Appendix A: Technical Notes

Introduction

The Community Need Index (CNI) is a composite measure designed to assess the need for HIV/AIDS-related services by ZIP code. There are two main issues pertaining to the measurement of service need by geography. First, the need for services is an abstract construct and thus cannot be observed directly. Furthermore, there is no known measure readily available to evaluate and compare the need for services in small geographic areas. Second, the need for services may emerge for different reasons and a single indicator may not fully capture multiple sources of service need. For instance, communities combating injection substance use may require treatment and prevention services that are markedly different from communities with elevated levels of unprotected sexual activity among teenagers. Using a multiple-indicator approach to these problems, the CNI project combined the values of nine ZIP-code level variables into a single index for every ZIP code located in each of the three CNI statistical areas in New York State. The CNI allows service providers to easily identify ZIP codes with the highest need for HIV/AIDS-related services. Providers may allocate their resources and effort more effectively according to the relative needs of the communities within their service areas.

The purpose of this appendix is two-fold. First, it explains the major steps taken in calculating the CNI. Second, it discusses issues that are critical to the interpretation of the data tables contained in this report. Selected examples of how to use the CNI in various applications are provided in Appendix B. Readers who want to analyze the data sets used in the CNI report series may refer to Appendix C for data documentation.

Need Indicators

The CNI is constructed using nine ZIP-code level need indicators. Each indicator is a health statistic expressed as a rate or a ratio that is based on three years of data, and it is presumed to be associated with one or more underlying risk factors for the transmission of HIV in a local community (Table A1).

Index Construction

The CNI is computed separately for each of the three CNI statistical areas in New York State. An independent statistical model is estimated for each CNI statistical area. ZIP codes with less than 500 residents are excluded from the estimation process. All indicators are standardized (i.e. they have a mean of zero and a standard deviation of one) and a statistical weight is computed for each standardized indicator. Table A2 shows the weights for three CNI statistical areas,

Table A1. Need indicators and the associated risk factors.

Indicator	*Presence of HIV/AIDS*	*Risk Factor*	
		Drug Use Behavior	*Sexual Behavior*
AIDS Cases	X		
HIV Cases	X		
Maternal Seroprevalence	X		
Female HIV Discharge	X		
Male HIV Discharge	X		
Cocaine Discharge		X	
Opioid Discharge		X	
Teenage Pregnancy			X
Sexually Transmitted Diseases			X

Table A2. Statistical weights for the nine indicators by CNI area.*

Indicator	*CNI Statistical Area*		
	NYC	*NYV*	*ROS*
AIDS Cases	0.141	0.140	0.132
HIV Cases	0.138	0.138	0.130
Maternal Seroprevalence	0.123	0.129	0.122
Female HIV Discharge	0.140	0.129	0.128
Male HIV Discharge	0.133	0.113	0.113
Cocaine Discharge	0.132	0.121	0.126
Opioid Discharge	0.130	0.140	0.102
Teenage Pregnancy	0.099	0.117	0.128
Sexually Transmitted Diseases	0.128	0.117	0.122

New York City (NYC), New York City Vicinity (NYV), and Rest of State (ROS), respectively. These weights are used to determine the CNI raw scores for every ZIP code, including those with less than 500 residents. They represent the relative effects of the indicators on the CNI within a statistical area. The calculation of these weights is discussed in Step 9 of The CNI Algorithm section below.

The CNI Algorithm

1. "Point ZIPs", ZIP codes that represent office buildings, post offices with mail delivery, or mail boxes but without residence population, are re-classified as parts of the surrounding ZIP codes. Data from these ZIP codes are also reassigned to the nearby ZIP codes that have population.

2. Indicator rates are created by dividing a numerator (number of cases) by a population denominator, as indicated in Table 1.2. Rates that include three years of data over a population are multiplied by a conventional number of persons to give a rate per 1,000 or 100,000, and are divided by three to give an average that approximates a one year rate. In the case of Newborn Seroprevalence the denominator also contains three years of data, so it is not necessary to divide by three. For any given indicator, a denominator with a value of zero in a ZIP code will be assigned with a rate of zero.

3. New York State is divided into three CNI statistical areas - New York City, New York City Vicinity, and Rest of State. The CNI for each CNI statistical area is computed separately.

4. ZIP codes with either less than 500 residents or less than 5 cases are excluded from weight estimation described in Step 5 through 8 below.

5. SAS PROC CORR is used to generate a correlation matrix for the ten indicators.

6. SAS PROC VARCLUS is run on the correlation matrix. PROC VARCLUS divides the nine indicators into varying numbers of clusters by maximizing the sum across clusters of the variance of the ten indicators that is explained by the cluster components. In addition, the procedure produces a set of scoring coefficients b_{ij} for the standardized indicators x_{ij} associated with a cluster $CLUS_j$, where i denotes any one of the n_j indicators (i=1,…, i, …n_j), and j denotes any one of the J clusters (j=1, …, j,…, J). The scoring coefficients will be used to create cluster scores, c_j, in the next step.

7. SAS PROC SCORE is used to create a set of cluster scores, c_1, c_2,…,c_J, for each ZIP code. A cluster score c_j is a linear combination of the products of the scoring coefficients b_{ij} and the standardized indicators x_{ij} (i.e. the sum of $b_{ij}x_{ij}$) associated with a cluster $CLUS_j$:

$$c_j = \sum_{i\ 1}^{n_j} b_{ij}x_{ij} = b_{1j}x_{1j} + b_{2j}x_{2j} + ... + b_{nj.j}x_{nj.j}$$

8. The variance of the indicators x_{ij} explained by the cluster $CLUS_j$ is labeled as k_j. The total variance of all nine indicators explained by all clusters is equal to $(k_1+k_2+\ldots+k_J)$. The relative proportion of variance explained by any single cluster $CLUS_j$ is equal to $k_j/(k_1+k_2+\ldots+k_J)$, or simplified as K_j.

9. CNI scores are computed for all ZIP code areas (including ZIP codes with less than 500 residents) as

10. $$\sum_{j=1}^{J}\sum_{i=1}^{n_j} K_j b_{ij} x_{ij} = \sum_{j=1}^{J} K_j \sum_{i=1}^{n_j} b_{ij} x_{ij} = \sum_{j=1}^{J} K_j c_j = K_1 c_1 + K_2 c_2 + \ldots. + K_J c_J$$

Notice that the products $K_j b_{ij}$ are equivalent to the weights listed in Table A2. They represent the relative effects of the indicators on the CNI within a statistical area.

11. The CNI scores are transformed by adding the absolute value of the lowest score to the original scores, multiplied by 100, and rounding the transformed scores to the nearest digit. This is done to avoid negative scores and scores with many decimal digits which would be more difficult to read and compare.

12. If the ZIP code has less than 50 people or less than six cases in the denominators of all rates, the CNI score is considered unreliable, and is assigned a value of zero which represents a low CNI score.

13. In each of the three statistical areas, ZIP codes are rank-ordered from the highest to the lowest CNI score. The top 20% of the ZIP codes (80th percentile and above) are classified as having high need for HIV/AIDS-related services; the next 20% of the ZIP codes are classified as having moderate need; and the remaining 60% are considered as low need areas. The CNI letter scores, H, M, and L, are used throughout the report to denote ZIP codes with high, moderate, and low need for services, respectively.

14. If necessary, ZIP codes are added or subtracted from a letter group so that no two ZIP code areas with the same score (before rounding) are assigned to different letter groups. The choice of adding or subtracting is made so that the change is applied to the smallest number of ZIP codes.

15. It should be noted that the CNI scores can be used to compare relative service need of two or more ZIP codes from the same CNI statistical area. However, ZIP codes from two different statistical areas (i.e. one ZIP code from NYC and the other from NYV) should not be compared directly since the two sets of CNI scores are computed based on different statistical parameters.

Data Suppression

The CNI is an index. It is a composite of nine rates that indicate presence of HIV, or of risk behavior that could lead to HIV. Each of the nine rates is based on data that are subject to confidentiality protocols. These protocols are instituted to avoid publication of rates with very small numerators or denominators which, in a small community, might lead to identification of individuals. The CNI component rates are published in Table 1.2. Rates in ZIP code areas with small numerators or denominators are suppressed according to protocols required by data owners. Those protocols are listed in Table A3. In cases where a single suppressed rate might be computed from a knowledge of census denominators and a county rate, an additional rate is suppressed.

The CNI itself is not suppressed, but is converted to "L" and to a number score of zero in ZIP code areas where the sum of the numerators of component rates is less than six, or where the population of the area is less than fifty. In most but not all of these cases, the CNI was "L" before this conversion. This conversion has the same effect and some of the same motivation as suppression. The conversion is also done because it is thought that CNI scores based on fewer cases are not reliable.

Table A3. Suppression protocols for CNI variables.

Rate	*Numerator (3 year totals)*	*Denominator*	*Suppress when:*	
			Numerator	*Denominator*
AIDS Cases	AIDS Cases aged13-64 np*	Population 13-64 np	n < 3	
HIV Cases	HIV Cases aged 13-64 np	Population 13-64 np	n < 3	
Maternal Seroprevalence	Newborns positive	Newborns tested		d < 50
Female HIV Discharge	Female HIV hospital discharge 13-64 np	Females 13-64 np	n < 6	d < 50
Male HIV Discharge	Male HIV hospital discharge 13-64 np	Males 13-64 np	n < 6	d < 50
Cocaine Discharge	Cocaine hospital discharge 13-64 np	Population 13-64 np	n < 6	d < 50
Opioid Discharge	Opioid hospital discharge 13-64 np	Population 13-64 np	n < 6	d < 50
Teenage Pregnancy	Pregnancies to females aged 10-17	Females aged 10-17	n < 6	d < 50
Sexually Transmitted Diseases**	STD cases aged 13-63	Population 13-64	n < 6	

* Non-prison population
** STD includes syphilis, gonorrhea, and chlamydia

Cautions When Using the CNI and the Indicator Rates

Some of the known limitations of the CNI are discussed in Section 1. Other technical issues that may affect the interpretation of the CNI and the indicator rates are ZIP code definitions.

ZIP codes are the lowest level of geography that are available on all data sets. However, ZIP codes were originally developed for mail delivery and not designed for data tabulation purpose. Problems arise when ZIP code definitions and boundaries are changed by the U.S. Postal Service. When this happened, attempts were made to reassign the numerators and the denominators to the appropriate areas. However, new ZIP codes may not always match the map boundary files in the CNI database. The vintage of the ZIP code boundary file (2002-03) used in mapping and tabulation was chosen to closely match the years of the data presented in this report. This will result in blank areas on the maps. Another problem is that while attempts have been made to determine addresses and ZIP codes of residence for all cases, there are occasions in which the data may be assigned to temporary residences or to post office boxes. It should be noted that the addresses of a hospital patient shows the ZIP code of the patient's residence, not the ZIP code of the hospital. Therefore, a ZIP code with a high CNI score or high indicator rates cannot be attributed to the mere presence of one or more hospitals in the community. If this happens frequently in a low-population area, the cases in the numerator can be larger than they should be given the denominator. For this reason, areas with low populations but many temporary residents or many mailboxes in associated ZIP codes may have rates that are higher than appropriate. A final problem is that a ZIP code may be missing from a county map or table. This may occur for several reasons: (1) The ZIP code is a "point" ZIP, that is, an office building or a post office with mail delivery or mail boxes but without population; (2) the ZIP code is associated with a small town that has a non-delivery post office and mail delivery from another ZIP code; it is assigned to the ZIP code of delivery; (3) the ZIP code has been changed by the Postal Service; (4) the ZIP code crosses one or more county boundaries. The ZIP code is assigned to a primary county, and does not appear on the maps and tables of the other counties with which it is associated.

History of the CNI

Previous editions of the Community Need Index were published in 1991, 1992, 1994, 1995, and 2000. The basic CNI model has been fairly consistent for all editions. Methodological modifications were introduced over the years to improve the utility of the CNI. The 1991 edition did not include sexually transmitted diseases as an indicator since the data were not available. Editions for 1992 through 2000 included ten variables as shown in Table A3. Based on the results of focus groups, variable correlations, data availability, and many formulations, the current version dropped one variable, combined two, and added another. Weighting by the relative proportion of variance explained by each cluster was introduced in 1992 to adjust for the relative contribution of the clusters to the overall CNI.

The number of years of data used for the numerator of each indicator varied over the years (Table A4). In 1992, 1994, and 1995, more years of data were added to improve stability of the indicator rates. In 2000 and 2006, the number of years of data used in model estimation were restricted to three for all indicators in order to show more recent patterns of service needs. The numerators and denominators for the newborn seroprevalence and the low birthweight indicators are from the same data years, whereas population-based denominators used U.S. Census figures or other population estimates as shown in Table A4. The parameter estimates of the 1992, 1994 and 1995 CNI model were based on ZIP codes with population of 4,000 or more. The 2000 and 2006 edition lowered the threshold to include ZIP codes with population of 500 or more.

Table A4. Published editions of the CNI and years of data included.

Indicator	*Edition*					
	1991	*1992*	*1994*	*1995*	*2000*	*2006*
HIV Cases						2001-03
AIDS Cases						2001-03
AIDS Cases: MSM	1985-89	1985-90	1985-91	1988-94	1994-96	
AIDS Cases: Other	1985-89	1985-90	1985-91	1988-94	1994-96	
Maternal / Neonatal Seroprevalence	1987-90	1987-91	1987-92	1987-94	1995-97	2001-03
Female HIV Discharge	1988-89	1989-90	1990-92	1991-93	1995-97	2001-03
Male HIV Discharge	1988-89	1989-90	1990-92	1991-93	1995-97	2001-03
Cocaine Discharge	1988-89	1989-90	1990-92	1991-93	1995-97	2001-03
Opioid Discharge	1988-89	1989-90	1990-92	1991-93	1995-97	2001-03
Teenage Pregnancy	1989	1989-90	1989-91	1991-93	1994-96	2000-02
Sexually Transmitted Diseases	n/a	1990	1990-91	1991-93	1995-97	2001-03
Low Birth Weight	1989	1989-90	1989-91	1991-93	1994-96	
Population	NPD* 90	NPD 90	Census 1990	Census 1990	Claritas** 1996	Census 2000

* National Planning Data 1990
** Population estimates for 1996 from Claritas, Inc.