,000 to 49,999	76	1,704,037	$184	$108	$192	$112
1,000 to 9,999	217	820,403	$90	$109	$91	$111
500 to 999	94	68,754	$17	$243	$17	$247
100 to 499	247	57,389	$27	$469	$28	$480
50 to 99	126	9,723	$8	$850	$9	$876
5 to 49	459	10,701	$26	$2,456	$27	$2,532
<5	38	87	$12	$136,191	$12	$140,027

1 Hazardous Liquid Accident Data. Form F 7000-1 (rev 7-2014), January 2010 to Present. Includes onshore and offshore pipeline incidents through June 2016.
2 Costs as reported by operator (not adjusted to constant dollars).
3 Costs restated to 2015 dollars based on the Bureau of Labor Statistics Consumer Price Index

Exhibit A-3 summarizes the values calculated based on the database we compiled for this study after adding data from other sources and making further adjustments described in Section 3 of this report. In Exhibit A-3, we show both the spill costs reported directly in dollar terms, as well as adjusted costs that include imputed values for non-monetized damages such as reported injuries and fatalities. In contrast to the values in Exhibit A-1 and Exhibit A-2, the final database with imputed costs shows higher unit cost for spills, at $303 per gallon.

Exhibit A-3: Pipeline Crude Oil Spill Quantities and Costs Based on Final Database[1]

Spill Size (Gallons)	Count	Total Quantity Released (Gallon)	Total Costs[2]		Total Costs with Imputed Values[2,3]	
			Total Cost (Million 2015$)	Cost per Gallon (2015$/gal)	Total Cost (Million 2015$)	Cost per Gallon (2015$/gal)
All	1,282	8,032,896	$1,849	$230	$2,431	$303
>100,000	17	4,809,756	$1,286	$267	$1,818	$378
50,000 to 99,999	8	552,048	$188	$340	$224	$407
10,000 to 49,999	76	1,704,037	$192	$112	$192	$113
1,000 to 9,999	217	820,403	$91	$111	$92	$112
500 to 999	94	68,754	$17	$247	$17	$247
100 to 499	247	57,389	$28	$480	$28	$480
50 to 99	126	9,723	$9	$876	$9	$876
5 to 49	459	10,701	$27	$2,532	$36	$3,396
<5	38	87	$12	$140,027	$15	$175,626

1 Data reflect information compiled from several sources, as detailed in Section 3. Includes onshore and offshore pipeline incidents from 2010 through June 2016.
2 Costs restated to 2015 dollars based on BLS Consumer Price Index.
3 Costs include monetized value of reported injuries, fatalities, net product lost, and travel delays.

Exhibit A-4 is the same as Exhibit A-3, except it includes only those pipeline incidents that were included in the final database (i.e., it excludes incidents that occurred offshore).

Exhibit A-4: Pipeline Crude Oil Spill Quantities and Costs Based on Final Database[1]

Spill Size (Gallons)	Count	Total Quantity Released (Gallon)	Total Costs[2]		Total Costs with Imputed Values[2,3]	
			Total Cost (Million 2015$)	Cost per Gallon (2015$/gal)	Total Cost (Million 2015$)	Cost per Gallon (2015$/gal)
All	1,256	7,993,766	$1,816	$227	$2,398	$300
>100,000	17	4,809,756	$1,286	$267	$1,818	$378
50,000 to 99,999	8	552,048	$188	$340	$224	$407
10,000 to 49,999	75	1,687,237	$176	$105	$177	$105
1,000 to 9,999	214	799,277	$86	$107	$86	$107
500 to 999	94	68,754	$17	$247	$17	$247
100 to 499	243	56,561	$25	$444	$25	$444
50 to 99	123	9,518	$8	$829	$8	$830
5 to 49	453	10,552	$26	$2,460	$35	$3,336
<5	29	64	$4	$63,947	$7	$112,193

1 Data reflect information compiled from several sources, as detailed in Section 3. Includes onshore pipeline incidents from 2010 through June 2016.
2 Costs restated to 2015 dollars based on BLS Consumer Price Index.
3 Costs include monetized value of reported injuries, fatalities, net product lost, and travel delays.

Finally, Exhibit A-5 shows statistics similar to those in Exhibit A-3 and Exhibit A-4, but for the rail incidents.

Exhibit A-5: Rail Crude Oil Spill Quantities and Costs Based on Final Database[1]

Spill Size (Gallons)	Count	Total Quantity Released (Gallon)	Total Costs[2]		Total Costs with Imputed Values[2,3]	
			Total Cost (Million 2015$)	Cost per Gallon (2015$/gal)	Total Cost (Million 2015$)	Cost per Gallon (2015$/gal)
All	439	1,707,470	$56	$33	$59	$34
>100,000	4	1,403,348	$36	$26	$38	$27
50,000 to 99,999	2	148,540	$3	$19	$3	$21
10,000 to 49,999	4	108,765	$10	$92	$10	$95
1,000 to 9,999	9	42,983	$5	$126	$6	$130
500 to 999	2	1,573	$0	$53	$0	$84
100 to 499	6	1,076	$0	$54	$0	$54
50 to 99	3	186	$0	$434	$0	$434
5 to 49	66	626	$0	$344	$0	$344
<5	343	374	$1	$1,734	$1	$3,195

1 Data reflect information compiled from several sources, as detailed in Section 3. Includes rail incidents from 2010 through June 2016.
2 Costs restated to 2015 dollars based on BLS Consumer Price Index.
3 Costs include monetized value of reported injuries, fatalities, net product lost, and travel delays.

Appendix B Incident Cause Consolidation

Exhibit B- 1 provides a crosswalk between the more detailed sub-causes associated with pipeline incidents as pulled into the database and the cause category that we assigned for this analysis. Exhibit B-2 provides a crosswalk between the four categories of rail incidents and the causes reported to PHMSA.

Exhibit B- 1: Pipeline Incident Causes and Categories

Cause Category	Sub-category Cause
CORROSION FAILURE	External Corrosion
	Internal Corrosion
EQUIPMENT FAILURE	Malfunction of Control/Relief Equipment
	Pump or Pump-related Equipment
	Threaded Connection/Coupling Failure
	Non-threaded Connection Failure
	Defective or Loose Tubing or Fitting
	Failure of Equipment Body (except Pump), Tank Plate, or other Material
	Other Equipment Failure
EXCAVATION DAMAGE	Excavation Damage by Operator (First Party)
	Excavation Damage by Operator’s Contractor (Second Party)
	Excavation Damage by Third Party
	Previous Damage due to Excavation Activity
INCORRECT OPERATION	Damage by Operator or Operator’s Contractor NOT Related to Excavation and NOT due to Motorized Vehicle/Equipment Damage
	Tank, Vessel, or Sump/Separator Allowed or Caused to Overfill or Overflow
	Valve Left or Placed in Wrong Position, but NOT Resulting in a Tank, Vessel, or Sump/Separator Overflow or Facility Overpressure
	Pipeline or Equipment Overpressured
	Equipment Not Installed Properly
	Wrong Equipment Specified or Installed
	Other Incorrect Operation
MATERIAL FAILURE OF PIPE OR WELD	Construction-, Installation-, or Fabrication-related
	Original Manufacturing-related (NOT girth weld or other welds formed in the field)
	Environmental Cracking-related
NATURAL FORCE DAMAGE	Earth Movement, NOT due to Heavy Rains/Floods
	Heavy Rains/Floods
	Lightning
	Temperature
	High Winds
	Other Natural Force Damage

Exhibit B- 1: Pipeline Incident Causes and Categories

Cause Category	Sub-category Cause
OTHER OUTSIDE FORCE DAMAGE	Nearby Industrial, Man-made, or Other Fire/Explosion as Primary Cause of Accident
	Damage by Car, Truck, or Other Motorized Vehicle/Equipment NOT Engaged in Excavation
	Damage by Boats, Barges, Drilling Rigs, or Other Maritime Equipment or Vessels Set Adrift or Which Have Otherwise Lost Their Mooring
	Routine or Normal Fishing or Other Maritime Activity NOT Engaged in Excavation
	Electrical Arcing from Otherv [sic] Equipment or Facility
	Previous Mechanical Damage NOT Related to Excavation
	Intentional Damage
	Other Outside Force Damage
OTHER INCIDENT CAUSE	Miscellaneous
	Unknown

Note: The above cause categories and sub-categories reflect possible cause categories available for operators when reporting incidents on PHMSA Accident Reports.

Exhibit B- 2: Rail Incident Causes and Categories

Assigned Cause Category	Original Cause	Original Cause Count
Derailment (19)	Derailment	19
Loose, Missing, or Broken Component (273)	Broken Component or Device	15
	Incorrectly Sized Component or Device	1
	Loose Closure Component or Device	183
	Missing Component or Device	28
	Defective Component or Device	46
Deterioration or Aging (29)	Deterioration or Aging	21
	Inadequate Maintenance	4
	Corrosion - Interior	2
	Threads Worn or Cross Threaded	2
Other Human Error (129)	Valve Open	41
	Human Error	16
	Improper Preparation for Transportation	14
	Inadequate Preparation for Transportation	20
	Misaligned Material Component or Device	25
	Overfilled	12
	Over-pressurized	1

Notes: The original cause count presented here includes the 2 rail incidents with derailment listed as a secondary cause, which have been re-assigned as having derailment as the primary original cause.
A sum of the cause counts does not reflect the total rail incidents, as there are 2 incidents with no causes or alternative causes listed.

Appendix C Narrative Descriptions of Selected Incidents

This appendix provides narrative descriptions of the ten incidents in the database that had the highest total costs, as well as the top ten by volume released and by unit costs.

Appendix C.1 Top Ten Incidents By Total Cost

20100181 Marshall, MI 7/25/2010 *Total Cost Rank: 1* *Quantity Released Rank: 6* *Unit Cost Rank: 335*	An NTSB report (NTSB, 2012) and an NRDA report (USFWS et al., 2015) provide detailed accounts of this incident as well as the operator and public response and natural resource damages resulting from the release. On July 25, 2010 a 30-inch Enbridge Energy pipeline in Calhoun County, Michigan ruptured during a planned shutdown. The rupture and crude oil release was not discovered for over 17 hours, during which time the operator attempted two line startups, causing 843,444 gallons of crude oil to be released into a surrounding wetland, Talmadge Creek, and ultimately Kalamazoo River. Ultimately, 38 miles of the Kalamazoo River were contaminated with oil, including wetlands, floodplain forests, residential properties, farmland, and commercial properties. The affected portion was closed to public access for two years, with some activities (including fishing and swimming) further restricted as late as 2014. Natural resource damages were extensive, including 1,560 acres of in-stream habitat oiled, 2,887 acres of floodplain oiled (with 299 acres having residual oil observed), 185 acres of upland habitat impacted by response actions, and wildlife deaths including 52 birds, 40 mammals, 106 reptiles, and 42 fish. Approximately 100,000 recreational user-days were lost. The day after the initial release, residents of six houses self-evacuated due to the crude oil odors from the spill, and the EPA and county health officials subsequently recommended evacuations of an additional 50 to 61 houses due to high concentrations of benzene in the air. The voluntary evacuation for these residents lasted from July 29 through August 12. There were no fatalities, but 320 people in the vicinity of the release reported some adverse health effects including headaches, nausea, and respiratory problems. In its incident report to PHMSA, the operator reported $23 million in public property damage, $126,118 in commodity losses, $3 million in operator property damage, $635 million in operator environmental remediation, $2.4 million in operator other costs, and $177 million in operator emergency response costs, for a total of $840.5 million ($913.6 million updated to 2015$). Additionally, the NPFC provided $64.9 million in reimbursements to government agencies for public emergency response and environmental remediation costs. Finally, Enbridge's 2014 annual financial report (Enbridge, 2015) states that the total expected costs for the incident are $1.2 billion. As such, we added $294.5 million in additional costs to bring the total cost of the incident to $1.2 billion. After updating to 2015$, the total cost of the incident was $1.3 billion.

20150224 Santa Barbara, CA 5/19/2015 *Total Cost Rank: 2* *Quantity Released Rank: 39* *Unit Cost Rank: 269*	The Refugio State Beach crude oil release (near Santa Barbara, California) of May 19, 2015 involved the release of 123,228 gallons from a 24-inch pipeline owned by Plains All American Pipeline. The company ultimately determined that the incident was caused by external corrosion. Crude oil released from the pipeline reached a culvert that leads to the Pacific Ocean, and as such the release had extensive impacts to the shoreline and nearshore ocean. Natural resource damage assessment is ongoing (California DFW, 2015), but an intermediate report states that 202 birds and 99 mammals were killed, while another 65 and 63, respectively, were oil and captured by responders for rehabilitation. Local, state, and federal agencies continue to assess the full scope of damages from the incident. No fatalities or injuries were reported as a result of the incident. In its report to PHMSA on the incident, Plains Pipeline reported $144,000 in commodity lost, $9.9 million in operator property damage, $22.4 million in operator environmental remediation, $19.8 million in operator other costs, and $90.7 million in operator emergency response costs, for a total of $142.9 million. However, the company's 2015 annual financial report (Plains All American Pipeline, 2016) reports that the total costs of the incident are approximately $269 million. We expect that up to $3 million of this total will be composed of fines and penalties imposed on the company. As such, the total cost of the incident is currently estimated at $266 million ($142.9 million in reported costs plus $123.1 million in additional costs).
20050287 Plaquemines, LA 8/30/2005 *Total Cost Rank: 3* *Quantity Released Rank: 4* *Unit Cost Rank: 996*	This incident involved the release of 991,788 gallons of crude oil on August 30, 2005 in Plaquemines County, Louisiana. The release was a result of extreme winds from Hurricane Katrina, which severely impacted the crude oil terminal of Chevron Pipe Line Company. According to NOAA (2015), the oil was initially contained in an adjacent retaining pond that was designed to catch spilled oil. However, additional adverse weather conditions a few weeks later resulted in the oil being washed into a nearby marsh. Booms and skimmers were used to remove some oil, but 4,000 gallons remained in the marsh after those efforts, damaging a "valuable ecosystem where saltwater from the Gulf of Mexico and freshwater from the Mississippi come together." A team of responders (including personnel from Chevron, NOAA, Coast Guard, USFWS, and Louisiana Department of Wildlife and Fisheries) ultimately decided to carry out *in-situ* burning of the remaining oil, which occurred on October 12 and 13, 2005.[70] In its incident report to PHMSA, Chevron reported $150 million in property damages ($182 million updated to 2015$). No other costs were reported for this incident.

70 For more information about the controlled burn, see Merten and Henry (2008).

20110262 Laurel, MT 7/1/2011 *Total Cost Rank: 4* *Quantity Released Rank: 53* *Unit Cost Rank: 259*	On July 1, 2011, a 12-inch ExxonMobil pipeline ruptured and released 63,378 gallons of crude oil into the Yellowstone River near Laurel, Montana. According to Montana DOJ (2013), the spill occurred during flood conditions and affected the river and its floodplain for 85 miles downstream of the release. Over 1,000 personnel worked on cleanup and shoreline assessment of approximately 11,000 acres along the affected reach, although little of the released oil was ultimately recovered. In its report to PHMSA, ExxonMobil reported \$7.5 million in public property damage, \$150,000 in commodity losses, \$5 million in operator property damage, \$22.4 million in other operator costs, and \$100 million in operator emergency response costs, for a total of \$135 million. Additionally, the NPFC amount for this incident was \$2.8 million (for public response and remediation). As such, the total cost of the incident was \$137.8 million (or \$145.2 million when updated to 2015\$).
20130151 Mayflower, AL 3/29/2013 *Total Cost Rank: 5* *Quantity Released Rank: 37* *Unit Cost Rank: 543*	This release occurred on March 29, 2013 in a residential area in Mayflower, Alabama, in a high consequence area (HCA). It involved the release of 133,980 gallons of crude oil from an underground Mobil Pipe Line Company pipeline. There were 83 people evacuated from the vicinity of the release. Although Mobil's report to PHMSA did not document any injuries from the release, air quality monitoring after the spill showed dangerous levels of benzene, and many people living near the release (but outside the mandatory evacuation zone) reported long-term adverse health impacts including respiratory problems, digestive problems, headaches, and many others (see Eifling, 2013). The costs of the incident reported to PHMSA include \$300,000 in lost commodity, \$900,000 in operator property damage, \$20.1 million in operator environmental remediation costs, \$5.3 million in other operator costs,[71] and \$64.7 million in operator emergency response costs. Additionally, the POLREP for the incident noted \$630,000 in public emergency response costs. As such, the total cost of the incident was approximately \$91.9 million (\$93.5 million updated to 2015\$).

71 According to the incident report, these costs included temporary housing and living expenses for affected residents.

20100221 Romeoville, IL 9/9/2010 *Total Cost Rank: 6* *Quantity Released Rank: 18* *Unit Cost Rank: 866*	This September 9, 2010 incident involved the release of 270,060 gallons of crude oil from an Enbridge Energy pipeline at an industrial park in Romeoville, Illinois. The spill was first reported by the Romeoville Fire Department which reported visible oil on the ground. The release occurred near several small businesses in the afternoon, and emergency response officials evacuated 470 people from the area. Enbridge reported $47.4 million in costs from the incident, including $12.2 million in public property damages, $70,000 in commodity losses, $2.8 million in operator property damages, $16.5 million in operator environmental remediation, $3.6 million in other operator costs, and $12.1 million in operator emergency response costs. Additionally, the NTSB report for the incident documented $550,000 in federal oversight and response costs, for a total of $48 million. Enbridge's 2014 annual financial report (Enbridge 2015) states that the costs for "emergency response, environmental remediation, and cleanup activities" were approximately $51 million (excluding fines and penalties), which is approximately $15.8 million higher than the operator costs for those categories included in the report to PHMSA. As such, we added $15.8 million in additional costs for this incident, for a total of $69.3 million (2015$).
20100146 Salt Lake City, UT 6/12/2010 *Total Cost Rank: 7* *Quantity Released Rank: 86* *Unit Cost Rank: 431*	The Salt Lake City Fire Department notified Chevron Pipeline of a crude oil release from a pipeline on June 12, 2010. In total, 33,600 gallons were released into the Red Butte Creek and subsequently flowed into Liberty Park Lake and the Jorden River. Subsequent examination of the pipeline show that the damage to the pipeline was caused by an electrical arc from a power company. Chevron reported $32.2 million in costs from the incident, including $153,554 in operator property damage, $13.6 million in operator environmental remediation,[72] $18.4 million in operator emergency response costs. Additionally, the NPFC amount was $141,159 (for public emergency response and environmental remediation), and we imputed $1,783 in product loss (since the operator did not include a cost for commodity lost). As such, the total costs of the incident were $35.2 million (2015$).
20100124 Delta Junction, AK 5/25/2010 *Total Cost Rank: 8* *Quantity Released Rank: 42* *Unit Cost Rank: 817*	This accident occurred on May 25, 2010 in Delta Junction, Alaska, and it involved the release of 108,360 gallons. It occurred after the pipeline was shut down for maintenance, and was attributed to incorrect operation. According to a news account (Holland, 2010), a lack of storage capacity at a pump station contributed to the spill. The released oil was contained to a containment area, but the pipeline remained out of service until May 28. The operator, Alyeska Pipeline Service, reported $28.9 million in costs from the incident, including $238,000 in commodity loss, $27 million in operator property damage, and $1.7 million in operator emergency response costs. Updated to 2015$, the total costs of the incident were $31.5 million.

[72] Utah Water Quality Board (2011) reports that Chevron had expended $26.9 million in remediation efforts as of July 31, 2011. However, it is not clear whether all of the costs were associated with incident 20100146, as the settlement also addresses a subsequent December 2010 release by Chevron in the same area.

20120098 New Lenox, IL 3/3/2012 *Total Cost Rank: 9* *Quantity Released Rank: 54* *Unit Cost Rank: 648*	On March 3, 2012 two cars collided with an Enbridge Energy oil pipeline in New Lenox, IL. The collision sheared a connector, causing a release of oil and a subsequent fire. Two people were killed in the incident, and another three were hospitalized with severe burns. The pipeline was shut down as a result of the incident, and 63,000 gallons of crude oil were released. Enbridge Energy reported $2.8 million in costs from the incident, including $124,500 in cost of commodity lost, $864,819 in property damage, $931,946 in environmental remediation, and $915,512 in operator emergency response costs. Additionally, the NPFC amount for public response and remediation was $7,040. Finally, assigning a value of $9.6 million for each of the two fatalities and $3.08 million for each hospitalizations results in an additional $28.4 million from this incident, for a total of $31.4 million (2015$).
X-2015030156 Fayette, WV 2/16/2015 *Total Cost Rank: 10* *Quantity Released Rank: 13* *Unit Cost Rank: 1,444*	On February 16, 2015, a CSX train carrying crude oil derailed in Fayette, West Virginia as a result of a broken rail. During the crash, 362,349 gallons of crude oil were released and there was a subsequent fire. There were no fatalities, but one injury was reported, a house was destroyed, and 300 people were evacuated from the area. Heat from the fire prevented crews from accessing the site for a day after the incident, and the fire continued to burn for several days (CNN Wire Service, 2015). CSX reported costs of $23.5 million, including $5 million in public property damage, $456,560 in commodity lost costs, $3 million in operator property damages, $5 million in operator environmental remediation costs, and $10 million in operator emergency response costs. Additionally, the POLREP for the incident indicates $200,000 in public emergency response and remediation costs. We also assigned a value of $240,000 to the injury reported for the incident, and $192,000 for the value of travel delays. As such, the total cost for the incident was $24.1 million.

Appendix C.2 Top Ten Incidents by Quantity Released

2006071 Cushing, OK 2/20/2006 *Total Cost Rank: 128* *Quantity Released Rank: 1* *Unit Cost Rank: 2,329*	This release occurred as a result of a leak in the bottom of a crude oil storage tank owned by Semcrude, L.P. It entailed the release of 2.06 million gallons of oil. According to the report to PHMSA, all oil was contained inside a dike and was recovered within 72 hours. The total costs of the incident, as reported to PHMSA, were $705,500 ($829,442 updated to 2015$) including $5,500 in commodity lost, $500,000 in operator property damage, and $200,000 in emergency response costs.

20080020 Denver City, TX 1/7/2008 *Total Cost Rank: 33* *Quantity Released Rank: 2* *Unit Cost Rank: 2,156*	On January 7, 2008, 1.32 million gallons of crude oil were released from a ConocoPhillips pipeline near Denver City, Texas. According to the operator's report to PHMSA, the release was caused by seam misalignment in a longitudinal seam. Response time was not reported to PHMSA, but Nalder (2010) and Fehling (2012) state that the oil was spilling from the pipeline for 24 hours before it was detected and shut down. In its report to PHMSA, ConocoPhillips documented $1,000 in public property damage, $2.9 million in commodity loss costs, $3.76 million in operator other costs, and $10,000 in public emergency response and remediation. The total costs of the incident were $6.68 million ($7.36 million in 2015$).
20050279 Plaquemines, LA *Total Cost Rank: 11* *Quantity Released Rank: 3* *Unit Cost Rank: 1,877*	This incident involved the release of 1.07 million gallons of crude oil in Plaquemines Parrish, Louisiana on September 2, 2005 from a Shell Pipeline Company LP pipeline. It was caused by flooding and winds from Hurricane Katrina,[73] which damaged an above-ground storage tank, releasing oil into a tank dike. According to the operator's report to PHMSA, there were several phases of response to the release, including removing oil from the tank dike, transferring the oil, cleaning tanks, demolition of the tanks and piping, and final cleanup. The operator reported costs of $5.44 million in public other costs (specified as tank demolition and site cleanup), $315,575 in cost of commodity lost, $1.55 million in operator property damage, $10.42 million in public emergency response, and $1.05 million in public environmental remediation, for a total cost of $18.77 million ($22.79 million in 2015$).
20050287 Plaquemines, LA 8/30/2005 *Total Cost Rank: 3* *Quantity Released Rank: 4* *Unit Cost Rank: 996*	See description in Appendix C.1.

73 Hurricane Katrina resulted in many large-scale releases of oil from pipelines and terminals, totaling approximately 11 million gallons. See Schleifstein (2010).

20130353 Mountrail County, ND 7/29/2013 *Total Cost Rank: 16* *Quantity Released Rank: 5* *Unit Cost Rank: 1,894*	Sometime between July 29 and August 9, 2013, a lightning strike created a hole in a Tesoro High Plains Pipeline Company pipeline in Mountrail County, North Dakota. The leak was not detected until September 29, 2013 when a farmer notified the company of oil reaching the surface in his wheat field. During the spill, 865,200 gallons of crude oil were released. According to Frosch (2013), the release did not affect wildlife, drinking water sources, or homes, but it affected over seven acres of farmland. The operator reported $16.99 million in costs for the incident, including $168,000 in public property damage, $1.42 million in commodity losses, $269,051 in operator property damage, $11.03 million in operator environmental remediation, and $4.1 million in operator emergency response. Total costs are $17.28 million when updated to 2015$.
20100181 Marshall, MI 7/25/2010 *Total Cost Rank: 1* *Quantity Released Rank: 6* *Unit Cost Rank: 335*	See description in Appendix C.1.
20060353 Borger, TX 12/3/2006 *Total Cost Rank: 383* *Quantity Released Rank: 7* *Unit Cost Rank: 2,338*	This release occurred at a ConocoPhillips Pipe Line Co crude oil terminal in Borger, Texas when a mixer became detached from the side of a tank into a containment dike on December 3, 2006. 630,000 gallons were released during the incident, but it was all contained within the dike. The operator reported total costs of $120,000 for the incident ($141,082 in 2015$), including $35,000 in cost of commodity lost, $5,000 in operator property damage, and $80,000 in cleanup and disposal.
20110210 Chico, TX 6/4/2011 *Total Cost Rank: 88* *Quantity Released Rank: 8* *Unit Cost Rank: 2,235*	This incident involved the release of 513,618 gallons of crude oil from an Enterprise Crude Oil Pipeline LLC tank in Chico, Texas on June 4, 2011.According to the incident narrative provided in the PHMSA report, the release was detected when the control center sent a tech to the tank to investigate crude oil levels that were falling faster than expected. The tech found that a tank mixer had failed and released crude oil into a dike, where all released oil was contained. In its report to PHMSA, the operator included $706,140 in commodity loss costs, $1,000 in operator property damage, $633,290 in operator environmental remediation, and $20,000 in operator emergency response costs, for a total cost of $1.36 million ($1.43 million in 2015$).

X-2014010238 Casselton, ND 12/30/2013 *Total Cost Rank: 66* *Quantity Released Rank: 9* *Unit Cost Rank: 2,163*	This incident involved the derailment of a BNSF Railway Company train carrying crude oil in Casselton, North Dakota, on December 30, 2013.[74] A westbound BNSF grain train derailed in front of the approaching eastbound BNSF crude oil train. When the crude oil train struck a car from the derailed grain train at a speed of 42 miles per hour, the lead locomotive and 20 rail cars derailed and ignited. 18 of the cars carrying crude oil were breached. 474,936 gallons of crude oil were released, and the operator reported that the incident resulted in the closure of a major artery for 61 hours. 1,500 people were evacuated from Casselton due to concerns about drifting smoke and fumes. The operator reported $1.31 million in public property damages, $718,100 in cost of commodity lost, $200,000 in operator property damage, $20,000 in operator environmental remediation, and $40,000 in operator emergency response costs, for a total of $2.29 million ($2.33 million in 2015$). Additionally, we added costs of $122,000 for the travel disruptions caused by the incident. As such, total costs were $2.45 million (2015$).
E-2013120116 Aliceville, AL 11/7/2013 *Total Cost Rank: 40* *Quantity Released Rank: 10* *Unit Cost Rank: 2,016*	On November 7, 2013, an Alabama & Gulf Coast Railway LLC train carrying crude oil derailed at 39 miles per hour in Aliceville, Alabama, releasing 455,520 gallons. The incident did not cause any fatalities, injuries, or evacuations, but it resulted in a fire and the closure of a major artery for 236 hours. In its report to PHMSA, the operator included $1 million in public property damages, $1 million in commodity losses, $1 million in operator property damages, and $1 million in operator environmental remediation, for total reported costs of $5 million. Additionally, the NPFC amount was $153,017 for public response and remediation, and we added $472,000 for the value of travel disruptions from the artery closure. As such, the total costs of the incident were $5.71 million (in 2015$).

Appendix C.3 Top Ten Incidents by Unit Cost

X-2009110204 Lodi, OH 10/28/2009 *Total Cost Rank:124* *Quantity Released Rank:2,340* *Unit Cost Rank: 1*	On October 28, 2009, a CSX train that included a loaded tank car of crude oil derailed in Lodi, OH. During the derailment, the crude oil tank car rolled 90 degrees. No oil releases were initially detected by the responders on-scene, but subsequent investigation revealed that approximately one cup of oil was leaked. The operator reported costs of $760,000 in operator property damage and $6,000 in emergency response costs for a total of $766,000 ($846,265 in 2015$). Based on a reported quantity released of 0.5 gallons, this equates to unit costs of $1.7 million.

74 The information in this summary is primarily derived from the incident report to PHMSA; see NTSB (2014) for additional details.

20150340 Jones Creek, TX 9/1/2015 *Total Cost Rank:168* *Quantity Released Rank:2,365* *Unit Cost Rank:2*	This incident was the result of a lightning strike on an Enterprise Crude Pipeline LLC crude oil tank in Jones Creek, Texas, which caused a fire. The operator reported a release of 0.42 gallons of crude oil. Emergency responders extinguished the fire, and subsequent inspections revealed the need to repair 13 plates and the damaged seal on the tank. After repairs were completed, the tank was returned to service on July 21, 2016. The operator reported $450 in lost commodity costs, $250,000 in operator property damage, and $278,708 in emergency response costs, for a total cost of $529,158. Based on a quantity of 0.42 gallons, unit costs from this release are $1.3 million.
20150448 Cohasset, MN 11/9/2015 *Total Cost Rank:230* *Quantity Released Rank:2,365* *Unit Cost Rank:3*	On November 9, 2015, a contractor discovered a crude oil leak from an Enbridge pipeline near Cohasset, Minnesota. The reported quantity released was 0.42 gallons, but Enbridge determined that the repair costs to address the leak (including replacement of the o-ring, bonnet studs, and nuts) exceeded the NRC reporting threshold. The operator reported $337,771 in operator property damage and $2,000 in emergency response costs for a total cost of $339,771. Unit costs are $809,979 (based on a quantity released of 0.42 gallons).
20150169 TX 2/19/2013 *Total Cost Rank:62* *Quantity Released Rank:2,101* *Unit Cost Rank:4*	This incident occurred during the planned replacement of 50 feet of a West Texas Gulf Pipeline Company pipeline in Freestone County, Texas. During the work on the pipeline, proper sealing procedures (using mud packs) were not followed by workers, which caused crude oil vapor to be discharged in the direction of welding activities, resulting in flash ignition of the vapor. 4.2 gallons of crude oil were released during the incident. The operator did not report any costs from the release, but there was one hospitalization, which we valued at $3.08 million, in addition to $10 in value of product lost. Based on a release of 4.2 gallons, this results in unit costs of $733,336.
X-2014050147 Vicksburg, MS 5/6/2014 *Total Cost Rank:1,608* *Quantity Released Rank:2,485* *Unit Cost Rank: 5*	During a routine FRA inspection in Vicksburg, Mississippi, a small quantity of released oil was discovered from a Kansas City Southern Railway Company (KSC) railcar. When the leak was detected, hazmat responders were dispatched by KCS to the scene. Ultimately, responders determined that the leak resulted from a packing gland nut that was not sufficiently tightened and involved less than 0.1 gallons. The operator reported emergency response costs of $5,000. Given a very small quantity released, the unit costs of the incident were $642,000.
X-2014050348 Vicksburg, MS 5/8/2014 *Total Cost Rank:1,608* *Quantity Released Rank:2,485* *Unit Cost Rank:5*	During a routine FRA inspection in Vicksburg, Mississippi, a small quantity of released oil was discovered from a Kansas City Southern Railway Company (KSC) railcar. When the leak was detected, hazmat responders were dispatched by KCS to the scene. Ultimately, responders determined that the leak resulted from a nozzle flange connection that was not sufficiently tightened and involved less than 0.1 gallons. The operator reported emergency response costs of $5,000. Given a very small quantity released, the unit costs of the incident were $642,000.

X-2014020204 Battle Creek, MI 1/18/2014 *Total Cost Rank:183* *Quantity Released Rank:2,216* *Unit Cost Rank:7*	During switching operations at the Battle Creek Yard in Battle Creek, Michigan, Illinois Central Railroad Company employees reported an odor of crude oil from a tank car. The company deployed a response team, which noted crude oil leaking from man-way bolts and valve plugs that were not sufficiently tightened. The operator reported emergency response costs of $10,736 plus $5 in cost of commodity lost, for a total of $10,741 ($10,754 in 2015$). Additionally, there were two non-hospitalization injuries resulting from the incident, which we valued at $480,000. Given a quantity released of one gallon, this results in unit costs of $490,754.
20150459 Jones Creek, TX 11/5/2015 *Total Cost Rank:317* *Quantity Released Rank:2,365* *Unit Cost Rank:8*	On November 5, 2015, Enterprise Crude Pipeline LLC personnel conducting maintenance on a crude oil pipeline in Jones Creek, Texas noticed a small quantity of crude oil leaking near a nipple fitting. Subsequent metallurgic analysis concluded that the leak was caused by fatigue cracking, and involved less than one cup of oil released. The operator reported $196,960 in operator property damages. Given a reported release quantity of 0.42 gallons, this results in unit costs of $468,952.
20130309 Longview, TX 8/10/2013 *Total Cost Rank:29* *Quantity Released Rank:1,698* *Unit Cost Rank:9*	On August 10, 2013, contractors were working to replace an ExxonMobil pipeline in Longview, Texas. The work involved cutting out a section of pipeline, displacing the line with a pig, and installing blinds to abandon the pipeline. The contractors successfully welded the flanges on one end of the pipeline, but when they began welding the other end, a flash fire occurred. 20 gallons of crude oil were released and overflowed the catch pan. According to the operator, three of the contractor employees have filed suit against ExxonMobil as a result of the incident. The operator reported $1,000 in environmental remediation expenditures, $3,800 in emergency response costs, and $50 in commodity losses. Additionally, there were three hospitalizations as a result of the incident, which we valued at $9.24 million. As such, total costs of the incident were $9. 24 million (in 2015$). Given a release of 20 gallons, this yields unit costs of $458,578.
20110457 Coalinga, CA 12/1/2011 *Total Cost Rank:219* *Quantity Released Rank:2,325* *Unit Cost Rank:10*	On December 1, 2011, a Shell operating station lost power, shutting down pumps as well as communications with control centers. When power was turned back on approximately 4 hours later, there was a fire which was later attributed to the release of crude oil and crude oil vapors from a pump seal failure. The operator reported that the release volume was 0.84 gallons of crude oil. The operator reported $340,000 in operator property damage, and we imputed $2 in the value of product lost. Updated to 2015$, the total cost of the incident was $358,258, for unit costs of $426,498.

Appendix D Supplementary Regression Model Specifications

This appendix provides the results of several regression model specifications in addition to those presented in Sections 5.2.3 and 5.2.2.

Appendix D.1 Alternative Pipeline Model with Full Incident Set

The main pipeline regression model described in Section 5.2.2 is based on pipeline incidents that occurred between January 2010 and June 2016, since those are the incidents with the most complete data available. Exhibit D-1 shows the results of that regression using the full pipeline dataset, including incidents between 2005 and 2009.

As shown in the exhibit, this model has different slope terms for the two pipeline datasets. The reason for splitting these variables is to allow the costs across the datasets to have different relationships with quantity released. The intercept is similar to the intercept in the main pipeline model (8.113, see Exhibit Exhibit 5-8), but is comparatively lower for older pipeline incidents, which would have a 1 value for the "transp_mode_pipe09" variable, effectively shifting the intercept down by 1.687.

The rest of the variables in the regression match those from the main model shown in Exhibit 5-8, and the coefficients for the variables are generally similar in direction and magnitude.

Exhibit D-1: Regression-based Analysis of Total Costs (Full Pipeline Data Set)

Variable	Regression Specification
intercept	8.155*** (0.260)
transp_mode_pipe09	-1.687** (0.578)
ln_qreleased_pipe09	0.333* (0.155)
ln_qreleased_pipe10	-0.106 (0.0890)
ln_qreleased_sq_pipe09	0.0134 (0.0106)
ln_qreleased_sq_pipe10	0.0486*** (0.00713)
pop800	0.000173*** (0.0000378)
watwet800_bi	0.349*** (0.0749)
spatial_missing	-0.484* (0.243)
et800_bi	0.679** (0.224)
lu_hddev_800	0.777 (0.427)
hca_bi	0.513*** (0.0947)
fire_bi	2.281*** (0.540)
underground_bi	1.080*** (0.0971)
pcause_corrosion	0.118 (0.0888)
pcause_natlforces	0.745** (0.226)
Observations	2,026
R^2	0.505
Root MSE	1.645

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Appendix D.2 Kitchen Sink Models

Exhibit D- 2 shows the results of the main regression models with additional explanatory variables, including an indicator for soil contamination, train speed, and an indicator for incidents occurring during winter months. We developed these additional models as an additional robustness check on the primary regression results; if the addition of more variables significantly impacts the direction and magnitude of the coefficients that are included in the primary regression model, it is an indicator that there may be some omitted variable bias driving the results.

The coefficients that are common to the models shown in Exhibit 5-8 and Exhibit D- 2 have the same direction and generally the same magnitude, indicating that omitted variable bias is not a significant concern in our regression analysis.

Exhibit D- 2: Regression-based Analysis of Total Costs (Kitchen-Sink Models)

Variable	Regression Specification	
	Rail	Pipeline
intercept	7.620*** (0.676)	8.042*** (0.262)
ln_qreleased	0.138*** (0.0283)	-0.170 (0.0893)
ln_qreleased_sq	0.0152** (0.00564)	0.0487*** (0.00716)
pop800	0.000225 (0.000332)	0.000141** (0.0000431)
watwet800_bi	1.012 (0.640)	0.258** (0.0901)
spatial_missing	-0.797 (0.472)	-1.101*** (0.146)
et800_bi	0.794 (0.666)	0.760* (0.362)
lu_hddev_800		0.926 (0.551)
hca_bi		0.529*** (0.105)
fire_bi		2.777*** (0.464)
underground_bi		1.046*** (0.119)
soilcont_bi		0.776*** (0.105)
winter	0.0358 (0.145)	0.162 (0.0967)
train_speed	0.0267 (0.0147)	
rcause_derail	1.492** (0.555)	
rcause_component	-0.275* (0.120)	
rcause_aging	-0.466** (0.152)	
pcause_corrosion		0.0327 (0.120)
pcause_natlforces		0.643* (0.267)
Observations	315	1,250
R^2	0.766	0.519
Root MSE	0.853	1.569

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Appendix D.3 Incident Cause Models

As described in Section 3.5.2, we consolidated the reported incident causes into 4 categories for rail incidents, and 8 for pipeline incidents. Based on the literature described in Section 2.4, we expect that some incident causes may result in systematically different incident costs. To identify these causes, we conducted regression analyses predicting total costs based on the cause, allowing both the intercept and the slope (i.e., the relationship between the total cost and the quantity released) to vary by cause. Exhibit D-3 summarizes the results of this analysis.

For pipeline incidents, the model excludes a binary for "other" causes, meaning that when the other seven cause binary variable values are set to zero, the model defaults to this "other" cause. The rail model excludes a binary for incidents caused by error; as such, if the binary variables for the other three causes are zero, the model defaults to the error cause. Relative to the error cause category, derailments result in substantially higher total costs, while the component and aging causes result in relatively lower costs. The aging cause results in the lowest costs overall, and has the lowest slope coefficient (in other words, as the quantity released increases, the total costs increase slowest for incidents caused by aging, and fastest for derailments).

Exhibit D-3: Regression-based Analysis of Total Costs (Cause Models)

Variable	Regression Specification	
	Rail	**Pipeline**
intercept	7.895*** (0.111)	9.402*** (0.772)
rcause_derail	3.166*** (0.815)	
rcause_component	-0.278* (0.124)	
rcause_aging	-0.482** (0.154)	
pcause_corrosion		-1.628* (0.812)
pcause_equipment		-2.110** (0.799)
pcause_excavation		-3.226** (1.108)
pcause_incorrectopp		-1.369 (0.893)
pcause_mwfail		0.237 (0.942)
pcause_natlforces		-0.189 (1.066)
pcause_outforce		-1.388 (1.584)
ln_qr_rcause_derail	0.288*** (0.0788)	
ln_qr_rcause_component	0.231*** (0.0434)	
ln_qr_rcause_aging	0.0795 (0.0653)	
ln_qr_rcause_error	0.239*** (0.0626)	
ln_qr_pcause_corrosion		0.522*** (0.0419)
ln_qr_pcause_equipment		0.443*** (0.0395)
ln_qr_pcause_excavation		0.613*** (0.0959)
ln_qr_pcause_incorrectopp		0.370*** (0.0845)
ln_qr_pcause_mwfail		0.439*** (0.0885)
ln_qr_pcause_natlforces		0.394** (0.126)
ln_qr_pcause_outforce		0.629*** (0.173)
ln_qr_pcause_other		0.268* (0.128)
Observations	315	1,250
R^2	0.733	0.406
Root MSE	0.904	1.744

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Appendix D.4 Combined Quadratic Log-Log Model

Exhibit D-4 shows the results of the regression analysis based on a pooled dataset of rail and pipeline releases (including the pipeline incidents since 2010).

Exhibit D-4: Regression-based Analysis of Total Costs (Combined)

Variable	Regression Specification
intercept_pipe	8.176*** (0.261)
intercept_rail	8.069*** (0.444)
ln_qreleased_pipe	-1.096*** (0.0889)
ln_qreleased_rail	0.161*** (0.0304)
ln_qreleased_sq_pipe	0.0478*** (0.00699)
ln_qreleased_sq_rail	0.00664 (0.00646)
pop800	0.000173*** (0.0000463)
watwet800_bi	0.301** (0.0917)
spatial_missing	-0.483 (0.422)
et800_bi	0.623 (0.336)
lu_hddev_800	0.635 (0.560)
hca_bi	0.494*** (0.106)
fire_bi	2.410*** (0.421)
underground_bi	1.142*** (0.124)
rcause_derail	1.971** (0.633)
rcause_component	-0.293* (0.119)
rcause_aging	-0.499** (0.154)
pcause_corrosion	0.0348 (0.122)
pcause_natlforces	0.684* (0.277)
Observations	1,565
R^2	0.587
Root MSE	1.491

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Appendix D.5 Unit Cost Model

The dependent variable in the regressions presented in Section 5.2.2 is the log of total cost. An alternative modeling approach would be to use the log of unit cost as the dependent variable. However, because of the properties of logarithms, this alternative approach is mathematically equivalent to our main regression specification, and would produce the same coefficients on all variables (with one minor exception).

To see this, write the unit model as:

$$\log\left(\frac{Total\ Cost_i}{Quantity_i}\right) = f(SpillChar_i, SiteChar_i, CostMethod_i)$$

The properties of logarithms imply that:

$$\log(Total\ Cost_i) - \log(Quantity_i) = f(SpillChar_i, SiteChar_i, CostMethod_i)$$

And so:

$$\log(Total\ Cost_i) = f(SpillChar_i, SiteChar_i, CostMethod_i) + \log(Quantity_i)$$

This implies that the coefficients from our main model are identical to those that would be estimated by a model with log unit cost as the dependent variable. Converting from log unit costs to log total costs simply requires adding the log of spill quantity to the right-hand side of the equation. The intuition is that adding the logarithm of quantity to the right-hand side (which is otherwise a model of log unit cost) is mathematically identical to multiplying the unit cost model by quantity (to calculate modeled total cost) and then taking a logarithm. Either way, the result is a model of log total cost.

Although this conversion involves some mathematical details, the main point is that it makes no difference whether we estimate regressions using log unit cost or log total cost as the dependent variable. Either specification can be used to derive the other.

To illustrate, Exhibit D-5 shows the combined quadratic log-log model (Exhibit D-4), except with unit costs as the dependent variable rather than total costs. All of the coefficients except the quantity variable coefficients are identical to the total cost model. Applying the unit cost model to an incident and multiplying the predicted costs by the quantity released would yield the same total costs as would be predicted by the total cost model for the same incident.

Exhibit D-5: Regression-based Analysis of Unit Costs (Combined)

Variable	Regression Specification
intercept_pipe	8.176*** (0.261)
intercept_rail	8.069*** (0.444)
ln_qreleased_pipe	-1.0956*** (0.0889)
ln_qreleased_rail	-0.839*** (0.0304)
ln_qreleased_sq_pipe	0.0477*** (0.00716)
ln_qreleased_sq_rail	0.00664 (0.00646)
pop800	0.000173*** (0.0000463)
watwet800_bi	0.301** (0.0917)
spatial_missing	-0.483 (0.422)
et800_bi	0.623 (0.336)
lu_hddev_800	0.635 (0.560)
hca_bi	0.494*** (0.106)
fire_bi	2.410*** (0.421)
underground_bi	1.142*** (0.124)
rcause_derail	1.971** (0.633)
rcause_component	-0.293* (0.119)
rcause_aging	-0.499** (0.154)
pcause_corrosion	0.0348 (0.122)
pcause_natlforces	0.684* (0.277)
Observations	1,565
R^2	0.610
Root MSE	1.491

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown below in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Appendix D.6 Alternative Functional Forms

The quadratic log-log functional form (which includes both the log of quantity and the log of quantity squared as independent variables) are the best fit for the rail and pipeline incident costs in the database, as shown in Exhibit 5-10. However, we also explored alternative functional forms for the models, including a linear form and a “binned” form. These alternative functional form models are the same as those presented in Exhibit 5-8, except that the quantity variables take on different forms.

First, we ran the models using only the log of quantity released to vary the slope, excluding the quadratic term. These results are shown in Exhibit D-6.

Exhibit D-6: Regression-based Analysis of Total Costs (Linear Log-Log)

Variable	Regression Specification	
	Rail	Pipeline
intercept	7.937*** (0.652)	6.760*** (0.150)
ln_qreleased	0.222*** (0.0309)	0.475*** (0.0232)
pop800	0.0000357 (0.000321)	0.000167*** (0.0000486)
watwet800_bi	1.066 (0.569)	0.345*** (0.0940)
spatial_missing	-1.141* (0.505)	-0.661*** (0.113)
et800_bi		0.946* (0.456)
lu_hddev_800		0.728 (0.569)
hca_bi		0.495*** (0.110)
fire_bi		3.410*** (0.454)
underground_bi		1.241*** (0.130)
rcause_derail	2.838*** (0.478)	
rcause_component	-0.264* (0.121)	
rcause_aging	-0.470** (0.160)	
pcause_corrosion		-0.119 (0.126)
pcause_natlforces		0.663* (0.294)
Observations	315	1,250
R^2	0.747	0.471
Root MSE	0.879	1.643

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Next, we also conducted the regression using a "binned" specification, in which total costs are modeled independently for a number of ranges of gallons spilled for each transportation mode, as shown in Exhibit D-7. This is a flexible way of modeling the effect of quantity released on costs, since it allows the data to determine the overall functional form between these two variables. The R^2 for these models are slightly less than the R^2 for the preferred models in Section 5.2.2. The model does allow flexibility in the relationship between costs and different ranges of gallons spilled. However, *within* each range of gallons spilled, the model holds unit costs constant—thus reducing the explanatory power of the model.

Exhibit D-7: Regression-based Analysis of Total Costs, Binned Functional Form

Variable	Regression Specification	
	Rail	**Pipeline**
intercept		14.42*** (0.488)
pop800	0.000204 (0.000366)	0.000155*** (0.0000466)
watwet800_bi	1.086 (0.666)	0.245** (0.0937)
spatial_missing	-0.880 (0.461)	-0.991*** (0.225)
et800_bi		0.688* (0.344)
lu_hddev_800		0.768 (0.577)
hca_bi		0.477*** (0.109)
fire_bi		2.730*** (0.485)
underground_bi		1.190*** (0.126)
rcause_derail	2.319*** (0.584)	
rcause_component	-0.221 (0.118)	
rcause_aging	-0.422** (0.147)	
pcause_corrosion		0.0137 (0.124)
pcause_natlforces		0.701* (0.284)
ln_qreleased_binA (0.001 to 0.009 gal)	8.312*** (0.687)	
ln_qreleased_binB (0.01 to 0.09 gal)	6.955*** (0.713)	
Ln_qreleased_binC (0.1 to 0.9 gal)	7.492*** (0.702)	-5.219*** (0.757)
Ln_qreleased_binD (1 to 9 gal)	7.685*** (0.699)	-6.465*** (0.529)
Ln_qreleased_binE (10 to 99 gal)	8.154*** (0.697)	-6.048*** (0.487)
Ln_qreleased_binF (100 to 999 gal)	9.142*** (0.780)	-4.949*** (0.489)
Ln_qreleased_binG (1,000 to 9,999 gal)	9.733*** (0.840)	-4.078*** (0.495)
Ln_qreleased_binH (10,000 to 99,999 gal)	10.56*** (1.031)	-2.462*** (0.521)
Ln_qreleased_binI (100,000 to 999,999 gal)	12.27*** (0.975)	
Observations	315	1,250
R^2	0.761	0.481
Root MSE	0.865	1.631

*Note: Each column represents a different regression model, where the dependent variable is the log of total cost. Each cell in the table shows a regression coefficient, with the corresponding standard errors shown in parentheses. All models are estimated using ordinary least squares, with Huber-White heteroskedasticity-robust standard errors. Observations with costs of $0 are excluded from the regressions. Significance is denoted by stars: * p<0.05, ** p<0.01, *** p<0.001.*

Appendix E Correlations between Variables

Exhibit E-1 provides correlations between the various explanatory variables used in the regressions. Variables are listed in the same order across the top and sides of the table. The value in each cell is a measure of how closely the variable in that column and row co-vary. A value of 1 indicates that two variables are perfectly positively correlated; a value of 0 indicates no correlation; and a value of -1 indicates a perfect negative correlation.

For example, *transp_mode_pipe09* has a correlation of 1 with itself, indicating (as expected) that it is perfectly correlated with itself. However, *transp_mode_pipe09* and *transp_mode_pipe10* have a correlation of -0.63, reflecting the fact that when one of these two dummy variables takes the value 1, the other always takes the value 0. (They are not perfectly negatively correlated because both take the value 0 when *transp_mode_rail* takes the value 1.)

Exhibit E-1: Correlation of Regression Variables

Variable Name	transp_mode_pipe09	transp_mode_pipe10	transp_mode_rail	ln_qreleased_pipe09	ln qreleased _pipe10	ln_ qreleased _rail	hca_bi	fire_bi	pop800	spatial_missing	underground_bi	et800_bi
transp_mode_pipe09	1											
transp_mode_pipe10	-0.63	1										
transp_mode_rail	-0.22	-0.61	1									
ln_qreleased_pipe09	0.95	-0.60	-0.21	1								
ln qreleased _pipe10	-0.49	0.78	-0.48	-0.47	1							
ln_ qreleased _rail	-0.06	-0.17	0.28	-0.06	-0.14	1						
hca_bi	0.06	0.17	-0.28	0.06	0.14	-0.08	1					
fire_bi	-0.04	0.00	0.04	-0.03	-0.06	0.32	0.00	1				
pop800	0.05	-0.03	-0.01	0.05	-0.02	0.00	-0.01	0.00	1			
spatial_missing	0.01	0.05	-0.07	0.00	0.07	0.01	0.23	0.00	0.00	1		
underground_bi	-0.12	-0.61	0.9	-0.12	-0.48	0.1	-0.27	-0.06	-0.01	-0.10	1	
et800_bi	0.21	0.11	-0.36	0.19	0.22	-0.1	0.15	-0.05	0.03	0.11	-0.32	1
watwet800_bi	0.07	-0.02	-0.04	0.06	0.01	0.06	0.13	0.05	0.00	0.04	-0.06	0.05
lu_hddev_800	-0.05	-0.21	0.31	-0.06	-0.18	0.09	0.00	0.01	0.02	-0.13	0.32	-0.12
people_evacuated	-0.03	0.13	-0.13	-0.02	0.07	-0.01	0.35	-0.01	-0.01	0.41	-0.17	0.06
soilcont_bi	-0.01	-0.04	0.07	-0.01	-0.01	0.32	0.00	0.25	0.00	0.01	-0.02	-0.01
surfwater_bi	0.36	0.12	-0.52	0.36	0.32	-0.11	0.12	-0.02	0.02	0.09	-0.47	0.40
train_speed	-0.22	-0.61	0.99	-0.21	-0.48	0.23	-0.28	0.00	-0.01	-0.07	0.91	-0.35
winter	0.08	0.00	-0.08	0.08	0.13	0.09	0.09	0.07	-0.01	0.03	-0.11	0.18
response_t~s	-0.05	-0.13	0.21	-0.05	-0.10	0.66	-0.06	0.35	0.00	0.02	0.01	-0.08
artery_clo~s	0.02	0.03	-0.06	0.02	0.04	0.01	0.00	-0.01	0.04	0.02	-0.05	0.02
rcause_der~l	-0.02	0.03	-0.02	-0.02	0.01	-0.01	0.01	-0.01	0.00	-0.01	-0.02	0.02
pcause_cor~n	-0.03	-0.08	0.13	-0.03	-0.06	0.51	-0.04	0.33	0.00	0.00	-0.02	-0.05
pcause_nat~s	-0.05	-0.14	0.22	-0.05	-0.11	0.66	-0.06	0.37	0.00	0.02	0.01	-0.08

Exhibit E-1: Correlation of Regression Variables (cont'd)

Variable Name	watwet800_bi	lu_hddev_800	people_evacuated	soilcont_bi	surfwater_bi	train_speed	winter	response_t~s	artery_clo~s	rcause_der~l	pcause_cor~n	pcause_nat~s
watwet800_bi	1											
lu_hddev_800	0.01	1										
people_evacuated	0.00	-0.02	1									
soilcont_bi	0.08	0.03	0.01	1								
surfwater_bi	0.07	-0.16	0.04	-0.02	1							
train_speed	-0.06	0.31	-0.13	0.05	-0.53	1						
winter	0.10	0.09	-0.01	0.08	0.17	-0.11	1					
response_t~s	0.07	0.07	-0.02	0.37	-0.07	0.16	0.08	1				
artery_clo~s	-0.03	-0.02	-0.01	0.04	0.05	-0.06	0.06	0.02	1			
rcause_der~l	-0.01	0.02	0.00	0.00	0.00	-0.02	0.05	0.00	-0.02	1		
pcause_cor~n	0.18	0.04	-0.02	0.27	0.00	0.04	0.13	0.57	0.00	0.00	1	
pcause_nat~s	0.07	0.07	-0.02	0.32	-0.08	0.17	0.1	0.88	0.01	0.00	0.55	1

Note: The only variable combination with a true zero correlation is fire × transp_mode_10. All the other zeros in the table represent correlation smaller than 0.01.

Appendix F Incident Frequency

As noted in Section 2.2, the primary purpose of this study is to determine the relationship between the total costs of an incident and the multitude of factors that may affect the magnitude of these costs. Incident probability is an important consideration in analysis of rulemakings regarding the transportation of crude oil, and is likely to be affected by many of the same factors. However, this study focuses primarily on the costs of an incident, conditional on its occurrence.

The regression model described in this report (see Section 5) can be used to estimate the expected costs of crude oil spill incidents and the benefits of improving the safety of crude oil transportation. A key input for this analysis is the set of projected incidents and therefore model application rests in part on the ability to project incident probabilities and characteristics.

This appendix presents a summary of available information and data on the pipeline and rail networks that transport crude oil, and the frequency of incidents involving the release of crude oil.

Appendix F.1 Pipeline Incidents

A variety of studies and reports estimate pipeline spill probabilities using empirical data; they do so generally either as a review of changes to a pipeline network safety over time, or to support a proposed method for improving maintenance efforts (Dawotola et al., 2012; Little, 2009; Hovey and Farmer, 1993; Hill and Director, 1993). For these analyses, spill probability is most frequently presented in the format of number of incidents per miles (or kilometers) per year.

Girgin and Krausmann (2016) present a slightly different probability analysis, focusing their empirical results on spills from crude oil pipelines that were caused by natural hazards. They find that between 2004 and 2012 in their supplemented PHMSA dataset, natural hazard spills occurred at a rate of 0.053 incidents per year per thousand kilometers for crude oil pipelines.

It is worth noting that these measures of probability do not account for the volume of each incident. Generally this is because these metrics are estimated to support network operators seeking to reduce the frequency of any type of spill, although for some authors volumetric data may not be available. When volume of spills is accounted for in spill probability, it can be done by characterizing separately the probability of different spill size thresholds.

In order to generate probability estimates, it is necessary to combine incident data with data on the universe, or network of crude pipelines and volume that they transport. The universe estimates are based on data from PHMSA, reflecting data reported by hazardous liquid pipeline system operators annually in Form PHMSA F 7000-1.1.[75]

[75] Downloaded on 10/19/2016 from: http://www.phmsa.dot.gov/pipeline/library/data-stats/annual-report-mileage-for-hazardous-liquid-or-carbon-dioxide-systems

In 2015, pipeline operators reported a total of 67,305 miles of onshore pipelines,[76] for pipelines that are primarily crude oil pipelines, or that had been used to transport crude oil that year. These same pipeline systems reported a total throughput for the year of 2.4 trillion barrel-miles of crude oil. Pipeline operators also reported a total of 23,226 miles of onshore pipeline transporting crude oil that could affect designated high consequence areas (HCAs).

Based on the available incident data described in Section 3, in 2015 there were an average of 0.0037 crude oil spill incidents per mile per year in the U.S.[77] Applying available data in the incident and universe datasets, we estimated two other types of incident probability measurements:

- Based on annual throughput, or volume transported, there was one incident per 10 billion barrel-miles transported in 2015.[78] This was equivalent to 8 barrels spilled per 100 million barrel-miles transported.
- In high consequence areas in 2015, there were 0.0036 HCA incidents per year per HCA mile of pipeline.

It is possible to derive other incident frequency metrics for subcategories of pipelines of interest to a given policy analysis using the incident data and Annual Reports.

Appendix F.2 Rail Incidents

While a broad literature base exists for estimating pipeline spills for crude oil, relatively few studies consider the probability of crude oil spills from rail transportation. One recent draft study by Environmental Research Consulting (ERC) for the state of Washington models the probability of future crude oil rail spills in Puget Sound based on the possible construction of a new refinery (ERC, 2016). The model seeks to estimate crude spill probability relative to throughput, defined as probability of a spill per million train-miles transported. It estimates that an 8.6 percent increase in crude-by-rail traffic would correspond with an increase in the annual spill frequency of any volume from crude rail transport between 0.0055 spills and 0.046 spills per year (ERC, 2016).

In a national analysis of the OSRP rule, PHMSA (2016a) considers crude rail derailment probability between 2004 and 2015 for the purpose of projecting the future frequency of spills from derailment. The report discusses spill probability estimates in terms of incidents per year and incidents per volume transported. Using Waybill Sample data for carloads transported, the report finds 12 derailments in 1.738

---

[76] This estimate of number of miles of onshore pipelines transporting crude oil in the U.S. in 2015 falls in between the 56,375 miles of crude pipeline in 2014 estimated by the Bureau of Transportation Statistics (2015), and the 73,082 miles of crude oil pipeline in 2015 estimated by PHMSA (2016b). This second value from PHMSA includes both onshore and offshore mile of crude oil pipelines.

[77] Calculated as 253 incidents divided by 67,305 miles of crude oil pipeline. Alternatively presented as 3.7 spills per thousand miles per year. For reference, these estimates are within the order of magnitude of comparable estimates for different pipeline systems in different time periods (Dawotola, 2012; Little, 2009; Hovey and Farmer, 1993).

[78] Although this seems quite infrequent, this estimate doesn't account for the size of the spill incidents. By comparison, the Association of Oil Pipe Lines and the American Petroleum Institute (2014) estimated that 9.3 billion barrels of crude were transported through onshore and offshore pipelines in the U.S. in 2014.

million carloads in 2014 and 2015,[79] for a derailment rate of 0.0069 per thousand carloads (PHMSA, 2016a).

Based on carload data provided in PHMSA (2016a) and rail incidents in the database, Exhibit F-1 shows that the number of reported releases of crude oil from rail cars has increased along with the volume transported during the years 2005 to 2014, followed by a decline in both trends in 2015. The number of spills declined more sharply in 2015 compared with the decline in the number of carloads transported. The Association of American Railroads (AAR, 2016b) highlights the implementation of regulations and voluntary safety improvements by the industry in 2014 and 2015, but it is probably too early to tell whether these improvements were factors in the sharp decline in the number of incidents in 2015 (and so far in 2016). Furthermore, the trend is not uniformly pointing to improving safety and the quantity of crude oil released from railcars does not follow the same decline in 2015. As shown in panel B of Exhibit F-1, the quantity released from rail cars has been much more variable than the volume transported or the number of incidents, and sharply *increased* in 2015 relative to 2014 (and that 2014 is the anomalous year).

Overall, rail incident trends seem to be primarily driven by large-magnitude incidents. Out of 42 rail incidents in 2015, there were four that involved the release of at least 20,000 gallons (ranging between 26,449 and 362,349 gallons).[80] Comparatively, 2014 had a higher rate of incidents, but only two incidents involving more than 20,000 gallons (one at 29,868 gallons and the other at 50,450).[81] In 2013, there were two incidents that involved the release of over 400,000 gallons (455,520 gallons and 474,936 gallons; the rest of the 2013 incidents involved 10,000 gallons or less).[82]

[79] Additionally, we have verified the OSRP Waybill Sample data, using comparable estimates by the Association of American Railroads (AAR), which collects data on the number of originated and terminated carloads of crude oil on U.S. class I railroads in its AAR Freight Commodity Statistics. Although the full dataset is not publicly available, data from 2008 through preliminary 2015 estimates are provided in its annual report and industry presentations (AAR, 2015; 2016a). The report and dataset also include values for the number of carloads of crude oil terminated on U.S. Class 1 railroads. These numbers are not necessarily identical, because some shipments along U.S. Class 1 railroads begin or end in Canada or on U.S. short line railroads. However, these two estimates track very closely with each other and with the Waybill Sample data, all within 5% variability across years.

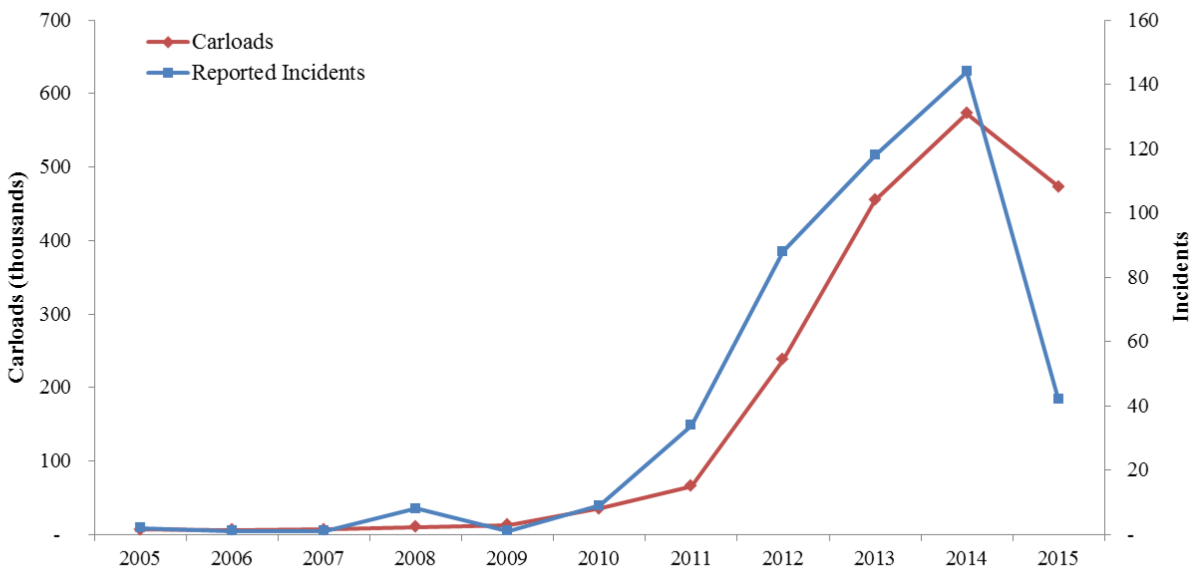
[80] X-2015080186 in Culberton, MT (26,449 gallons), X-2015050385 in Heimdal, ND (98,090 gallons), X-2015040049 in Galena, IL (110,543 gallons), and X-2015030156 in Mount Carbon, WV (362,349 gallons)

[81] X-2014050225 in Lynchburg, VA (29,868 gallons) and X-2014020292 in New Augusta, MS (50,450 gallons)

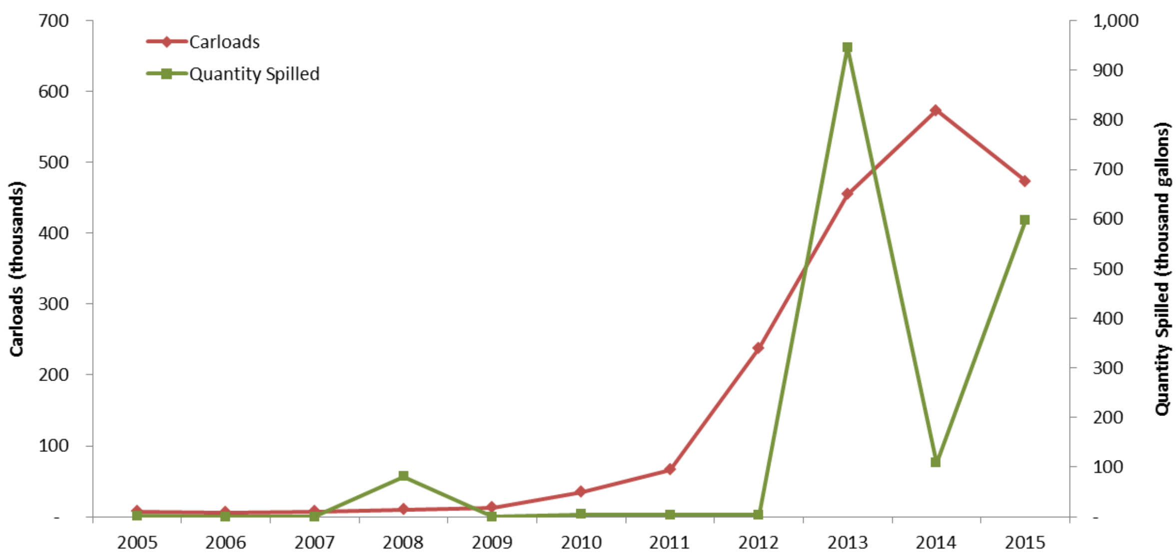
[82] E-2013120116 in Aliceville, AL (455,520 gallons) and X-2014010238 in Casselton, ND (474,936 gallons)

Exhibit F-1: Relationship of Carloads of Crude Oil Transported and Number of Crude Oil Releases and Quantity of Crude Oil Released from Railcars, 2005-2015

(A) Relationship of Carloads Transported and Number of Reported Incidents



(B) Relationship of Carloads Transported and Quantity of Crude Oil Released from Railcars



Source: Based on carload data from PHMSA (2016a) and number of included rail incidents in the database.
Note: carload data are based on Surface Transportation Board Waybill data for 2005 through 2013; PHMSA (2016a) derived data for 2014 and 2015 based on U.S. Energy Information Administration (U.S. EIA) production forecasts in millions of barrels per year and then converted to carloads. As such, the carload data for the last two years of data are projections and not actual data; the carloads may be over- or underestimated in those years. Recent data from U.S. EIA (2016a) suggest that the projected trends for 2014 and 2015 are reasonable; the barrels transported increased from 296.1 million in 2013 to 382.0 million in 2014 (an increase of 29%), then decreased to 318.8 million in 2015 (a decline of 17%).

U.S. EIA (2016a) provides the total crude oil volume shipped by rail between January 2010 and June 2016. Exhibit F-2 summarizes the number of rail incidents and volume released per million gallons transported via rail, based on rail shipment data and the incidents included in the database assembled for this study (see Section 3 for details). The table shows significant variation in spill rate across the years; the ratio of gallons spilled per million gallons transported range from less than 1 in 2012, to nearly 1,000 in 2013.

Exhibit F-2: Rail Incident Rate Based on Volume Transported

Year	Gallons Transported (Millions)	Incident Count	Incidents per Million Gallons Transported	Gallons Spilled (Thousands)	Gallons Spilled per Million Gallons Transported
2010	999	9	0.009	4.9	4.92
2011	1,780	34	0.019	3.9	2.21
2012	6,386	88	0.014	3.8	0.59
2013	12,435	118	0.009	945.5	76.04
2014	16,045	144	0.009	108.5	6.76
2015	13,389	42	0.003	598.4	44.69
2016	3,994	4	0.001	42.5	10.63
Total	**55,007**	**439**	**0.008**	**1,707**	**31.03**

Source: Gallons transported based on U.S. EIA (2016a). Data for 2016 cover the period of 1/1/2016 through 6/30/2016.

Appendix G Quantitative Damage Information

Exhibit G-1 summarizes available quantitative information about damages to natural habitats and wildlife from crude oil releases.

Exhibit G-1: Quantitative Habitat Impacts from Crude Oil Releases

Incident Number	Incident Name	Miles Oiled	Acres of Habitat Oiled					Wildlife Impacted	Wildlife Killed					Wildlife Oiled (and Rehabilitated)					Recreational User Days Lost	Reference
			Water Body	Wetland	Upland	In-Stream	Floodplain		Fish	Birds	Mammals	Reptiles	Amphibians	Fish	Birds	Mammals	Reptiles	Amphibians		
20100181	Marshall	38			185	1,560	2,887		42	52	40	106							100,000	USFWS et al. (2015)
20110262	Yellowstone	85																		Montana DOJ (2013)
20150224	Refugio Beach									202	99				65	63				CA DFW (2015)
20080272	Marathon pipeline			7	14			✓												Marathon Pipeline LLC (2011)
20140152	Mid-Valley	1						57												U.S. EPA Region V (2014)
20100221	Romeoville											32					141			NTSB (2013)
20130151	Mayflower			✓						✓										U.S. EPA (2015)
20110410	Centurion		11				160													U.S. EPA Region VI (2011)
20130130	Lion Oil Magnolia	4.5	✓																	U.S. EPA Region VI (2013)
20140202	Delek	1							8	1		3					3	4		U.S. EPA Region VI (2014a)
20140394	Mid Valley	4.1						490												U.S. EPA Region VI (2014b)
20150048	Bridger	59																		U.S. EPA Region VIII (2015a)
20150277	Plains	8	578																	U.S. EPA Region V (2015a)

Exhibit G-1: Quantitative Habitat Impacts from Crude Oil Releases

Incident Number	Incident Name	Miles Oiled	Acres of Habitat Oiled					Wildlife Impacted	Wildlife Killed					Wildlife Oiled (and Rehabilitated)					Recreational User Days Lost	Reference
			Water Body	Wetland	Upland	In-Stream	Floodplain		Fish	Birds	Mammals	Reptiles	Amphibians	Fish	Birds	Mammals	Reptiles	Amphibians		
20050073	Kentucky River	10																		Schreiner (2005)
X-2015040049	BNSF Galena	1																		U.S. EPA Region V (2015b)
X-2015050385	Heimdal			0.25																U.S. EPA Region VIII (2015b)