Methods

A. A Rural Definition for New England

The first step in any analytic assessment of data along rural lines is to select or develop a rural definition that reflects the nature of the inquiry, as well as the rural nature of the area being studied. In planning the study, the New England Rural Health RoundTable (NERHRT) examined the range of existing rural definitions to determine which, if any, were of utility for this undertaking. The data advisory group discussed a series of principles felt to be of importance for any definition used. To be considered, the group agreed that the definition must

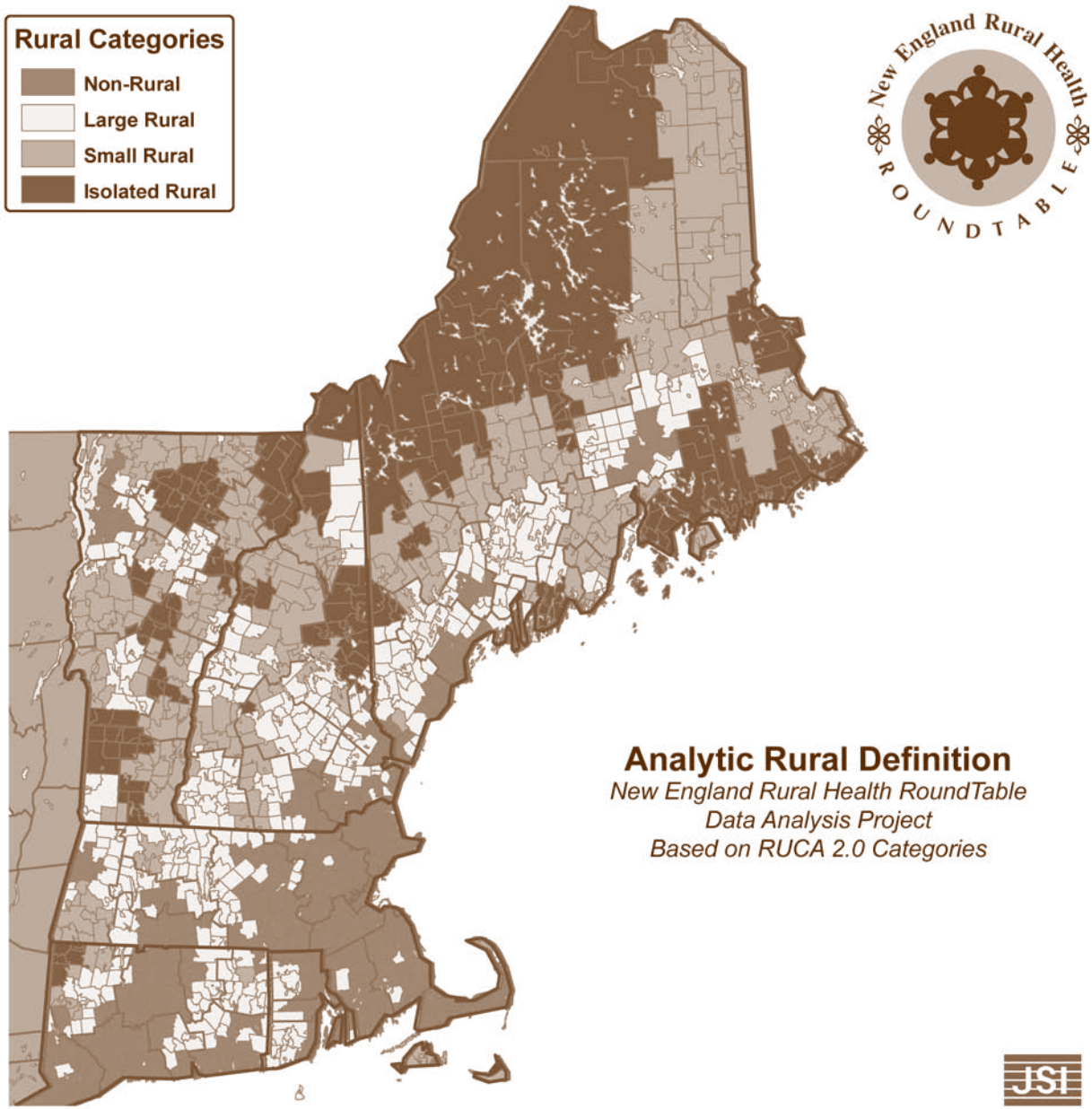
- be based on methods broadly acknowledged as relevant for rural definitions;
- be based on objective data, independent of underlying health-related statistics;
- treat rural communities as discrete groupings on a continuum, allowing for sub-analysis of more and less rural areas of the region;
- reflect the range of rural communities that exist within the New England states;
- produce geographic clusters that do not overly fragment the region;
- maintain a population size in the most rural tier that is statistically meaningful; and
- be capable of being transformed to conform with the structure/content of existing and available data.

It was determined that the Rural Urban Commuting Area (RUCA) codes (version 2.0), developed by the Federal Office of Rural Health Policy and WWAMI Rural Research Center, provided the best ability to meet these objectives. The RUCA codes are a series of over 30 designations, assigned to each zip code based on a combination of community size and the proportion of the population that commutes to a Census defined Urbanized Area or Urban Cluster. The RUCA codes are explicitly a non-linear set of designations intended to be grouped according to the nature of the rural study. According to the creators of the RUCA taxonomy at the WWAMI Rural Research Center, "The large number of codes facilitate the aggregation of the codes to fit specific needs of those using them for health, demographic, geographic, and other types of uses."[1]

Starting with one of the general grouping strategies recommended, the data advisory group engaged in a series of meetings to review and modify the categories to reflect the rural nature of communities in New England. Ultimately, the group agreed upon a clustering strategy for the RUCA codes yielding a four-tier rural definition for the study. This includes a non-rural tier, as well as three rural tiers, designated large rural, small rural, and isolated rural to follow common RUCA naming conventions. By further combining the three 'rural' tiers into a single group, this approach also yields a two-tier definition which permits the aggregate statistics (termed 'all-rural') to be compared to those for the non-rural areas. See Appendix A for detail on the RUCA codes and their assignment to the NERHRT analytic rural definition.

The map in figure 1 shows how the analytic definition applies across the region at the zip code level. Note that the rural tier definitions are applied regionally. As such, they combine populations across state lines and not all tiers exist in all states. A full color version of the map is available on the inside cover.

Figure 1



The table in figure 2 shows the results of the definition in terms of population, land mass, and resulting population density for each of the defined tiers. Note that 83% of the land mass of the New England region is rural based on this definition, yet this area encompasses only 20% of the population in the region. The average population density for the non-rural tier is nearly twenty times as high as the density in the combined rural tiers. As we move down through the rural tiers, the sparcity of the population continues to increase, with the isolated rural tier representing only 2% of the New England population, yet 27% of the land. The figures clearly demonstrate the rural nature of much of the New England region. It also highlights the need to focus on the unique health care challenges faced by relatively small portions of the total population which might otherwise be overlooked.

83% of the land mass of the New England region is rural based on this definition, yet this area encompasses only 20% of the population in the region.

Figure 2

NERHRT Rural Definition: Population, Land Area, and Density

Zip Code Level Definition	Population	Percent of Population	Land Area	Percent of Land Area	Average Population Density
Isolated Rural	246,628	2%	17,265	27%	14
Small Rural	782,233	6%	21,117	34%	37
Large Rural	1,734,076	12%	13,945	22%	124
Non-Rural	11,140,122	80%	10,471	17%	1,064
All Rural	**2,762,937**	**20%**	**52,326**	**83%**	**53**
Non-Rural	**11,140,122**	**80%**	**10,471**	**17%**	**1,064**

While the zip code based definition provided the best level of detail for aggregating RUCA codes, most data is not directly available at the zip code level. As such, it was necessary to transform this base definition into a series of alternate geographic units. These include the following:

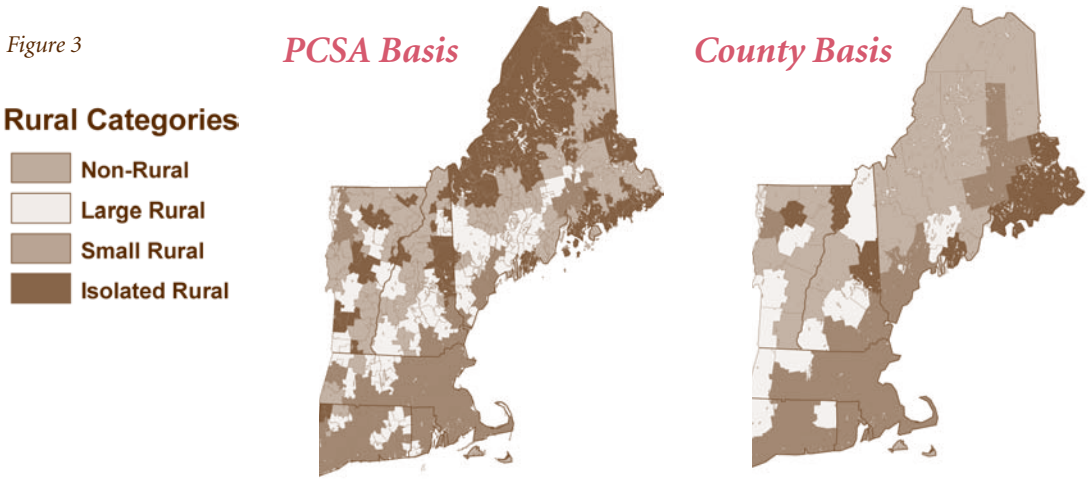
Zip Code Tabulation Areas (ZCTAs) – Census representations approximating one or more zip code areas. Reported for the first time in the 2000 decennial census, this definition is very close to the base (zip code) definition.

Primary Care Service Areas (PCSAs) v.2.0 – Aggregations of ZCTA units intended to represent rational service areas for the delivery of primary care. These units and underlying statistics were created by the Dartmouth Center for Evaluative Clinical Studies, under contract to the Health Resources and Services Administration (HRSA) by analyzing patient origin and destination information in the Medicare outpatient claims data.

Counties – Census-defined county boundaries

Because these alternate geographic units do not directly conform to the underlying zip code boundaries, transformations were performed by coding a detailed population-block layer according to the rural-definition tier of the zip code they are located in and then re-aggregating population units based on the alternate boundaries. As such, the population of a given PCSA or county may come from several different underlying rural tiers. The resulting population was then analyzed to determine which tier represented the greatest proportion of the residents (all-rural was first compared to non-rural before the sub-rural tiers were assigned). Because a portion of the population/area was forced to 'shift' between the tiers with different definitions, the level of 'rurality' varied slightly between the definitions, with the combined rural tiers representing 17% of the PCSA based definition and 15% of the county-based transformation, compared to 20% at the zip code level. Data was always collected at the finest level of detail possible. The maps in figure 3 show the PCSA and county transformations of the definition.

Figure 3



B. Analytic Method

Once created, the rural definitions developed for the project were used to aggregate a wide array of available data into statistics describing the distinct rural tiers and permitting comparison between the tiers. Where possible, statistics were aggregated directly to obtain the numerator (incidence) and denominator (population) for a given rate. Where only precalculated rates were available, statistics were aggregated using weighted means based on population. Survey data was aggregated at the case level, with results weighted to reflect the population at the state level where the data was collected. The report examines both crude and age-adjusted statistics, based on data availability and the relevance of age. Crude rates accurately reflect the per-capita statistics for the populations in a given area, while age-adjusted rates attempt to better isolate the impact of 'rurality' where age differences are a dominant factor.

Because the report is focused on describing rural health statistics across the rural continuum, each measure is compared across the four tiers individually (except where limited by data availability), as well as for the three rural tiers combined compared to the non-rural tier (termed the two-tier definition). The accompanying charts in this report show both the two-tier rural statistics, as well as the sub-category four-tier rural statistic analysis. The sub-category statistics are of particular importance as many differences are masked or diluted by variations within the rural groupings.

The report focuses on areas where notable differences or disparities in rural health statistics were identified. Please refer to Appendix B for a full listing of the detailed statistics examined and for the results of the sub-categories within the rural tiers.

C. Data Sets

The focus of the study is an examination of the health and health-related characteristics of rural communities in New England, and how they differ internally and compared with the non-rural areas of the region. This broad analytic approach necessitates that a wide variety of data sources be incorporated, in order to describe the region as completely as possible. As such, the data for this report is derived from a range of disparate sources, often available at different levels of geographic specificity and covering somewhat different time frames. Appendix B describes each of the data sources used for the study and provides details concerning the parameters and specificity of the data. The available data provided information concerning the three basic components of the health care system: the population; the delivery system; and the resulting patterns of access, utilization and outcomes that result from the interaction of the first two.

> ***"Not everything that can be counted counts, and not everything that counts can be counted."***
>
> **-Albert Einstein**

There were several requirements for data accessibility and content in order for the data set to be incorporated into the study. First, the data needed to be available uniformly across the entire New England region. Second, the data needed to be available at the county level or below to allow the application of the rural definitions. Third, due to resource limitations of the project, the data had to be available from a single source across the region—as opposed to procurement from individual state data repositories for example— and without an extensive research application process. Due to these constraints, readers will likely identify statistics from data sets or rural studies conducted in their state or local area that are not included in this study. Some of the findings from these sources have been incorporated by reference in this report where relevant and appropriate. In addition, there are many areas of interest or concern to the health care environment in rural communities for which statistics are not available. Due to these constraints, the absence of statistics on a particular topic should not be considered to imply that there are not valid concerns in these areas. Also, the statistics that are reported here are intended to be a starting point for identification and description of rural health characteristics and disparities, but additional depth of analysis may be required to more fully explore the issues identified.