3. Risk Analysis Overview

Natural hazard risk, in the most general terms, is often defined as the likelihood (or probability) of a natural hazard event happening multiplied by the expected consequence if a natural hazard event occurs. The generalized form of a risk equation is given in Equation 1.

Equation 1: Generalized Risk Equation

$$Risk = Likelihood\ x\ Consequence$$

3.1. Risk Calculation

In the National Risk Index, risk is defined as the potential for negative impacts as a result of a natural hazard. The risk equation behind the National Risk Index includes three components: a natural hazards risk component, a consequence enhancing component, and a consequence reduction component. EAL is the natural hazards risk component, measuring the expected loss of building value, population, and/or agriculture value each year due to natural hazards. Social Vulnerability is the consequence enhancing component and analyzes demographic characteristics to measure the susceptibility of social groups to the adverse impacts of natural hazards. Community Resilience is the consequence reduction component and uses demographic characteristics to measure a community's ability to prepare for, adapt to, withstand, and recover from the effects of natural hazards. The Social Vulnerability and Community Resilience components are combined into one Community Risk Factor (CRF) which is multiplied by the EAL component to calculate risk using Equation 2.

Equation 2: Generalized National Risk Index Risk Equation

$$Risk = Expected\ Annual\ Loss \times Community\ Risk\ Factor$$

$$where\ Community\ Risk\ Factor = f\left(\frac{Social\ Vulnerability}{Community\ Resilience}\right)$$

The Risk Index values form an absolute basis for measuring risk within the National Risk Index. They are used to generate all Risk Index scores and ratings. As described in Equation 2 above, Risk Index values are determined by multiplying EAL by the CRF, which is a scaling factor unique to each community, that varies as a function of the community's Social Vulnerability and Community Resilience values. All Risk Index values are calculated at the Census tract level. County and county-equivalent (e.g., boroughs, municipios, parishes) values are calculated by summing the values from their tracts.

The function, $f(\cdot)$ is atransformation that maps the ratio between Social Vulnerabilty and Community Resilience to CRF values. This mapping is constructed so that higher social vulnerability and lower community resilience, relative to all other communities at the same level (county or Census tract), result in higher Risk Index values for a given level of EAL. For more details regarding the way the components are estimated before entering the risk equation, see Sections 4.1 Social Vulnerability, 4.2 Community Resilience, 4.3 Community Risk Factor, and 4.4 Expected Annual Loss, respectively.

3.2. Values, Scores and Ratings

The National Risk Index provides three different types of results for Risk and each component used to derive Risk: EAL, Social Vulnerability, and Community Resilience:

- **Values.** Values for Risk and EAL are in units of dollars, representing the community's average economic loss from natural hazards each year. For Social Vulnerability and Community Resilience, values are the index values for the community provided by the source data sets.
- **Scores.** Scores represent the national percentile ranking of the community's component value compared to all other communities at the same level (county or Census tract).
- **Ratings.** Ratings are provided in one of five qualitative categories describing the community's component value in comparison to all other communities at the same level. Rating categories range from "Very Low" to "Very High."

In the risk equation, each component is represented by a score that represents a community's national percentile ranking relative to all other communities at the same level (county or Census tract). The composite Risk Index score is calculated to measure a community's risk to all 18 hazard types. The Risk Index score is a community's national percentile ranking in risk compared to all other communities at the same level. The Risk Index score and EAL score are provided as both composite scores from the summation of all 18 hazard types, as well as scores where each specific hazard type is considered separately.

All calculations are performed separately at two levels—county and Census tract—so scores are relative only within their level. It must be stressed that scores are relative, representing a community's relative ranking among all other communities for a given component and level. Scores are national percentiles, not absolute measurements, and should be expected to change over time either by their own changing measurements or changes in other communities.

All scores are constrained to a range of 0 (lowest possible value) to 100 (highest possible value). To achieve this range, scores for each component are determined by the given value's percentile ranking in the natonal distribution (rounded to the nearest hundredth).

For every score, there is also a qualitative rating category that describes the nature of a community's component value in comparison to all other communities at the same level, ranging from "Very Low" to "Very High." Rating categories are relative and for risk and EALthere are no specific numeric values that determine the rating. For example, a community's Risk Index score for one hazard could be 8.9 with a rating of "Relatively Low," while its Risk Index score for another hazard may be 11.3 with a rating of "Very Low." The rating is intended to classify a community for a specific component in relation to all other communities at the same level.

To determine ratings, an unsupervised machine learning technique known as k-means clustering or natural breaks is applied to each score. This approach divides all communities into groups or clusters such that the communities within each cluster are as similar as possible (minimized variance or inertia) while the clusters are as different as possible (maximized variance).

K-means clustering for rating designation is performed in the National Risk Index processing database using the k-means clustering algorithm in the Python library scikit-learn.[1] The algorithm is initialized with the following parameters:

- Number of clusters (n_clusters): 5
- Maximum iterations (max_iter): 500
- Number of times the algorithm is run with different centroid seeds to arrive at the output with the least inertia (n_init): 20
- Relative tolerance in the cluster centers of consecutive iterations to declare convergence (tol): 1×10^{-15}
- Random number generation seed for centroid initialization (random_state): 42

All other parameters are defaults. The algorithm works by selecting five random initial scores, one for each cluster centroid, and then the rest of the scores are iteratively assigned to a cluster based on proximity to the centroid. Cluster centroids are updated in each iteration as the mean value of the scores within each cluster. The algorithm stops when it completes the maximum number of iterations or the centroid calculations converge within the established tolerance, whichever occurs first. The finalized cluster of the lowest scores is assigned the rating “Very Low,” the next lowest cluster receives a rating of “Relatively Low,” and so on.

Ratings for Social Vulnerability and Community Resilience have specific numerical boundaries based on national percentiles. The ratings are dividied into quintiles as described below:

- **Very High:** 80^{th} to 100^{th} percentiles
- **Relatively High:** 60^{th} to 80^{th} percentiles
- **Relatively Moderate:** 40^{th} to 60^{th} percentiles
- **Relatively Low:** 20^{th} to 40^{th} percentiles
- **Very Low:** 0^{th} to 20^{th} percentiles

[1] See scikit-learn clustering documentation retrieved from https://scikit-learn.org/stable/modules/clustering.html#k-means.

The National Risk Index provides Risk Index values, scores and ratings for counties (and county-equivalents) and Census tracts for all 50 states and DC. EAL scores and ratings are available for all 50 states, DC, AS, MP, GU, PR, and VI. Social Vulnerability scores and ratings are available for all 50 states, DC, and PR. Community Resilience scores and ratings are available for all 50 states and DC.

In the application's maps and data visualizations, standard color schemes have been applied to the qualitative ratings. Risk Index ratings are represented using a diverging blue (Very Low) to red (Very High) color scheme. Ratings for EAL, Social Vulnerability, and Community Resilience are represented using sequential color schemes (e.g., single color at various intensities). Higher EAL, higher Social Vulnerability, and/or lower Community Resilience increase overall risk. In general, darker shading in the map layers represents a higher contribution to overall risk. When source data are not available or a score cannot be calculated, then additional ratings are used and shown in white or shades of gray. The standard color schemes are shown in Figure 3 with several illustrative examples of EAL, Social Vulnerability, Community Resilience, and risk scores and rating categories.

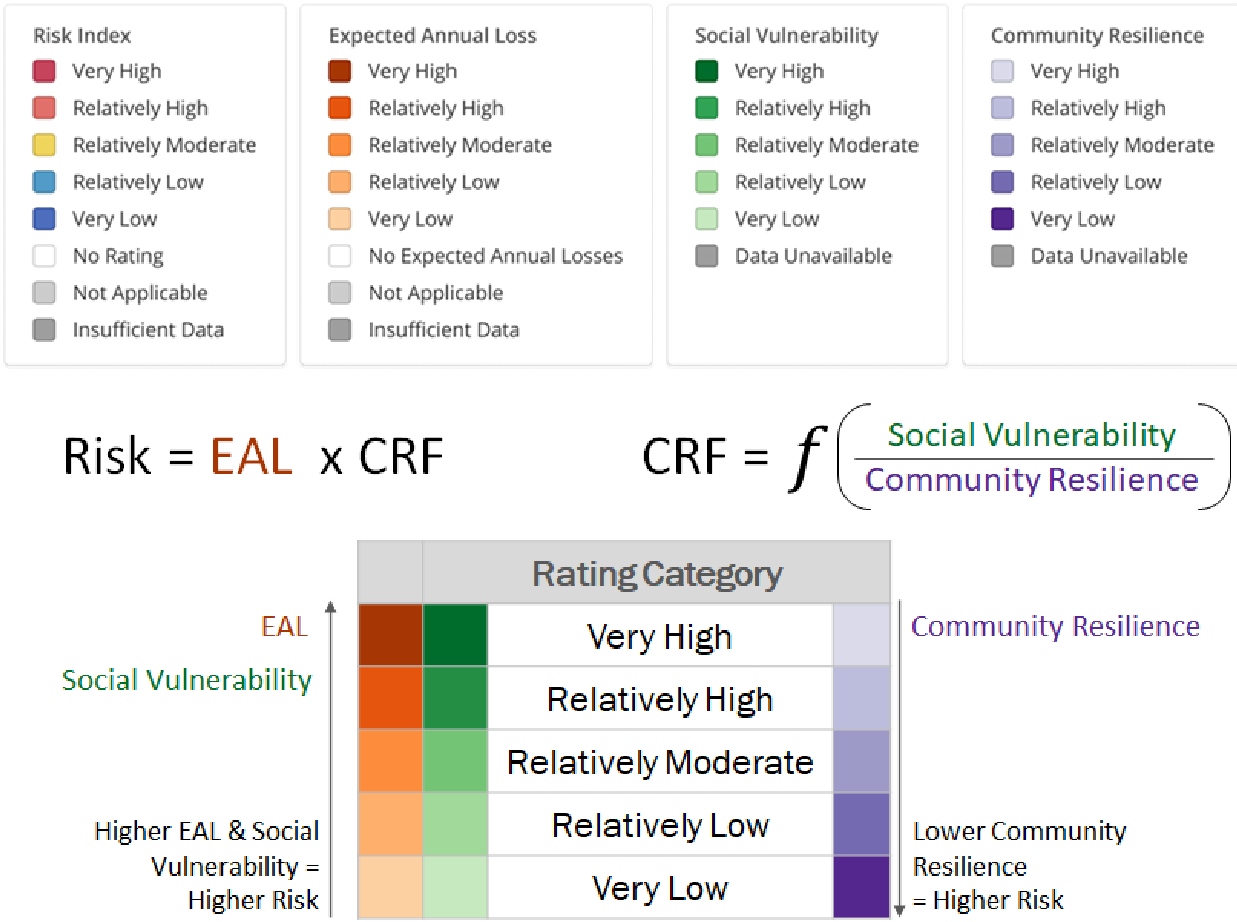

Illustration of Risk Component Scores

County	EAL	Social Vulnerability	Community Resilience	Risk
County 1	99.66	78.84	23.65	99.68
County 2	99.87	37.43	78.36	99.65
County 3	99.51	73.07	70.85	99.55
County 4	97.59	98.82	1.15	98.31
County 5	97.09	53.85	45.23	96.85
County 6	81.40	99.71	1.30	87.75
County 7	77.41	71.96	0.86	84.98
County 8	52.40	12.35	94.30	47.98
County 9	47.38	50.06	33.90	48.49
County 10	9.19	28.13	75.84	7.86

Figure 3: National Risk Index Qualitative Rating Legend and Illustration of Risk Component Scores

Component values of 0 (zero) or missing EAL values ("nulls") receive ratings that reflect the logic behind the score. A community where the EAL is zero either has no building value, population, or agriculture value exposed to the hazard type, or has a calculated annualized frequency of zero for the hazard type. These communities are displayed in the application as having "No EALes" for the designated hazard type, and .they will have "No Rating" for their Risk Index for that hazard type.

In collaboration with subject matter experts most familiar with a hazard type and the source data, a priori definitions of hazard type applicability have also been applied to help distinguish between where no risk exists for the hazard type and where the hazard type is deemed not able to occur. For example, Avalanche EAL is not computed for areas with no mountainous terrain. These communities are displayed in the application as "Not Applicable" for EAL computation for the designated hazard type.

Finally, if a factor used to calculate the EAL of a Census tract or county for a hazard type has a null value, the community is rated as "Insufficient Data." For example, certain hazard types, such as Wildfire, Lightning, and Landslide, only have source data used to determine annualized frequency or exposure available for the conterminous U.S., meaning that both Alaska and Hawaii are rated as "Insufficient Data" to compute the EAL for those hazard types. Census tracts and counties without Social Vulnerability or Community Resilience data are given a rating of "Data Unavailable". If a community has "Insufficient Data" to compute EAL or "Data Unavailable" for Social Vulnerability or Community Resilience, the community is given a Risk Index rating of "Insufficient Data." When a hazard type is not applicable or there are insufficient data for a community, EAL for that hazard type is simply not included in the community's final summation and scoring. A summary of non-numerical ratings is provided in Table 2.

Table 2: Definitions of Ratings without Numerical Scores

Rating	*EAL*	*Social Vulnerability*	*Community Resilience*	*Risk Index*
Not Applicable	Community is not considered at risk for hazard type.	n/a	n/a	Community is not considered at risk for hazard type.
Data Unavailable	n/a	Social Vulnerability data are not available.	Community Resilience data are not available.	n/a
Insufficient Data	Hazard source data are not available.	n/a	n/a	Social Vulnerability, Community Resilience, or hazard source data are not available.
No EAL	Hazard type exposure or annualized frequency is zero.	n/a	n/a	n/a
No Rating	n/a	n/a	n/a	EAL is zero.

3.3. Assumptions and Limitations

The National Risk Index dataset and application are meant for planning purposes only and are intended for use as a tool for broad, nationwide comparisons. Nationwide datasets used as inputs are in many cases not as accurate as locally available data. Users with access to local data for each risk component should consider substituting those data to calculate a more accurate EAL value at the local level.

The National Risk Index does not consider the intricate economic and physical interdependencies that exist across geographic regions. The user should be mindful that hazard impacts in surrounding counties or Census tracts can cause indirect losses in a location regardless of the location's risk profile.

The periods of record vary across hazard types and risk components with the most recent source datasets including a period of record up to 2022. It should be noted that the EAL values represent an extrapolation based on a snapshot in time. Extending source data collection beyond that time may result in varying Census tract or county EAL values due to changes in recorded hazard type severity and annualized frequency, as well as fluctuations in local economic value and/or population density.

Most of the hazard types use an annualized frequency model to determine EAL. This makes it difficult to accurately estimate EAL for high consequence, low frequency events. Certain rare hazard types (such as Earthquake, Hurricane, Tsunami, and Volcanic Activity) benefit from using a probabilistic model that estimates the likelihood of a hazard occurrence over an extended period of time, which can then be annualized. Of these, only Earthquake has probabilistic source data that are sufficient for accurately estimating EAL.[2] More details on the general methods used to calculate EAL are described in section 5 and the specific approaches for each hazard type are detailed in sections 6 through 23.

Best available nationwide data for some risk factors are rudimentary. More sophisticated risk analysis methodologies are available but require more temporally and spatially granular data for hazard exposure, annualized frequency, and historic loss ratio (HLR) measurements.

The Methodology makes various efforts to control for possible discrepancies in source data but cannot correct for all accuracy problems present in that data. The National Risk Index's processing database is a complex system, and localized inaccuracies in source data have the potential to propagate. Therefore, the National Risk Index and its components should be considered a baseline measurement and a guideline for determining natural hazard risk but should not be used as an absolute measurement of risk.

[2] Federal Emergency Management Administration (FEMA). (2017). Hazus Estimated AELs for the United States: FEMA Publication 366. Retrieved from https://www.fema.gov/sites/default/files/2020-07/fema_earthquakes_hazus-estimated-annualized-earthquake-losses-for-the-united-states_20170401.pdf.