Appendix

To inform the analysis in this study, we created a custom database of possible rail corridors, and metropolitan areas with demographic, employment, and transit data. Variables in the database are:

- Population
- Employment
- Transit ridership
- Population and employment within areas served by transit
- Air ridership along the corridor
- Highway congestion.

The steps to calculate these are outlined below.

Selection of start and end points for corridors

To create a list of start and end points for corridors, we identified the center of each metro area. First, we took the Census Bureau's list of Metropolitan Statistical Areas for the lower 48 states (Census, 2000). For each metro area, we identified the most populated city. If an Amtrak station is located in the city, we took the station to be the metro's center point for analysis.

Where no existing station existed for that metro area, we took the center point of the city. Three hundred sixty-four center points were identified.

Calculation of corridors

Using a map of the active passenger and freight national rail network (National Atlas of Transportation Data, 2007), the shortest travel paths between MSA centers were calculated for all MSA pairs within 600 miles. In Florida and California, the proposed HSR alignments were used to calculate the distances. We calculated 12,645 corridors. If a start or end point was not directly located on the rail system, the closest section of rail network was used as a starting point.

Metro profiles

For each metropolitan area, we assembled demographic data, using Census 2000 and population projections from the 2010 Complete Economic and Demographic Data Source from Woods and Poole Economics. Employment data were drawn from Bureau of Economic Analysis (BEA) 2007 estimates and Woods and Poole projections.

For each metro area, we calculated the total within 2, 10, and 25 miles of the MSA's center. Although summary data are available for metropolitan areas, we are specifically interested in the area of each region that is directly adjacent to existing or potential rail. When a tract or zip code is not completely enclosed within the 2, 10 or 25 mile study area, population was proportionally allocated based on area.

Population data

The population information describes the current and projected population in each metro area. Population data was calculated using census tracts from Census 2000, with projections taken from Woods and Poole 2010. Projections are at the county level, and apportioned to tracts based on the share of 2000 tract population compared to 2008 county population estimates.

Employment data

Employment information describes the labor mix in the metropolitan area, including total employment and employment in the following sectors:

- Finance and insurance
- Real estate, rental and lease
- Arts, entertainment and recreation
- Accommodation and food services

Employment was calculated on zip code tabulation areas (ZCTA) with base 2007 estimates from BEA. The zip-level employment data do not include government sector employment, so we used Woods and Poole county data to estimate the share of government employment at the zip level, and estimated the complete employment by zip. Where a ZCTA is not completely enclosed within the 2, 10 or 25 mile study area, employment was proportionally allocated based on area. Employment projections to 2040 were estimated from the county to zip based on share of employment in 2008.

How metro data was aggregated onto corridors

Once the corridors were calculated, we identified all urban centers located along the route, so that the total demographics for any corridor can be calculated to assess the multiple metropolitan areas it serves.

To calculate the demographic profile along each corridor between each city pair, we added up the data for each metro along that route. Each data point was counted only once per corridor, since in some locations the 25 and 10 mile study areas overlap (e.g. many corridors in the Midwest).

Transit system data

The corridor database includes ridership, and population and employment located near to mass transit.

We used the 2009 American Public Transit Association Fact Book to look up annual ridership. With this data, we identified all metros with non-bus mass transit.

Where available, we used the National Atlas of Transportation to identify the total coverage of routes. In cases where the transit networks were not completely represented (e.g. recent extensions to the Portland light rail), we used aerial photographs in Google Earth to locate the route, and add it to our map. Once mapped, we used the map to identify and add up population by tract within 0.5 mile of heavy and light rail, and 2 miles of commuter rail stations. We carried out the same calculation for employment, at the same distances.

Air data

Air market data was obtained from the T-100 segment market data from the Federal Aviation Administration's Trans-Stats data set and described total volume between major airports. We assigned each airport to the closest city, and then calculated the total volume of travel between all destinations along each rail corridor. For example, for the New York–Washington, DC corridor, we added up the total volume of flights between these cities along the corridor: New York City, Philadelphia, Baltimore, Washington, DC.

Traffic data

The Federal Highway Administration publishes the Freight Analysis Framework data set, including estimates of volume to capacity ratio (VCR) for 2002 and 2035 on the interstate and major road network. We calculated the shortest path by road between the end points for each corridor, excluding minor rural and urban arterials. For each section of road along the calculated path, we take the estimated VCR in 2002 and 2035, and add up the values from each segment to give a percentage breakdown of the whole corridor by VCR class (< 0.25, 0.25-0.5, 0.5-0.75, 0.75-1.0, >1).

These calculations were carried out in ArcGIS, a desktop mapping and analysis tool. Network calculations were carried out with pgRouting in PostGIS, a spatial database system.

Corridor Scoring Method

Each of these corridors was given a composite score based on a subset of the criteria described above. Only corridors that passed through one of the 49 metro regions in Table 69 were selected out for scoring. This reduced the total corridors from 12,645 to 7,870.

TABLE 69

Metro Regions for which Corridors were Selected

Megaregions	Metro Regions Included
Northeast	Baltimore, Boston, Hartford, New York, Philadelphia, Providence, Richmond, Washington DC
Florida	Jacksonville, Miami, Orlando, Tampa
Piedmont	Atlanta, Birmingham, Charlotte, Greensboro, Greenville, Raleigh
Cascadia	Portland, Seattle
Front Range	Albuquerque, Denver, Salt Lake City
Texas	Austin, Dallas, Houston, New Orleans, Oklahoma City, San Antonio, Tulsa
Southwest	Fresno, Las Vegas, Los Angeles, Phoenix, Riverside, Sacramento, San Diego, San Francisco, San Jose, Tucson
Great Lakes	Chicago, Detroit, Minneapolis, Saint Louis, Cleveland, Pittsburgh, Cincinnati, Kansas City, Milwaukee, Indianapolis

Preparing Data for Equation

Prior to applying an equation to the data to create the composite corridor score, we standardized the data such that every entry in the data base was a relative rank between zero and one.

First, each criterion was divided by the total length (in miles) of the corridor. This step resulted in the data being on a per mile basis, which allows for comparison between corridors of varying lengths. Without this step, longer corridors with more data points would have had an advantage over shorter corridors.

$$\text{Value}_n \,/\, \text{Length of Corridor}_n$$

Each criterion was given a rank of zero to 7,870 based on their relative value.

$$\text{Rank}\,(\text{Value}_n/\text{Length}_n)$$

These ranks were then converted to a value between 0 and 1 by dividing the rank by the maximum rank in each category and subtracting that result from 1. This yielded a number between 0 and 1 for each entry with the highest value 1 and lowest 0.

$$1 - (\text{Rank}_n \,/\, \text{Maximum Rank})$$

Final Equation

The final equation was then applied to these adjusted corridor ranks.

$$\text{Corridor Score} = 3*(RP+ECBD) +2*(TCE+TCP+CP+CE+RPGE+RAM) + (CRP+CTC+SF+ST)$$

TABLE 70

Criteria Used to Develop Corridor Score

Primary Factors: Weighted 3X	
Regional Population (25Miles)	(RP)
Employment CBD (2Mile)	(ECBD)
Secondary Factors: Weighted 2X	
Transit Connectivity Employment	(TCE)
Transit Connectivity Population	(TCP)
City Population (10 Mile)	(CP)
City Employment (10 Mile)	(CE)
Regional Population Growth Factor	(RPGF)
Regional Air Market	(RAM)
Tertiary Factors: Weighted 1X	
Commuter Rail Connectivity Population	(CRP)
Corridor Traffic Congestion	(CTC)
Share of Financial Workers	(SF)
Share of Workers in Tourism Industry	(ST)