Root Causes: Relative Contribution Methodology

The model used to quantify the relationships and relative contributions of various social factors to mental health outcomes in the Root Causes project is inspired by, but not identical to, the multivariate "Tangled Threats" analysis described by Milstein B & Homer J, Milbank Quarterly 98(2):372-398, 2020.

The Root Causes model used county-level data for all 64 Colorado counties. Variables and sources are listed below under "Variable Definitions."

As in the Tangled Threats analysis, a series of stepwise multivariate linear regressions were performed across the counties. For a given outcome variable, the model first included all factors that might plausibly help to explain it. ("Plausibility" was based on common knowledge and health and social science literature.) Variables were eliminated when statistical significance was low or the polarity was not logical — for example, when the model reported a positive coefficient when a realistic explanation would indicate the coefficient should be negative. The most efficient set of predictors for an outcome variable was found in this way and by maximizing the R-squared value.

The model identified both "proximal," or direct, impacts of social factors, as well as indirect impacts. For example, mental distress was best predicted by seven direct factors: income disparity, poverty, unemployment, uninsurance, food access, physical inactivity, and insufficient sleep. These are the proximal factors.

Moving up the chain of causation, regressions were performed to identify influences on those seven factors, using the same statistical approach. For example, physical inactivity is partially predicted by the rate of primary care physician per 100,000 people, making primary care physicians a "stage 1 distal" factor for mental distress. The primary care physician rate is, in turn, partially predicted by college attendance rate, making education a "stage 2 distal" factor for mental distress. A full visual map of these factors is displayed on the main Root Causes website.

The positioning of these factors in the multiple chains of cause and effect is not always simple. For example, food environment was identified as both a proximal factor and a stage 1 distal factor for mental distress, working both directly and through insufficient sleep. Literature supports the theory that food insecurity directly worsens mental health and also interferes with sleep. The regression analysis demonstrates that these two effects are likely separate phenomena: if food affected mental distress only via sleep, then it would not have emerged as a significant factor (separate from sleep) in the proximal regression analysis.

The analysis continues this way up the chain of causation, with additional stages of regression, as long as significant factors continued to be revealed. Ultimately, 23 different factors were found significant at one stage or another. These are included in the model in addition to the three mental health data metrics used as outcomes. Some have multiple causal branches, as in the case of food access. Some factors, like college attendance and income disparity, may go through several causal stages on their way to ultimately impacting mental health outcomes.

Of these 23 factors, 18 are potentially modifiable, while five are non-modifiable demographic factors. Demographic factors included in the analysis but excluded from the visual map are the percentage of the population ages 65 or older, under age 18, Black or Hispanic/Latinx, rural, or non-English speaking. These variables were selected because they are commonly used as demographic controls and remained significant in the model.

(It's important to note that some factors one might expect to see as explanatory are not included here because they lacked statistical significance in combination with the other explanatory variables. That is, even if these excluded factors have some correlation with the outcome variable, they were found to lack independent explanatory power after taking other stronger explanatory factors into account. To some degree, this may depend on the specificity of available metrics.)

This analysis controls for these demographic factors but focuses on the 18 potentially modifiable factors when examining relative contributions to mental health outcomes. Relative contributions are calculated using three types of information:

1. The "starting value" is the population average for each of the mental health outcomes and the 18 modifiable factors. This value was calculated for Colorado overall, as well as for five geographical regions of Colorado.

2. A "best feasible value" was established to determine how much each of these factors could reasonably be expected to improve with policy or community interventions. This was set as the value of the "best" county in Colorado with a population of 10,000 or more. For example, the best feasible college attendance rate was set at 88.2%, the value for Douglas County, the highest in the state.

3. Variables were linked algebraically to factors above them in the chain of causation through the regression coefficients found previously.

To identify relative contribution, we first identified the percentage improvement in each of the three mental health outcomes that was "feasible" based on the Colorado county with the best mental health outcome. For example, the minimum for mental distress is (again) Douglas County, with a rate of 9.2%, compared with 11.8% for Colorado overall. If the Colorado state rate moved from 11.8% to 9.2%, this would constitute a 22.0% improvement.

By setting all seven proximal factors to their best feasible value, the regression analysis indicates that Colorado could (according to the regression analysis) achieve a 19.8% improvement in mental distress. Thus, all but 2.2 percentage points of the 22.0% "feasible" improvement could be gained by changing these seven factors.

A modifiable factor's **contribution** to improvement of a mental health outcome is defined as the percentage improvement in the outcome *if only that one factor* were moved from its starting value to its best feasible value. The **relative contribution** of the factor is defined as its contribution alone divided by the contributions of all proximal factors combined; that is, if all proximal factors were moved to their best feasible values.

Note that the contributions of these factors are not necessarily additive--one factor may have an impact that overlaps that of another factor. In such cases, their combined contribution will be less than the sum of their individual contributions.

Appendix A: Variable Definitions

Variable: Income disparity ratio
Source: 2014-2018 American Community Survey via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Ratio of income at the 80th percentile to income at the 20th percentile

Variable: College attendance rate
Source: 2014-2018 American Community Survey via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of adults ages 25-44 with some post-secondary education

Variable: 30+ minute commute alone
Source: 2014-2018 American Community Survey via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Among all workers, the percentage that commute alone in their car more than 30 minutes

Variable: Racial segregation index (white/non-white)
Source: 2014-2018 American Community Survey via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: An index ranging from 0 (complete integration) to 100 (complete segregation). This value is the percentage of either non-white residents or white (non-Hispanic/Latinx) residents that would have to move to different Census tracts in order to produce a distribution that matches the county's.

Variable: Emergency room presence
Source: 2019 Colorado Department of Public Health and Environment Facility List
Website: https://www.colorado.gov/pacific/cdphe/find-and-compare-facilities
Definition: Presence of emergency room in the county

Variable: Mental health facility presence
Source: 2019 Colorado Department of Public Health and Environment Facility List
Website: https://www.colorado.gov/pacific/cdphe/find-and-compare-facilities
Definition: Presence of a mental health facility in the county

Variable: Household poverty rate
Source: 2014-2018 American Community Survey
Website: https://data.census.gov/cedsci/advanced
Definition: Percentage of households with incomes below 100% of the federal poverty level

Variable: Unemployment rate
Source: 2018 Bureau of Labor Statistics via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population ages 16 and older seeking work but unemployed

Variable: Uninsured rate
Source: 2017 Small Area Health Insurance Estimates via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population under age 65 without health insurance

Variable: Social association rate
Source: 2017 County Business Partners via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Number of membership associations per 10,000 population

Variable: Primary care physicians per 100,000
Source: 2017 Area Health Resource File/American Medical Association via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Ratio of population to primary care physicians (per 100,000 residents)

Variable: Food environment index
Source: 2015 and 2017 USDA Food Environment Atlas, Map the Meal Gap from Feeding America via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Index weighting two aspects of the food environment: (1) percentage of the population that is low-income and does not have close access to a grocery store, and (2) percentage of the population without access to a reliable food source in the past year.

Variable: Exercise access rate
Source: 2010 and 2019 Business Analyst / Delorme map data / ESRI / US Census Tigerline Files via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population with adequate access to locations for physical activity

Variable: Juvenile arrest rate
Source: 2017 State and County Juvenile Court Case Counts via 2020 County Health Rankings

Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Rate of delinquency cases per 1,000 juveniles

Variable: Physical inactivity rate
Source: 2016 United States Diabetes Surveillance System via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of adults age 20 and over reporting no leisure-time physical activity

Variable: Violent crime rate
Source: 2016 and 2018 Uniform Crime Reporting (FBI) via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Number of reported violent crime offenses per 100,000 population

Variable: Insufficient sleep rate
Source: 2016 Behavioral Risk Factor Surveillance System via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of adults who report fewer than 7 hours of sleep on average

Variable: Severe housing problem rate
Source: 2012-2016 Comprehensive Housing Affordability Strategy (CHAS) data via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of households with at least 1 of 4 housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities

Variable: Mental distress rate
Source: 2017 Behavioral Risk Factor Surveillance System via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of adults reporting 14 or more days of poor mental health per month

Variable: Days of poor mental health
Source: 2017 Behavioral Risk Factor Surveillance System via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Average number of mentally unhealthy days reported in past 30 days (age-adjusted)

Variable: Suicide rate

Source: 2014-2018 National Center for Health Statistics - Mortality Files via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Number of deaths due to suicide per 100,000 population

Variable: Percentage of population under age 18
Source: 2018 Census Population Estimates via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population below 18 years of age

Variable: Percentage of population age 65 and older
Source: 2018 Census Population Estimates via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population ages 65 and older

Variable: Percentage of population that is Black or Hispanic/Latinx
Source: 2018 Census Population Estimates via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population that is non-Hispanic Black or Hispanic/Latinx (any race)

Variable: Percentage rural
Source: 2018 Census Population Estimates via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population living in a rural area

Variable: Percentage of population not proficient in English
Source: 2014-2018 American Community Survey via 2020 County Health Rankings
Website: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation
Definition: Percentage of population that is not proficient in English

Appendix B: Tables

COLORADO OVERALL (64 COUNTIES)	Starting Value	Best Feasible Value	Coefficients
Income disparity ratio	4.4	3.0	-
College attendance rate	71.7	88.2	-
30+ minute commute alone	26.7	10.6	-
Racial segregation index	28.6	19.3	-
Emergency room presence	0.982	1	-
Mental health facility presence	0.894	1	-
Household poverty rate (r-squared=0.65)	10.9	3.5	-
Coefficient from income disparity (p=.002)			3.19
Coefficient from college attendance rate (p=.006)			-0.120
Unemployment rate (r-squared=0.55)	3.3	1.8	-
Coefficient from income disparity (p=.03)			0.40
Uninsured rate (r-squared=0.52)	8.7	3.8	-
Coefficient from household poverty rate (p=.003)			0.21
Coefficient from college attendance rate (p=.28)			-0.027
Social association rate (r-squared=.10)	8.8	32.8	-
Coefficient from 30+ minute commute alone (p=.15)			-0.108
Primary care physicians per 100,000 (r-squared=.15)	82	144	-
Coefficient from college attendance rate (p=.03)			0.955
Food environment index (r-squared=.21)	8.4	9.2	-
Coefficient from household poverty rate (p=.03)			-0.045
Exercise access rate (r-squared=.49)	90.5	100	-
Coefficient from household poverty rate (p=.006)			-1.07
Coefficient from college attendance rate (p=.21)			0.212
Juvenile arrest rate (r-squared=.44)	14.2	6.8	-
Coefficient from food environment index (p=2x10-7)			-8.40
Coefficient from unemployment rate (p=.43)			0.79
Physical inactivity rate (r-squared=.62)	15.5	9.5	-
Coefficient from juvenile arrest rate (p=.004)			0.18
Coefficient from uninsurance rate (p=.19)			0.29
Coefficient from primary care physicians per 100,000 (p=.25)			-0.022

Coefficient from exercise access rate (p=.003)			-0.109
Coefficient from college attendance rate (p=.30)			-0.053
Violent crime rate (r-squared=.18)	326	51	-
Coefficient from unemployment rate (p=.11)			27.8
Insufficient sleep rate (r-squared=.66)	27.5	23.0	-
Coefficient from social association rate (p=.11)			-0.048
Coefficient from violent crime rate (p=.02)			0.0033
Coefficient from college attendance rate (p=.002)			-0.055
Coefficient from primary care physicians per 100,000 (p=.009)			-0.014
Coefficient from food environment index (p=.002)			-0.95
Severe housing problem rate (r-squared=.46)	16.5	7.7	-
Coefficient from uninsurance rate (p=.007)			0.52
Coefficient from unemployment rate (p=.004)			1.64
Coefficient from racial segregation index (p=.10)			0.067
Coefficient from household poverty rate (p=.11)			0.172
Mental distress rate (r-squared=.85)	11.8	9.2	-
Coefficient from uninsurance rate (p=1×10^{-5})			0.126
Coefficient from insufficient sleep rate (p=.04)			0.070
Coefficient from income disparity ratio (p=.05)			0.26
Coefficient from poverty (p=.0008)			0.057
Coefficient from food environment (p=4×10^{-5})			-0.28
Coefficient from unemployment (p=.09)			0.137
Coefficient from physical inactivity (p=.04)			0.030
Days of poor mental health (r-squared=.79)	3.8	2.9	-
Coefficient from uninsurance rate (p=.0001)			0.035
Coefficient from insufficient sleep rate (p=.008)			0.030
Coefficient from income disparity ratio (p=.13)			0.067
Coefficient from household poverty rate (p=.04)			0.011
Coefficient from food environment index (p=.001)			-0.071
Coefficient from unemployment rate (p=.17)			0.035
Coefficient from physical inactivity rate (p=.05)			0.0089
Suicide rate (r-squared=.62)	20.4	13.9	-
Coefficient from housing problem rate (p=.19)			0.441
Coefficient from days of poor mental health (p=.09)			5.99
Coefficient from emergency room presence (p=.007)			-9.93
Coefficient from mental health facility presence (p=.30)			-2.39

Region 1: Denver Metro Area	
7 counties, population ~3.20 million	**Starting Value**
Income disparity ratio	4.2
College attendance rate	74.2
30+ minute commute alone	30.4
Racial segregation index	29.8
Emergency room presence	1.000
Mental health facility presence	0.978
Household poverty rate	9.8
Unemployed rate	3.1
Uninsured rate	8.3
Social association rate	8
Primary care physicians per 100,000	92
Food environment index	8.5
Exercise access rate	97.1
Juvenile arrest rate	13.4
Physical inactivity rate	14.3
Violent crime rate	352
Insufficient sleep rate	27.1
Severe housing problem rate	16.5
Mental distress %	10.5
Days of poor mental health	3.5
Suicide rate	17.7

Region 2: Northwest and Mountain Resorts

9 counties, population ~230,000	Starting Value
Income disparity ratio	3.8
College attendance rate	66.7
30+ minute commute alone	21.3
Racial segregation index	26.6
Emergency room presence	0.994
Mental health facility presence	0.265
Household poverty rate	9
Unemployed rate	3
Uninsured rate	11.8
Social association rate	10.1
Primary care physicians per 100,000	83
Food environment index	8.6
Exercise access rate	92.4
Juvenile arrest rate	12.4
Physical inactivity rate	13.3
Violent crime rate	150
Insufficient sleep rate	25.4
Severe housing problem rate	19.1
Mental distress %	10.7
Days of poor mental health	3.5
Suicide rate	19.2

Region 3: Northeast and Eastern Plains

12 counties, population ~780,000	Starting Value
Income disparity ratio	4.2
College attendance rate	69.5
30+ minute commute alone	25.6
Racial segregation index	25.5
Emergency room presence	0.960
Mental health facility presence	0.883
Household poverty rate	11.4
Unemployed rate	2.9
Uninsured rate	8.7
Social association rate	8.3
Primary care physicians per 100,000	67
Food environment index	8.4
Exercise access rate	83
Juvenile arrest rate	17.3
Physical inactivity rate	16.6
Violent crime rate	223
Insufficient sleep rate	26.5
Severe housing problem rate	15.8
Mental distress %	11.1
Days of poor mental health	3.7
Suicide rate	20.8

Region 4: Southeast

14 counties, population ~1.05 million	Starting Value
Income disparity ratio	4.3
College attendance rate	69.8
30+ minute commute alone	20.8
Racial segregation index	28.3
Emergency room presence	0.966
Mental health facility presence	0.906
Household poverty rate	8.5
Unemployed rate	4.2
Uninsured rate	8.1
Social association rate	8.8
Primary care physicians per 100,000	59
Food environment index	7.6
Exercise access rate	81.7
Juvenile arrest rate	16.4
Physical inactivity rate	18.1
Violent crime rate	402
Insufficient sleep rate	29.4
Severe housing problem rate	16.2
Mental distress %	11.9
Days of poor mental health	3.9
Suicide rate	26.3

Region 5: Southwest and San Luis Valley

22 counties, population ~450,000	Starting Value
Income disparity ratio	4.5
College attendance rate	63.3
30+ minute commute alone	15.7
Racial segregation index	26.3
Emergency room presence	0.920
Mental health facility presence	0.600
Household poverty rate	15.4
Unemployed rate	3.7
Uninsured rate	10.9
Social association rate	9.9
Primary care physicians per 100,000	93
Food environment index	7.8
Exercise access rate	78.2
Juvenile arrest rate	16
Physical inactivity rate	17.2
Violent crime rate	227
Insufficient sleep rate	26.5
Severe housing problem rate	17.1
Mental distress %	12
Days of poor mental health	3.8
Suicide rate	28.5