

Supplementary Information

Carbon dioxide pipelines are disproportionately located in marginalized communities in the United States

Julia A. Davis¹, Nafiseh Salehi¹, Liqing Li², Majid Shafiee-Jood^{1,*}

¹ Department of Civil and Environmental Engineering, University of Virginia, Charlottesville, Virginia, 22904, USA

² Department of Agricultural Economics, Texas A&M University, College Station, Texas, 77843, USA

* Corresponding author. Email: ms2dm@virginia.edu

Table of Contents

Supplementary Notes.....	3
1. LASSO Variable Selection for Binary Logistic Regression Models	3
2. Statistical Significance of Selected Demographic Variables.....	5
3. Estimations of Pipeline Accident Frequency Rates Along Future Pipeline Networks.....	5
4. Environmental Impact of CO ₂ Pipeline Accident.....	6
5. Location of CO ₂ Emissions from Pipeline Accidents.....	7
6. Description of CO ₂ Pipeline Accident Costs.....	7
7. CO ₂ Pipeline Accidents and Impact Breakdown by Cause Type.....	8
Supplementary Figures.....	12
Supplementary Tables.....	17
SI References.....	23

Supplementary Notes

S1. LASSO Variable Selection for Binary Logistic Regression Models

We use binary logistic regression models to assess the correlations between the presence of an existing, proposed, or 2030 projected CO₂ pipeline in a census tract, and the following demographics of that tract. These variables are measured from a five-year estimate from the American Community Survey between 2017-2021 and include: population, population density, urban/rural binary indicator, percentage of high-income households, percentage of occupied housing units that are owned, percentage of housing units that are vacant, and percentage of the population in high poverty, in low poverty, is white, is Black, is American Indian/Native American, is Asian, is Native Hawaiian/Pacific Islander, is defined as other race, has completed no schooling, has completed standard high school as their highest education level, has completed standard high school, or received a GED or alternative credential, as their highest education level, has some college experience as their highest education level, has an Associate's or Bachelor's degree as their highest education level, has a Bachelor's degree as their highest education level, and has Master's, professional school, or Doctorate degree as their highest education level (1). The selection of these covariates is first due to the theory that certain population density, housing, economic, race, and education related demographics are potentially correlated with being spatially close to pipeline infrastructure, based on prior literature. For example, natural gas pipelines were found to more likely be proposed in areas with lower-educated, economically poor, and non-Black and non-Hispanic residents (2). Existing natural gas pipelines were found to be located in areas with a greater Social Vulnerability Index (3).

The covariates used in each pipeline regression model are also chosen based on the Least Absolute Shrinkage and Selection Operator, or LASSO, regression. LASSO incorporates L1 regularization to enhance the model's predictive accuracy and interpretability. By adding penalty equivalent to the absolute value of the magnitude of coefficients to the loss function, LASSO effectively shrinks some coefficients to zero, thereby performing variable selection. The objective of LASSO is to minimize the following function:

$$\min_{\beta} \left(\frac{1}{2n} \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \varphi \sum_{j=1}^p |\beta_j| \right) \quad [\text{Eq. S1}]$$

Where:

- x_i equals the predictor variables for the i^{th} observation.
- y_i equals the response variable for the i^{th} observation.
- β is the vector of coefficients.
- n is the number of observations.
- p is the number of predictor variables.
- φ is the regularization parameter that dictates the strength of the penalty applied to the coefficients.

The L1 penalty is the second term, represented by:

$$\varphi \sum_{j=1}^p |\beta_j| \quad [\text{Eq. S2}]$$

The LASSO constrains the sum of the absolute values of the model parameters (α_z , β_z , and γ_z in Eq. S3, Eq. S4, and Eq. S5, respectively), excluding the intercept, α_0 , β_0 , and γ_0 (again, in Eq. S3, Eq. S4, and Eq. S5, respectively) (4). This method is particularly advantageous in high-dimensional datasets where multicollinearity is present, as it not only mitigates overfitting but also simplifies the model by excluding non-contributory predictors, leading to a more parsimonious representation of the underlying data structure.

The LASSO regression signifies certain variables do not improve the fits for each of the three pipeline network regression models, which can be viewed in Supplementary Fig. 1. Therefore, these variables are not included in the regression models of the existing, proposed, and 2030 pipeline regression models. Specifically, Within the existing CO₂ pipeline network, in descending order of relevance, the important variables include population density,

homeownership rates, urban indicator, high income rates, share of highest education level of a master's/professional school/doctorate degree, white race, low poverty rates, share of highest education level of college experience, Asian race, vacancy rates, Native American/American Indian race, share of highest education level of an associate's or bachelor's degree, Black race, and share of highest education level of a standard high school or a GED alternative credential. We find that the share of highest education level of no formal schooling, share of highest education level of high school education, share of highest education level of a bachelor's degree, high poverty rates, other races, and Native Hawaiian/Pacific Islander race are not important variables to improve the existing network model's balance of complexity and predictability. In the proposed CO₂ pipeline network, in descending order of relevance, the important variables include population density, share of highest education level of high school education, share of highest education level of an associate's or bachelor's degree, other races, Black race, share of highest education level of college experience, homeownership rates, Asian race, share of highest education level of a bachelor's degree, Native American/American Indian race, high poverty rates, high income rates, Native Hawaiian/Pacific Islander race, share of highest education level of no formal schooling, and vacancy rates. We find that the share of highest education level of a standard high school or a GED alternative credential, share of highest education level of a master's/professional school/doctorate degree, urban indicator, low poverty rates, and white race are not important variables in improving the proposed network model's predictability. Lastly, for the 2030 projected CO₂ pipeline network, in descending order of relevance, the important variables include population density, urban indicator, high income rates, Black race, share of highest education level of a bachelor's degree, vacancy rates, share of highest education level of college experience, Native Hawaiian/Pacific Islander race, share of highest education level of high school education, homeownership rates, Native American/American Indian race, low poverty rates, high poverty rates, and Asian race. Conversely, white race, share of highest education level of a master's/professional school/doctorate degree, share of highest education level of a standard high school or a GED alternative credential, share of highest education level of an associate's or bachelor's degree, other races, and the share of highest education level of no formal schooling are not important variables to select for the 2030 projected regression model.

As indicated, some variables are found to be unimportant in one model and not the others; for example, the variable indicating low poverty rates is deemed to be important for the existing and 2030 projected pipeline regression models, but not important for the proposed pipeline regression model. Thus, we include this variable in the existing and projected regression models but exclude it from the proposed regression model. Eq. S3, Eq. S4, and Eq. S5 indicate the equations with the important variables for the existing, proposed, and 2030 projected regression models, respectively.

$$\begin{aligned}
 \text{Existing Pipeline}_i = & \alpha_0 + \alpha_1(\text{Population Density}_i) + \alpha_2(\text{Homeownership Rate}_i) + \\
 & \alpha_3(\text{Urban Indicator}_i) + \alpha_4(\text{High Income Household Rate}_i) + \\
 & \alpha_5(\text{Master, Professional, or Doctorate Degree Rate}_i) + \alpha_6(\text{White Race Rate}_i) + \\
 & \alpha_7(\text{Low Poverty Rate}_i) + \alpha_8(\text{Some College Experience Rate}_i) + \alpha_9(\text{Asian Race Rate}_i) + \\
 & \alpha_{10}(\text{Vacancy Rate}_i) + \alpha_{11}(\text{American Indian Race Rate}_i) + \\
 & \alpha_{12}(\text{Associate or Bachelor Degree Rate}_i) + \alpha_{13}(\text{Black Race Rate}_i) + \\
 & \alpha_{14}(\text{GED or High School Degree Rate}_i) + \delta_1 \text{County}_1 + \dots + \delta_k \text{County}_k
 \end{aligned}
 \tag{Eq. S3}$$

$$\begin{aligned}
 \text{Proposed Pipeline}_i = & \beta_0 + \beta_1(\text{Population Density}_i) + \beta_2(\text{High School Degree Rate}_i) + \\
 & \beta_3(\text{Associate or Bachelor Degree}_i) + \beta_4(\text{Other Race Rate}_i) + \beta_5(\text{Black Race Rate}_i) + \\
 & \beta_6(\text{Some College Experience Rate}_i) + \beta_7(\text{Homeownership Rate}_i) + \\
 & \beta_8(\text{Asian Race Rate}_i) + \beta_9(\text{Bachelor Degree Rate}_i) + \beta_{10}(\text{American Indian Race Rate}_i) + \\
 & \beta_{11}(\text{High Poverty Rate}_i) + \beta_{12}(\text{High Income Household Rate}_i) + \\
 & \beta_{13}(\text{Native Hawaiian Race Rate}_i) + \beta_{14}(\text{No Educational Experience Rate}_i) + \\
 & \beta_{15}(\text{Vacancy Rate}_i) + \vartheta_1 \text{County}_1 + \dots + \vartheta_k \text{County}_k
 \end{aligned}
 \tag{Eq. S4}$$

$$\begin{aligned}
\text{Projected Pipeline}_i = & \gamma_0 + \gamma_1(\text{Population Density}_i) + \gamma_2(\text{Urban Indicator}_i) + \\
& \gamma_3(\text{High Income Household Rate}_i) + \gamma_4(\text{Black Race Rate}_i) + \gamma_5(\text{Bachelor Degree Rate}_i) + \\
& \gamma_6(\text{Vacancy Rate}_i) + \gamma_7(\text{Some College Experience Rate}_i) + \\
& \gamma_8(\text{Native Hawaiian Race Rate}_i) + \gamma_9(\text{High School Degree Rate}_i) + \\
& \gamma_{10}(\text{Homeownership Rate}_i) + \gamma_{11}(\text{American Indian Race Rate}_i) + \\
& \gamma_{12}(\text{Low Poverty Rate}_i) + \gamma_{13}(\text{High Poverty Rate}_i) + \gamma_{14}(\text{Asian Race Rate}_i) + \\
& \theta_1 \text{County}_1 + \dots + \theta_k \text{County}_k
\end{aligned}
\tag{Eq. S5}$$

An in-depth, interactive ArcGIS Web Map displaying all of the most significant demographic variables with respect to each pipeline network can be viewed here: <https://arcg.is/1LS8Sj2>.

S2. Statistical Significance of Selected Demographic Variables

While many of the variables in each of the CO₂ pipeline network regression models indicate a significant association with the presence of a pipeline, some do not. Our existing pipeline regression model shows that, on average, there is not a statistically significant relationship (at the 95% level or above) between the presence of an existing CO₂ pipeline in a census tract, and that tract's share of highest education level of a master's/professional school/doctorate degree, white race, low poverty rates, share of highest education level of college experience, Asian race, vacancy rates, Native American/American Indian race, share of highest education level of an associate's or bachelor's degree, Black race, and share of highest education level of a standard high school or a GED alternative credential.

Within the proposed pipeline regression model, we do not find a statistically significant relationship (at the 95% level or above) between the presence of a proposed CO₂ pipeline in a census tract, and that tract's share of Asian race, share of highest education level of a bachelor's degree, Native American/American Indian race, high poverty rates, high income rates, Native Hawaiian/Pacific Islander race, share of highest education level of no formal schooling, and vacancy rates.

The 2030 projected pipeline regression model indicates that there is not a statistically significant relationship (at the 95% level or above) between the presence of a projected CO₂ pipeline in a census tract, and that tract's share of highest education level of a bachelor's degree, vacancy rates, Native Hawaiian/Pacific Islander race, homeownership rates, Native American/American Indian race, low poverty rates, high poverty rates, and Asian race.

S3. Estimations of Pipeline Accident Frequency Rates Along Future Pipeline Networks

We utilize the prior CO₂ pipeline accident data from PHMSA (5) in order to not only understand the current accident rates but also estimate the possible future number of accidents along the proposed and 2030 projected pipeline networks, should these routes be constructed and operate under similar considerations and conditions to the existing infrastructure. To do so, we first test the correlation between the location of a pipeline accident and that area's demographic makeup to understand if there would be a significant association between a demographic variable and the likelihood and frequency of an accident. As the results show in Supplementary Table S4, we find that very few variables correlate with our accident outcomes of interest. This indicates that CO₂ pipeline accidents occur along pipeline routes generally independent of nearby demographic makeup; accidents are usually caused by failures related to the pipeline's physical properties, construction, or maintenance, which may fail random to the area's demographic composition. Therefore, we decide to estimate the number of potential future accidents by extrapolating the rate of accidents per mile between 1994 and 2023 along the existing CO₂ pipeline network onto the proposed and projected pipeline mileages of the two future

networks. We find that, within the existing network, given 121 accidents over 30 years and 5,300 cumulative pipeline miles, there are about 0.001 accidents per mile per year, equating to about 0.023 accidents per mile and about 4 accidents per year, given our available data.

$$\frac{121 \text{ accidents}}{5,300 \text{ miles}} = 0.0228 \text{ accidents per mile over 30 years}$$

We apply these rates to the expected mileage from the proposed and 2030 projected pipeline routes, which are about 3,580 and 11,800 miles, respectively. From our calculations, we estimate that, assuming completed networks in construction and operation, the proposed CO₂ pipeline route will experience about 2.7 accidents each year, or 82 accidents over a 30-year period.

$$0.0228 \text{ accidents per mile} \times 3,580 \text{ cumulative proposed miles} = 81.732 \text{ accidents over 30 years}$$

$$\frac{81.732 \text{ accidents}}{30 \text{ years}} = 2.7 \text{ accidents per year along proposed network}$$

The projected pipeline route is expected to have about 9 accidents per year, or about 269 accidents over a 30-year period.

$$0.0228 \text{ accidents per mile} \times 11,800 \text{ cumulative projected miles} = 269.396 \text{ accidents over 30 years}$$

$$\frac{269.396 \text{ accidents}}{30 \text{ years}} = 9 \text{ accidents per year along projected network}$$

We choose to present our estimations in both rates per year and total numbers over a specific time period to provide a clear comparison in rates and frequencies between our existing data and our expected accident data along future pipelines. Our estimations assume that these time periods are similar 1994-2023 in relation to the CO₂ pipelines, and that the pipelines are constructed, operated, and maintained with similar materials and processes to the pipelines already in operation. With more robust pipeline infrastructure and stricter operation and maintenance regulations, it is possible that these estimates may decrease as it becomes less likely to have an accident.

S4. Environmental Impact of CO₂ Pipeline Accident

Pipeline accidents have, and will likely continue to have, significant impacts on the areas' environmental wellbeing. We find that areas that have experienced pipeline accidents have been exposed to a total of over 17.4 million gallons of CO₂ emissions between 1994 and 2023, averaging to over 143,000 gallons of CO₂ per accident. While 4 accidents did not release any CO₂, over half of the accidents released at least 1,000 gallons of CO₂, and the largest release from an accident was almost 3 million gallons. Considering how much CO₂ emissions would additionally be released from accidents along the future routes, we find that over 11.7 million and 38.7 million gallons of CO₂ emissions would likely be released from proposed and projected pipeline accidents, respectively, equating to about 301,000 and 1.3 million additional gallons of CO₂ emissions from the accidents each year.

S5. Location of CO₂ Emissions from Pipeline Accidents

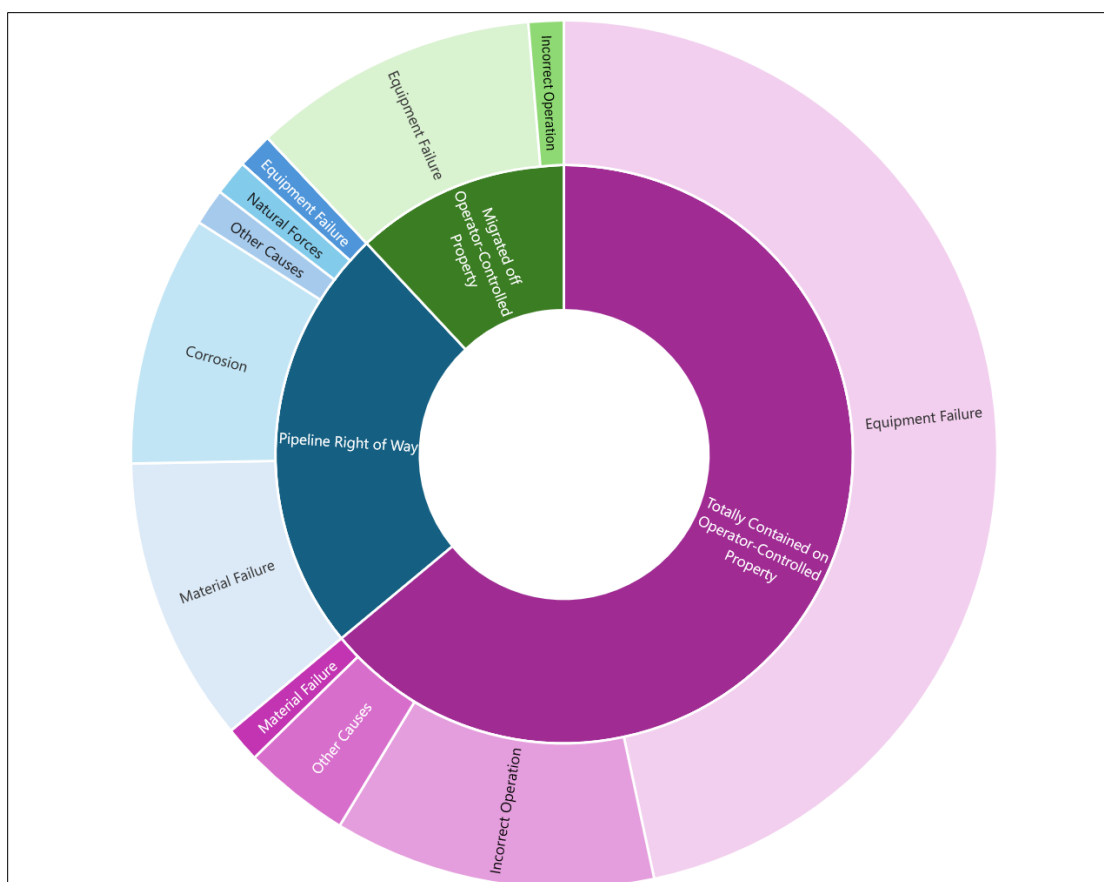
When an accident occurs, the released CO₂ emissions can either be located on the pipeline's right of way, totally contained on the Operator-controlled property, or flow/migrate off the Operator-controlled property. Generally, when CO₂ emissions released from accidents are located farther from the source point of accident, there is greater opportunity for these emissions to affect the surrounding communities, making these accidents potentially more dangerous. Between 2010 and 2023, PHMSA recorded the locations of the 75 accidents – in 18 accidents (24%), the released CO₂ emissions stayed along the pipeline right of way; in 48 accidents (64%), the released emissions were totally contained on the Operator-controlled property; in 9 accidents

(12%), the CO₂ emissions originated on the Operator-controlled property but then flowed/migrated off the property (Supplementary Fig. S1). In consideration of the total estimated proposed and projected pipeline accidents, about 7 and 23 accidents, respectively, will have their released CO₂ emissions migrate off the Operator-controlled properties of the pipelines over the course of a 14-year period, or about 0.5 and 1.6 accidents between the two networks each year.

S6. Description of CO₂ Pipeline Accident Costs

We find that there are considerable costs resulting from pipeline accidents. Across the 121 accidents between 1994 and 2023, the CO₂ pipeline accidents have an average associated cost of about \$60,100 per accident, totaling to about \$7.3 million in damage. The most expensive accident cost about \$4 million in total damages. We find that the estimated 82 proposed and 269 projected pipeline accidents would cause an estimated additional \$4.9 million and \$16.2 million, respectively, in total costs over a 30-year period, averaging about \$164,000 and \$540,000 in total costs per year.

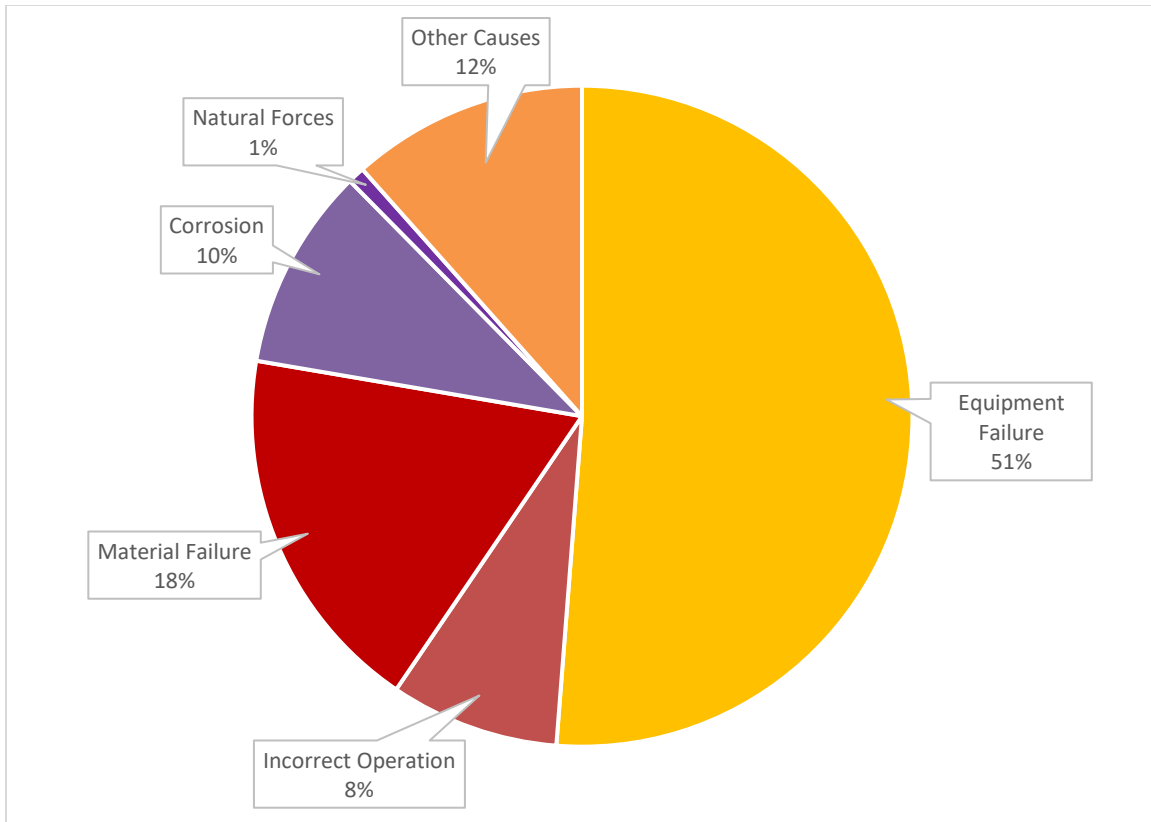
Since 2002, total cost calculations from accidents along existing CO₂ pipeline routes comprise of an accident's estimated costs of public and non-Operator private property damage, commodity lost, Operator's property damage and repairs, emergency response, environmental remediation, and other costs. In addition to estimating the total costs of previous accidents, we further investigate the estimated costs of public and non-Operator private property damage, emergency response, and environmental remediation, as these factors closely relate to how nearby individuals in a community can be economically impacted. The 111 accidents between 2002-2023 cost an estimated average of about \$2,200 per accident in public and non-Operator private property damage, totaling over \$240,000 in damage to non-Operator property. Therefore, given the estimated additional proposed and projected pipeline accidents, we find that accidents along the proposed pipeline network may cause an additional about \$163,000 in public and non-Operator private property damage over a similar 22-year period (\$7,400 per year); projected pipeline accidents are estimated to cause an additional about \$540,000 in public and non-Operator private property damage (\$24,500 per year). Emergency response to the prior accidents also proved to be costly: each accident is estimated to have caused over \$2,800 in expenses for emergency response. In total, about \$316,000 have been spent on emergency responses to the 111 accidents between 2002 and 2023, averaging to over \$14,000 a year. We estimate that emergency response would cost about \$214,000 and \$704,000 to address the expected accidents along the proposed and projected networks, respectively, or an additional about \$9,700 and \$32,000 for each pipeline network per year. Surprisingly, only two of the 111 previous accidents required environmental remediation with associated costs. These instances of remediation cost a total of \$3,000; thus, on average, it is expected that a CO₂ pipeline accident will cause damages of about \$27 in environmental remediation, an insignificant average expense compared to the non-Operator property damage and emergency response related costs. Still, we estimate that environmental remediation in response to pipeline accidents would cost about \$2,025 from proposed pipeline accidents and about \$6,680 from projected pipeline accidents in total over 22 years, or an additional \$92 and \$304 per year from each pipeline network. Accidents caused by natural forces made up almost all – over 93.38% – of the costs associated with in public and non-Operator private property damages between 2002 and 2023, and over 64.96% of the costs associated with emergency response. The two accidents that had environmental remediation-related expenses were caused by material failure, making up 100% of the \$3,000 of environmental remediation expensed.



Supplementary Fig. S1: Shares of CO₂ pipeline accidents whose emissions were located on the pipeline right of way, totally contained on the Operator-controlled property, and that migrated off the Operator-Controlled property of the pipeline. These shares are further divided by the cause of each accident. The green section indicates the proportion of pipeline accidents whose emissions were located on the pipeline right of way. The purple portions of the chart represent the proportion of pipeline accidents whose emissions were totally contained on the Operator-controlled property. The blue part of the chart represents the proportion of pipeline accidents whose emissions migrated off the Operator-Controlled property of the pipeline.

S7. CO₂ Pipeline Accidents and Impact Breakdown by Cause Type

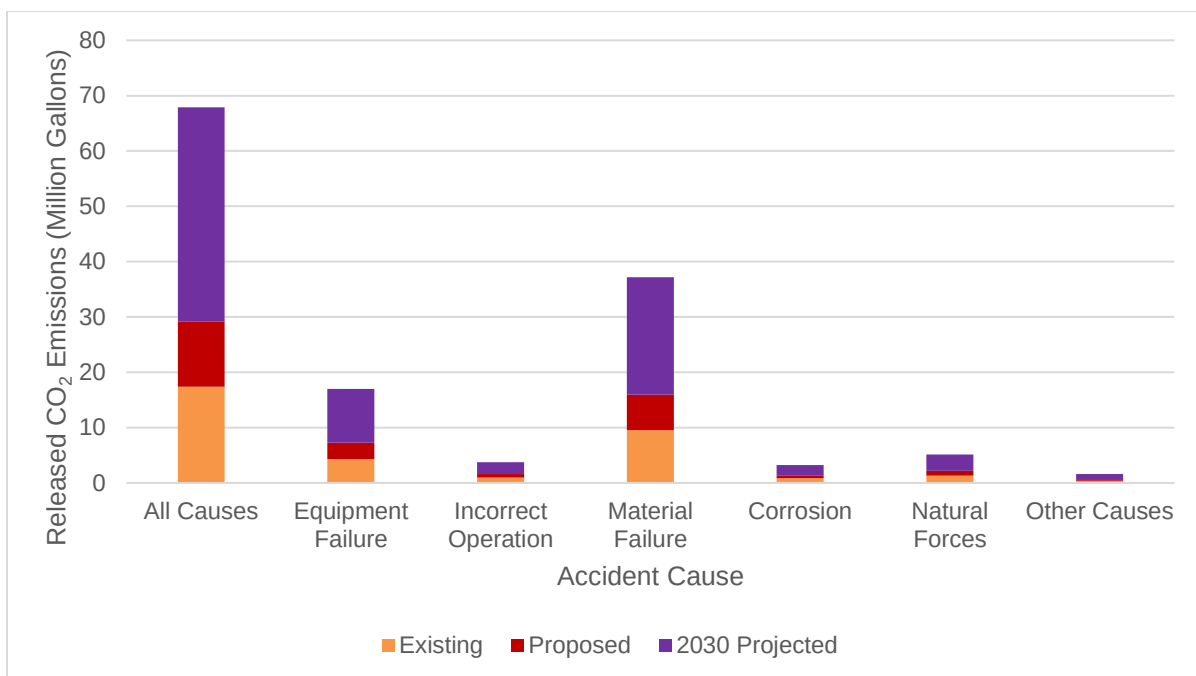
CO₂ pipeline accidents have been caused by one of six general reasons: equipment failure, incorrect operation, material failure, corrosion, natural forces, or other reasons (Supplementary Fig. S2) (5). The first four causes directly relate to the physical properties, construction, and operation and maintenance of pipelines, which can be influenced through state and federal CO₂ pipeline regulations. We find that, of the 121 accidents, equipment failure is the most common – 62 accidents, over half (51.24%) of all accidents, were caused by equipment failure. Natural force damage is least likely to cause an accident (0.83%), with only one accident being caused by heavy rain and flooding; although it should be noted that, because this is an external cause, it is plausible for this rate to change among accidents from the proposed and projected pipeline networks, as they are located in different geographical regions where there may be varying frequency and severity of extreme weather/natural events. Material failure, incorrect operation, and corrosion caused about 18.18%, 8.26%, and 9.91% of accidents, respectively. Other causes of pipeline accidents occurred about 11.57% of the time.



Supplementary Fig. S2: The proportion of the total number of accidents that were due to each accident cause type. This ratio is the same among the existing, proposed, and 2030 projected pipeline networks, as the estimations of the accidents from the proposed and projected pipeline networks are derived from the existing pipeline accident rate information.

To estimate the number of accidents caused by each type (equipment failure, incorrect operation, material failure, corrosion, natural forces, or other reasons) within the proposed and projected pipeline networks, we again applied the proportion of each cause type accident within the existing pipeline infrastructure to the future routes. We then were able to calculate how much environmental and economic risk and damage resulted from each cause type of accident by applying the rates of the current accident data to the estimated number of proposed and projected pipeline accidents.

Notably, the different causes of pipeline accidents have different related rates of released CO₂ emissions per accident. For example, while material failure is only likely to cause an accident less than one-fifth of the time, material failure-caused accidents released the highest rates of CO₂, responsible for over half (54.74%) of the total CO₂ emissions released from all accidents (Supplementary Fig. S3). In other words, the 22 material failure-caused accidents released a total of over 9.5 million gallons of CO₂. Other causes, on the other hand, caused just over 11% of accidents, but these accidents released only about 2.40% of the total CO₂ emissions (Supplementary Fig. S3). We estimate that over 90% of emissions released from pipeline accidents also come from equipment failure, incorrect operation, material failure, and corrosion caused accidents. About 87.69% of the additional proposed and projected pipeline accidents are estimated to be caused by equipment failure, incorrect operation, material failure, or corrosion, and these accidents are estimated to contribute over 90.02% of the additional gallons of CO₂ emissions.



Supplementary Fig. S3: Amount of CO₂ emissions released from accidents along the existing, proposed, and 2030 projected pipeline networks over a 30-year period, and the breakdown of the released CO₂ amounts among each pipeline network, dependent on the cause of the accident. The green bars represent the millions of gallons of CO₂ emissions released in total from existing pipeline accidents, for each cause of the accidents. The red and the purple show the same information for the estimated accidents along the proposed and projected pipeline networks, respectively.

In consideration of the location of CO₂ emissions, we find that accidents resulting from the most common cause – equipment failure – are the most likely type of accidents to have their released CO₂ emissions be totally contained on the Operator-controlled property, and the most likely type to have their emissions flow/migrate off the Operator-controlled property; however, CO₂ emissions from accidents caused by equipment failure are the least likely to be located directly on the pipeline right of way. Specifically, 35 of the 48 accidents (72.92%) where the CO₂ emissions were totally contained on the Operator-controlled property, and 8 of the 9 accidents (88.89%) where the emissions migrated off the Operator-controlled property, were caused by equipment failures (Supplementary Fig. S1). The other accident where the released CO₂ emissions migrate off the Operator-controlled property was caused by incorrect operation, again suggesting the positive impact that thorough equipment and operation related pipeline legislation may have in preventing accidents with spread released emissions. Within the estimated future pipeline accidents over a 14-year period, we expect that about 6 of the 7 proposed pipeline accidents with emissions that migrated off the Operator-controlled properties are caused by equipment failure, and about 21 of the 23 projected pipeline accidents that have emissions migrate off the Operated-controlled properties are caused by equipment failure. None of the accidents caused by material failure, corrosion, natural forces, and other causes, have emissions that have migrated off the Operated-controlled properties.

Similar to the environmental impact of CO₂ pipeline accidents, accidents resulting from certain causes have greater associated total costs than others. While accidents caused by natural forces are relatively rare (0.83%), when these accidents happen, they are the most expensive in related damages, accounting for over 54.26%, or \$3.95 million, of the total costs across all

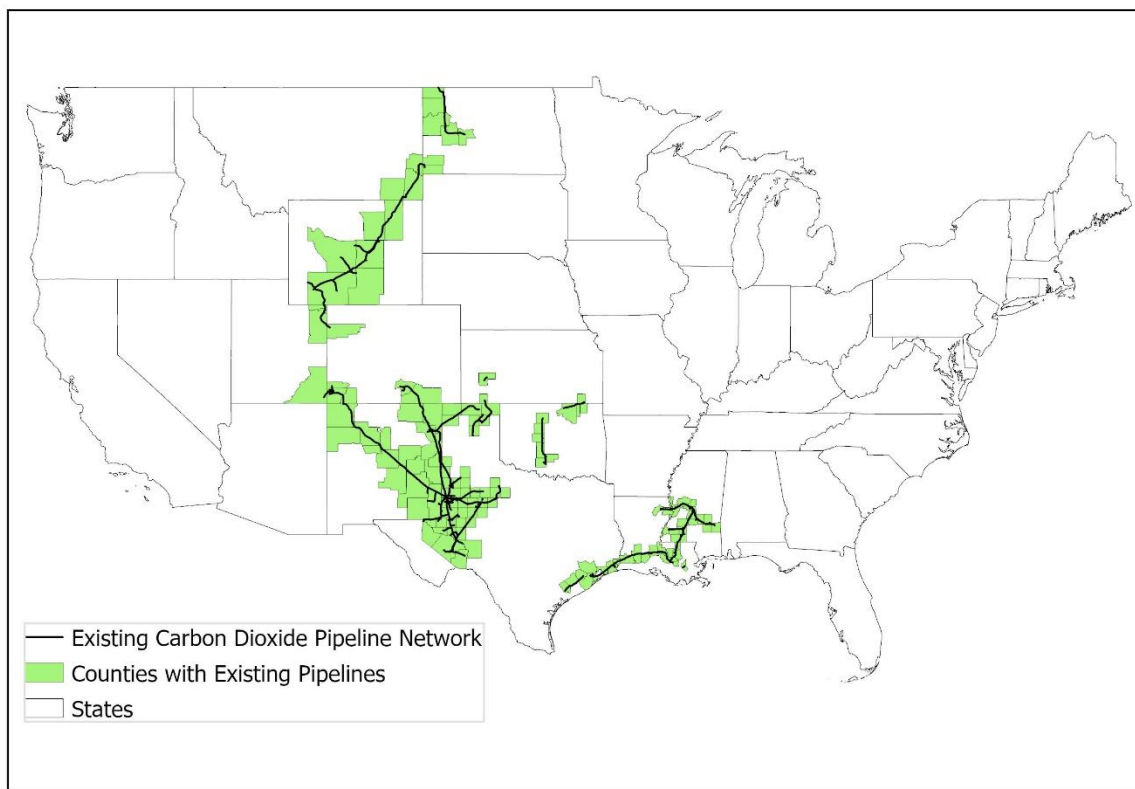
previous accidents (Supplementary Table S1). Conversely, accidents caused by equipment failure only account for about 8.29% of the total costs of all previous accidents, even though equipment failure causes over half of the accidents. We find that in the future, about 39.19% of the total costs from accidents along the proposed and projected pipelines are estimated to be from accidents caused by equipment failure, incorrect operation, material failure, or corrosion. These cause types of accidents along the proposed pipeline network are estimated to result in \$1.93 million in total damages (\$87,500 per year), and \$6.35 million in total damages (\$290,000 per year) from the projected pipeline route across 22 years (Supplementary Table S1).

Supplementary Table S1: *Estimated environmental damages and total costs by accident cause types, along the existing, proposed, and 2030 projected CO₂ pipeline routes. The estimates of gallons released CO₂ emissions are over a 30-year period (the existing estimates are specifically between 1994 and 2023), and the estimates of dollars of total costs are over a 22-year period (the existing estimates are specifically between 2002 and 2023). The estimates are rounded to the nearest whole number, and the rates are rounded to the nearest hundredth.*

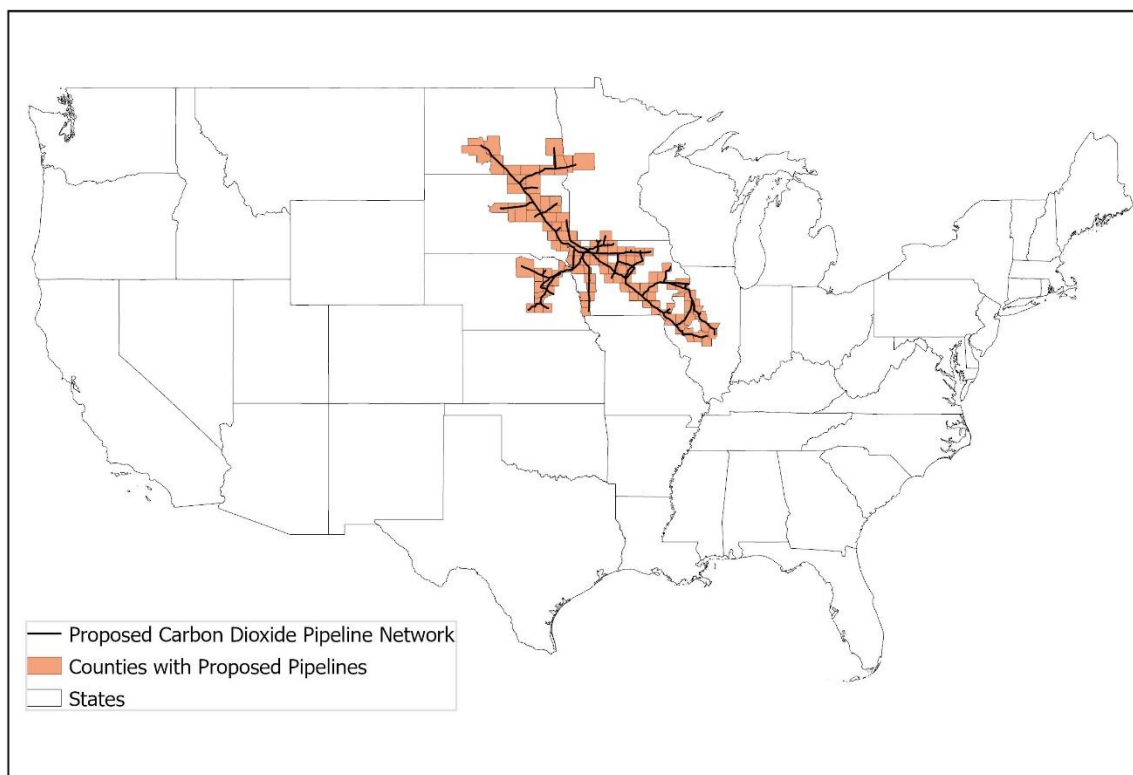
	Released CO ₂ Emissions				Total Costs			
	Rates (%)	Existing (Gal)	Proposed (Gal)	Projected (Gal)	Rates (%)	Existing (\$)	Proposed (\$)	Projected (\$)
Equipment Failure	24.99	4,347,744	2,936,778	9,679,883	8.29	603,245	407,475	1,343,074
Incorrect Operation	5.54	963,716	650,963	2,145,631	7.72	561,804	379,483	1,250,809
Material Failure	54.74	9,525,145	6,433,966	21,206,927	20.72	1,507,872	1,018,525	3,357,148
Corrosion	4.76	827,715	559,098	1,842,837	2.46	178,646	120,670	397,740
Natural Forces	7.58	1,319,010	890,954	2,936,664	54.25	3,947,009	2,666,093	8,787,680
Other	2.40	417,619	282,090	929,794	6.56	477,524	322,554	1,063,167
Total	100%	17,400,95	11,753,83	38,741,73	100%	7,276,10	4,914,80	16,199,61

Overall, most of these accidents and their impacts are, and likely will be, caused by equipment failure, incorrect operation, material failure, and corrosion. These four causes of accidents encapsulate the physical properties, construction, and operation and maintenance of pipelines, all of which can be regulated, strengthened, and monitored through state and federal legislation. Thus, more extensive regulations specific to the equipment, makeup, and operations of CO₂ pipelines may lower the expected additional accidents, and the negative consequences of these accidents, along newly constructed pipelines.

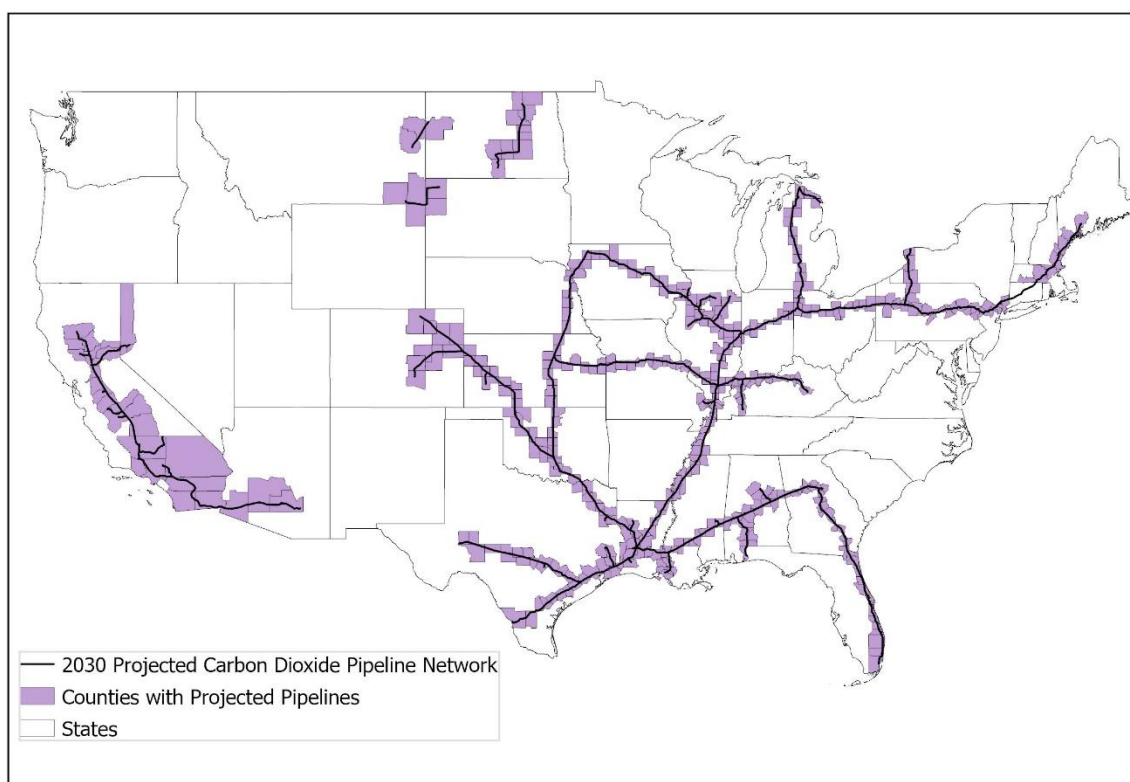
Supplementary Figures



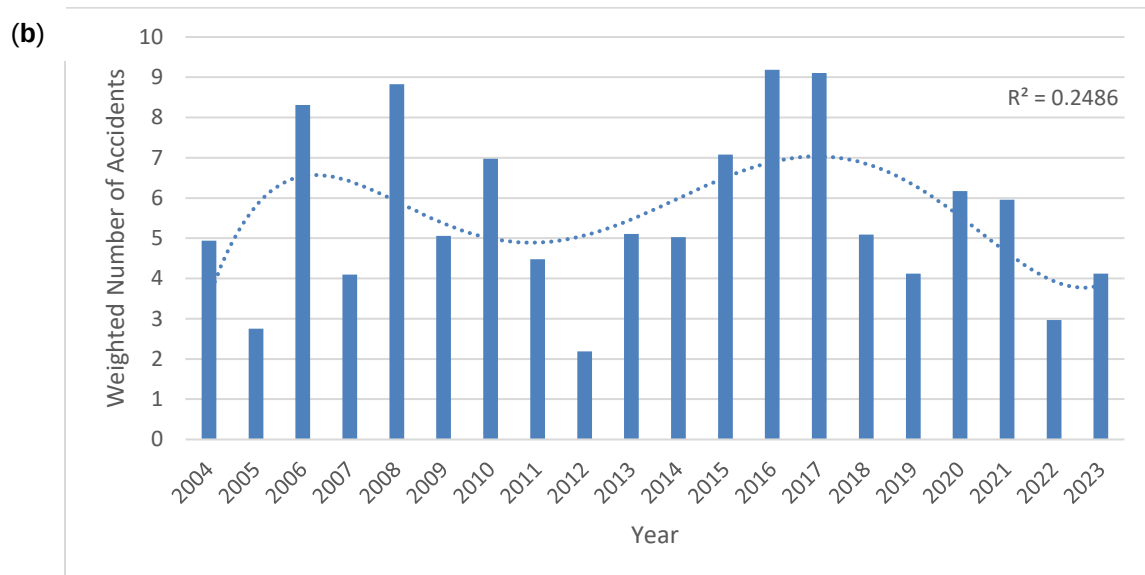
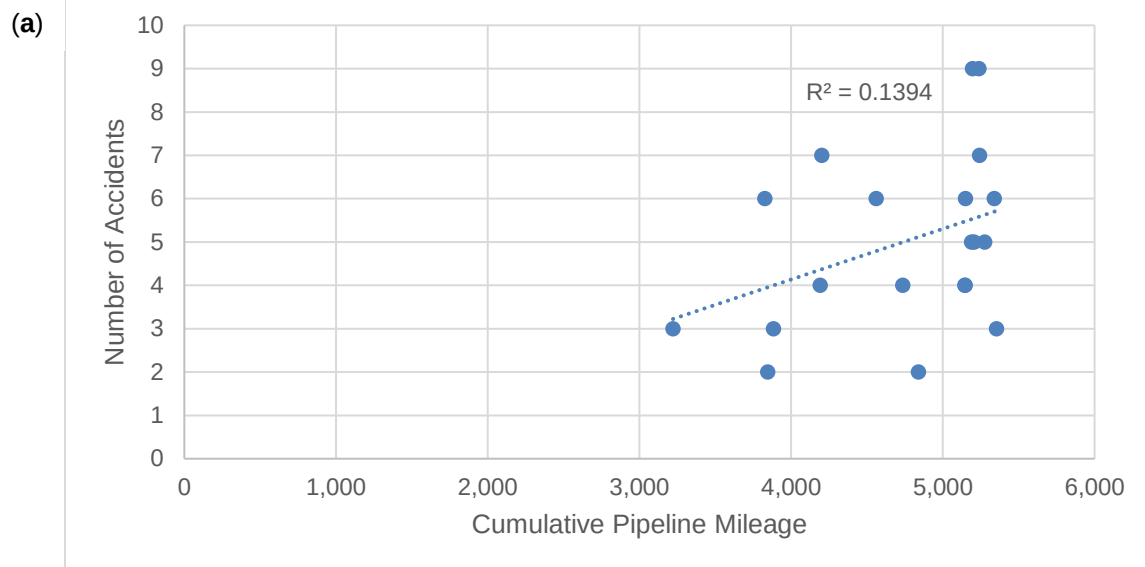
Supplementary Fig. S4: Map of the existing CO₂ pipeline network, totaling about 3,500 miles, and the counties in the respective sample.



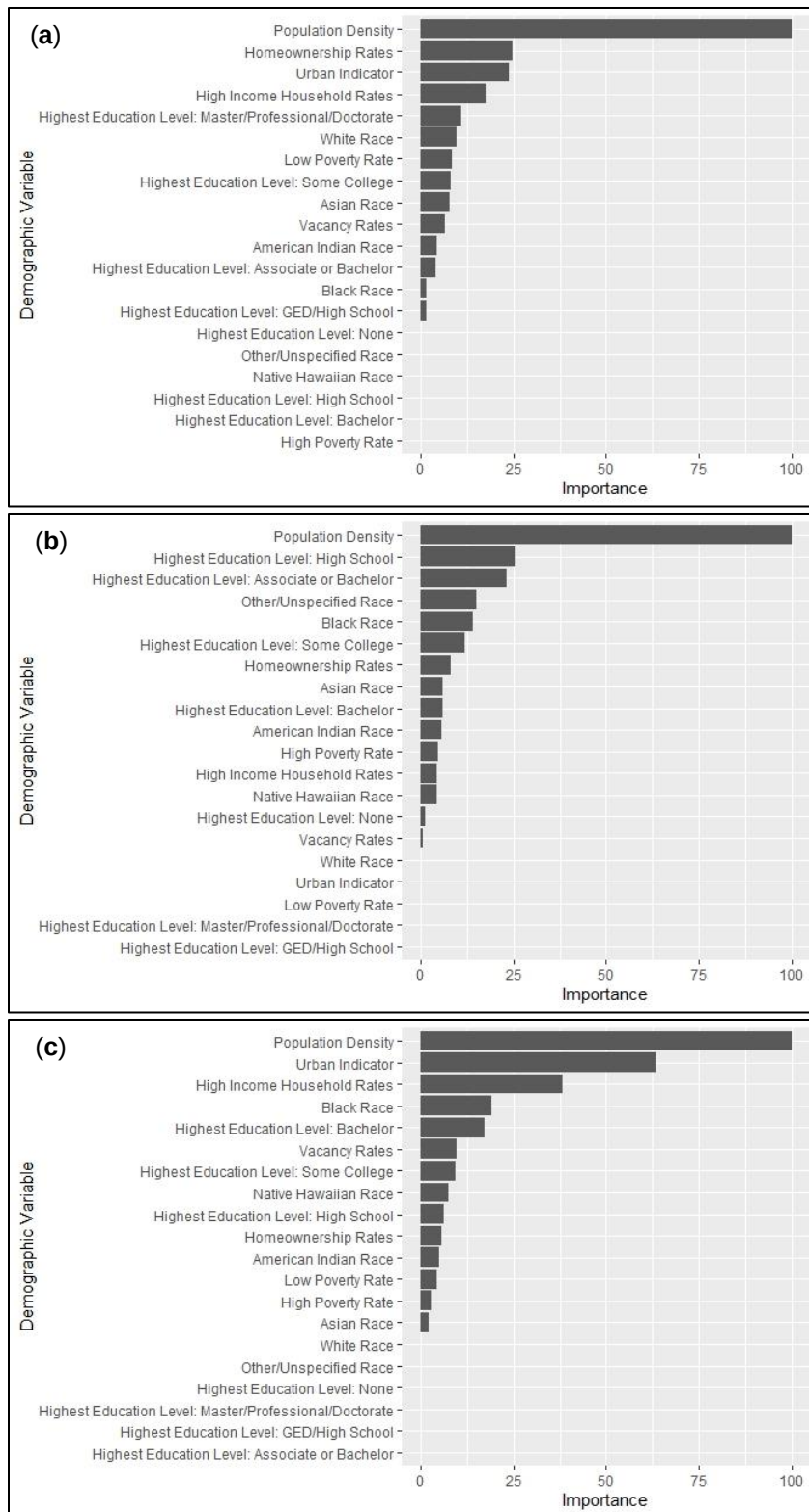
Supplementary Fig. S5: Map of the proposed CO₂ pipeline network and the counties in the respective sample. This network is composed of three proposals as new pipeline constructions, totaling about 3,580 miles.



Supplementary Fig. S6: Map of the 2030 projected CO₂ pipeline network, totaling about 11,800 miles, and the counties in the respective sample.



Supplementary Fig. S7: Comparison of the number of pipeline accidents and existing pipeline network mileage across time. Subfigure (a) depicts the number of accidents that occurred each year between 2004 and 2023, in comparison to the cumulative mileage of CO₂ pipeline infrastructure. We find a low, but positive, correlation between the pipeline mileage and the number of accidents. Subfigure (b) shows the weighted number of accidents each year, accounting for the proportion of total pipeline mileage of the existing pipeline network in use each year.



Supplementary Fig. S8: Ranking of importance of variables from the LASSO regression tool, for each pipeline network dataset. (a) shows the relative importance of each variable for the existing CO₂ pipeline network dataset, (b) shows the relative importance for the proposed CO₂ pipeline network dataset, and (c) shows the relative importance for the projected CO₂ pipeline network dataset.

Supplementary Tables

Supplementary Table S2: Descriptive statistics of census tracts in the United States, existing CO₂ pipeline network dataset, proposed CO₂ pipeline network dataset, and 2030 projected CO₂ pipeline network dataset. Across each variable, the pipeline datasets are statistically significant from one another at the 95-99% level (apart from the high school education level between the existing and projected pipeline networks, which are not statistically different from one another).

Variable	United States	Restricted Sample – Existing	Restricted Sample – Proposed	Restricted Sample – Projected
Population	3943.49	3976.70	3631.51	4079.98
Population Density (per 1,000 sq. miles)	5430.41	3440.88	1635.57	5307.22
Urban Indicator (urban = 1)	0.64	0.66	0.45	0.74
High Income Household (%)	8.83	7.57	5.30	9.20
High Poverty (%)	6.48	7.40	5.35	6.39
Low Poverty (%)	68.69	64.83	70.90	68.24
Homeownership (%)	64.56	63.67	70.28	61.69
Vacancy (%)	11.23	12.13	10.02	9.65
White Race (%)	38.89	35.95	45.86	37.54
Black Race (%)	9.96	12.47	3.25	9.49
American Indian/Native American Race (%)	0.63	2.11	0.35	0.51
Asian Race (%)	3.49	3.06	1.09	3.98
Native Hawaiian/Pacific Islander Race (%)	0.13	0.05	0.06	0.10
Other Race (%)	3.71	4.75	1.11	5.50
Highest Education Level (%): None	1.67	2.27	1.06	2.14
Highest Education Level (%): High School	23.12	22.19	26.12	22.38
Highest Education Level (%): GED/High School Degree	27.18	26.68	30.54	25.98
Highest Education Level (%): Some College Experience	20.07	21.11	21.87	20.29
Highest Education Level (%): Associate/Bachelor Degree	28.48	25.69	42.74	28.19
Highest Education Level (%): Bachelor Degree	19.80	17.92	18.55	19.71
Highest Education Level (%): Master/Professional/Doctorate Degree	12.51	10.88	8.89	12.13
Observations	83393	3051	1341	20911

Supplementary Table S3: *P-values from two-sample t-tests, comparing the averages of the demographic variables between the existing, proposed, and projected CO₂ pipeline network datasets.*

Variable	P-Value (Existing, Proposed)	P-Value (Existing, Projected)	P-Value (Proposed, Projected)
Population	3.8e-10	0.006626	<2.2e-16
Population Density (per 1,000 sq. miles)	<2.2e-16	<2.2e-16	<2.2e-16
Urban Indicator (urban = 1)	<2.2e-16	<2.2e-16	<2.2e-16
High Income Household (%)	<2.2e-16	2.552e-14	<2.2e-16
High Poverty (%)	<2.2e-16	1.873e-13	2.271e-08
Low Poverty (%)	<2.2e-16	<2.2e-16	1.818e-07
Homeownership (%)	<2.2e-16	3.101e-05	<2.2e-16
Vacancy (%)	1.086e-14	<2.2e-16	0.001697
White Race (%)	<2.2e-16	9.951e-14	<2.2e-16
Black Race (%)	<2.2e-16	<2.2e-16	<2.2e-16
American Indian/Native American Race (%)	<2.2e-16	<2.2e-16	6.293e-08
Asian Race (%)	<2.2e-16	1.831e-13	<2.2e-16
Native Hawaiian/Pacific Islander Race (%)	0.5753	8.242e-06	0.002143
Other Race (%)	<2.2e-16	4.167e-08	<2.2e-16
Highest Education Level (%): None	<2.2e-16	0.02582	<2.2e-16
Highest Education Level (%): High School	<2.2e-16	0.3216	<2.2e-16
Highest Education Level (%): GED/High School Degree	<2.2e-16	0.001579	<2.2e-16
Highest Education Level (%): Some College Experience	0.0001015	9.238e-09	2.71e-16
Highest Education Level (%): Associate/Bachelor Degree	2.053e-05	<2.2e-16	2.593e-14
Highest Education Level (%): Bachelor Degree	0.04678	9.196e-16	0.0002558
Highest Education Level (%): Master/Professional/Doctorate Degree	7.783e-14	2.311e-09	<2.2e-16

Supplementary Table S4: Binary logistic regression models for the existing, proposed, and 2030 projected pipeline networks, to estimate the likelihood of having a CO₂ pipeline in census tracts. Significance codes: 0 '****', 0.001 '***', 0.01 '**', 0.05 '*', 0.1, ' ' 1.

Variable	Existing Network Model (S.D.)	Proposed Network Model (S.D.)	Projected Network Model (S.D.)
Population Density	-0.0005837*** (0.0001889)	-0.002528**** (0.0003934)	-0.0001178**** (0.0000116)
Urban Indicator	-1.2757709*** (0.4072720)		-0.9084**** (0.07458)
High Income Household	-0.0420157** (0.0168392)	0.0574* (0.03388)	-0.02515**** (0.004502)
High Poverty		-0.0247 (0.03957)	0.005333 (0.005419)
Low Poverty	0.0074203 (0.0096441)		-0.0006999 (0.002960)
Homeownership	0.0255670**** (0.0077344)	0.03387** (0.01416)	0.001268 (0.00179)
Vacancy	-0.0162707* (0.0090785)	-0.005382 (0.01882)	-0.0007248 (0.00281)
White Race	0.0656975 (0.0408606)		
Black Race	0.0192799 (0.0211085)	0.1296**** (0.03078)	-0.0165**** (0.002260)
American Indian/Native American Race	0.0283081 (0.0208341)	0.1612 (0.1587)	-0.01896* (0.011)
Asian Race	-0.0172999 (0.0533223)	-0.1093 (0.1219)	0.006774 (0.004788)
Native Hawaiian/Pacific Islander Race		-1.807 (1.246)	-0.1297 (0.08189)
Other Race		0.2393*** (0.0823)	
Highest Education Level: None		-0.05161 (0.1395)	-0.006249 (0.01262)
Highest Education Level: High School		0.1001**** (0.03033)	0.008898** (0.004253)
Highest Education Level: GED/High School Degree	0.0034609 (0.0168709)		

Supplementary Table S4 (continued): Binary logistic regression models for the existing, proposed, and 2030 projected pipeline networks, to estimate the likelihood of having a CO₂ pipeline in census tracts. Significance codes: 0 '****', 0.001 '***', 0.01 '**', 0.05 '*', 0.1, ' ' 1.

Variable	Existing Network Model (S.D.)	Proposed Network Model (S.D.)	Projected Network Model (S.D.)
Highest Education Level: Some College Experience	-0.0287516* (0.0166613)	0.07001* (0.0359)	-0.01548*** (0.004773)
Highest Education Level: Associate/ Bachelor Degree	0.0151908 (0.0160090)	0.08953** (0.04061)	
Highest Education Level: Bachelor Degree		-0.02746 (0.0414)	-0.01299 (0.004999)
Highest Education Level: Master/Professional/ Doctorate Degree	0.0046500 (0.0225147)		
Constant	-6.8237387 (2.0823405)	-10.13 (2.699)	0.3697 (0.8169)
Observations	3051	1173	20911
County Fixed Effects	Yes	Yes	Yes

Supplementary Table S5: Regression models for the existing pipeline network to estimate the likelihood of having, and frequency of, a CO₂ pipeline accident in census tracts. Accident Model 1 shows the results from a linear regression analysis to estimate the number of accidents in a census tract based on our demographic variables of interest. Accident Model 2 shows the results from a binary logistic regression analysis to estimate the probability of having an accident in a census tract, dependent on our demographic variables of interest. Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 ' ', 0.1 ' ', 1.

Variable	Accident Model 1 (S.D.)	Accident Model 2 (S.D.)
Population Density	-0.00002636 (0.00007108)	-0.02980*** (0.01098)
Urban Indicator	-0.002668 (0.2427)	12.37 (2806)
High Income Household	-0.01688 (0.01058)	-0.2095 (0.507)
High Poverty	-0.003041 (0.009266)	0.01721 (0.09023)
Low Poverty	-0.005686 (0.005574)	-0.02588 (0.006063)
Homeownership	-0.001871 (0.004503)	-0.1957*** (0.07126)
Vacancy	-0.0005132 (0.005662)	-0.1011 (0.07221)
White Race	-0.005904 (0.02879)	0.05466 (0.3147)
Black Race	-0.004337 (0.01429)	0.03078 (0.1567)
American Indian/Native American Race	-0.003009 (0.01456)	0.03637 (0.1512)
Asian Race	0.02982 (0.04620)	0.3674 (0.6303)
Native Hawaiian/Pacific Islander Race	-0.05269 (0.2473)	3.632 (3.617)
Other Race	-0.005178 (0.02867)	0.5949 (0.4295)
Highest Education Level: None	0.01399 (0.02851)	-0.05991 (0.3105)
Highest Education Level: High School	-0.02541* (0.01499)	-0.1454 (0.1261)
Highest Education Level: GED/High School Degree	0.01646 (0.01636)	0.1792 (0.1453)

Supplementary Table S5 (continued): Regression models for the existing pipeline network to estimate the likelihood of having, and frequency of, a CO₂ pipeline accident in census tracts. Accident Model 1 shows the results from a linear regression analysis to estimate the number of accidents in a census tract based on our demographic variables of interest. Accident Model 2 shows the results from a binary logistic regression analysis to estimate the probability of having an accident in a census tract, dependent on our demographic variables of interest. Significance codes: 0 '****', 0.001 '***', 0.01 '**', 0.05 '*', 0.1, ' ' 1.

Variable	Accident Model 1 (S.D.)	Accident Model 2 (S.D.)
Highest Education Level: Some College Experience	-0.001786 (0.009670)	-0.02626 (0.09993)
Highest Education Level: Associate/Bachelor Degree	-0.01403 (0.01248)	-0.005743 (0.1589)
Highest Education Level: Bachelor Degree	0.02416* (0.01272)	0.1284 (0.1420)
Highest Education Level: Master/Professional/Doctorate Degree	0.005943 (0.01271)	0.3145 (0.1330)
Constant	1.149 (1.562)	-8.146 (0.0001561)
Observations	342	342
County Fixed Effects	Yes	Yes

Supplementary Table S6: Marginal effects of highly (95%) significant variables within the existing, proposed, and 2030 projected CO₂ pipeline networks.

CO ₂ Pipeline Network	Variable	Marginal Effect (percentage point change)	Standard Deviation	Baseline Statistic (percent)
Existing	High-Income Households	0.24	0.0035	7.57
	Urban Indicator	7.42	0.1078	66
	Population Density	4 (per 1,000 people per sq. mile)	0.00005	3440.88 (people per sq. mile)
	Homeownership	0.15	0.0022	63.67
Proposed	Black race	1.14	0.0125	3.25
	High School Education	0.88	0.0097	26.12
	Associate's/Bachelor's Degree Education	0.79	0.0087	42.74
	Population Density	20 (per 1,000 people per sq. mile)	0.0002	1635.57 (people per sq. mile)
	Homeownership	0.30	0.0033	70.28
2030 Projected	High-Income Household	0.20	0.0033	9.10
	College Education	0.12	0.0018	20.29
	High School Education	0.07	0.0011	22.38
	Black Race	0.08	0.0006	9.49
	Urban Indicator	7.13	0.0008	0.74
	Population Density	1 (per 1,000 people per sq. mile)	0.0649	5307.22 (people per sq. mile)

Data Availability: The complete dataset, including demographic data and spatial data of CO₂ pipeline routes, is publicly available as an ArcGIS Online Web Map at: <https://arcg.is/1LS8Sj2>. Additionally, the required input datasets used in this study are publicly available in the study's GitHub repository at: <https://github.com/UVA-CHNS/CO2Pipeline>.

Code Availability: The regression analysis was conducted using RStudio. The code for the regression analysis, along with the required input datasets, is publicly available in the study's GitHub repository at: <https://github.com/UVA-CHNS/CO2Pipeline>.

Supplementary References

1. S. Manson, *et al.*, IPUMS National Historical Geographic Information System: Version 18.0 [dataset]. [Preprint] (2023).
2. J. Strube, B. C. Thiede, W. E. “Ted” Auch, Proposed Pipelines and Environmental Justice: Exploring the Association between Race, Socioeconomic Status, and Pipeline Proposals in the United States*. *Rural Sociol* **86**, 647–672 (2021).
3. R. E. Emanuel, M. A. Caretta, L. Rivers, P. Vasudevan, Natural Gas Gathering and Transmission Pipelines and Social Vulnerability in the United States. *Geohealth* **5** (2021).
4. M. Allaire, H. Wu, U. Lall, National trends in drinking water quality violations. *Proc Natl Acad Sci U S A* **115**, 2078–2083 (2018).
5. P. and H. M. S. A. Office of Pipeline Safety U.S. Department of Transportation, Distribution, Transmission & Gathering, LNG, and Liquid Accident and Incident Data. [Preprint] (2024).