**Behavioral Health Index -2**

**Technical Brief**

Derek A. Chapman, PhD1

Jong Hyung Lee, MS1

Roy T. Sabo, PhD2

*1Center on Society and Health, Virginia Commonwealth University*

*2Department of Biostatistics, Virginia Commonwealth University*



August 2021

**Project Overview**

This technical brief describes the data sources and methods utilized in the 2021 Behavioral Health Index (BHI) project completed by the VCU Center on Society and Health (CSH) with funding by the Virginia Department of Behavioral Health and Developmental Services (VDBHDS).

***Background***

The burden of illness (e.g., the number of people suffering from mental illness and/or substance use disorder) is ideally determined by multiplying the population by the prevalence rate, but accurate prevalence data are lacking for most behavioral health conditions. No population-representative surveys are undertaken in the Commonwealth to accurately ascertain the prevalence of any mental illness or substance use disorder at the county/city level or smaller geography. The data maintained by a Community Service Board (CSB) describes the prevalence among the client base served by the CSB and may overestimate or underestimate the actual prevalence rate in the CSB’s service area. People who seek care at the CSB may not be representative of the entire population.

The initial BHI developed in 2020 used mentally unhealthy days reported in the CDC’s Behavioral Risk Factor Surveillance System (BRFSS) to estimate the burden of illness based on data accessibility at the time of development. While the outcome measure of mentally unhealthy days provided a global measure of mental health, it did not fully estimate the population in need of CSB services as it did not capture individuals with substance use disorders and excluded children.

Thus, the purpose of the current project was to use the Virginia All-Payer Claims Database (APCD) to extend the prior BHI work by:

* Computing cruderatesof (a) severe emotional disturbance (SED) among children ages 8 to 17 years; (b); serious mental illness (SMI) among adults 18 years and older; and (c) substance use disorder (SUD) at any age for each Zip Code and Community Service Board (CSB) service area in Virginia
* Computing a CSB-level BHI in Virginia using Zip Code Tabulation Area (ZCTA)-level Census indicator data linked to Zip Code-level APCD data for each of the three health outcomes (SMI, SED, SUD)
* Computing a summary BHI which modeled Census indicator data against rates of having any one of the three outcomes (SMI, SED, or SUD)

***Computing Crude Rates***

***Zip Code Prevalence Rates***

*Numerators*

Counts of unique individuals with at least one claim[[1]](#footnote-1) by Zip Code for each of the following four categories were computed from the Virginia All Payers Claims Database (APCD) data for calendar year 2018:

* Severe Emotional Disturbance (SED) among children ages 8 to 17 years
* Serious Mental Illness (SMI) among adults ages 18+ years
* Substance Use Disorder (SUD) among individuals of any age
* Combined SED, SMI, or SUD (MULTI) among individuals of any age

A detailed list of ICD-10-CM codes used to define each category can be found in the Excel sheet entitled “BHI2\_SupplementalTables.xlsx”.

*Population denominators*

Population counts for each Zip Code Tabulation Area (ZCTA)[[2]](#footnote-2) were derived from the U.S. Census Bureau’s 2018 American Community Survey (ACS) 5-year data.[[3]](#footnote-3) Three different population counts were created for each ZCTA: total population for the denominator of SUD and MULTI rates, 18+ years for SMI, and 8 to 17 years for SED. Since population estimates for the 8 to 17 years category did not exist in the ACS data, we had to estimate the population size for the SED denominator. This was done by “removing” 5- to 7-year-olds from the ACS reported 5- to 17-year-old population count. Under the assumption that the population distribution was relatively equal in the 5-to-9-year age range, we added two-fifths of the ACS reported 5-to-9-year population (i.e., multiplied it by 0.4) to the ACS reported 10-to-17-year category to create the final 8-to-17-year counts.

*Crude Prevalence Rates*

Crude rates per 1,000 population for each ZCTA were computed by dividing the counts of unique persons with at least one claim for each outcome (SED, SMI, SUD, MULTI) by the corresponding population denominator.

According to the basic guidelines for computing binomial confidence intervals, it is recommended to avoid using the simple normal approximation method when a product value (proportion with the health outcome (*p*), proportion without the health outcome (*q*), and sample size (*n*) multiplied together) is between 0 and 5.[[4]](#footnote-4) Thus, we flagged any ZCTAs (assigning 1 if it met the criteria, otherwise 0) accordingly to compute more appropriate version of confidence intervals for the health outcome prevalence rates.

A simple normal approximation method (i.e., simple Wald interval) is appropriate when sample size (*n*) is large, and proportion of health outcome interest (*p*) is neither extremely small nor extremely large. This Wald interval is not appropriate for small values of *p* since it leads to inaccurate estimates.[[5]](#footnote-5) So, a reliable statistical approach was adopted to compute 95% confidence intervals for a binomial proportion using the Clopper-Pearson method based on the beta distribution.[[6]](#footnote-6)

The following equations were used especially for those localities with either very small or large proportion of health outcome occurrence (*p*) or relatively low population size (n < 40)5,[[7]](#footnote-7)

where

 = lower limit of the confidence interval

 = upper limit of the confidence interval

 *n* = sample size

 *k* = occurrence of health outcome cases

 *α* = level of significance

On the other hand, when sample size is sufficient based on the central limit theorem, adopting *Z*-value from the normal distribution is an appropriate approach for estimating the binomial distribution. However, as previously mentioned confidence interval may contain values below zero or above one when *p* is extremely close to 0 or 1, respectively. To avoid this situation, numerous statisticians have recommended adopting the adjusted Wald confidence interval.4 So, the following equation was applied for sample size greater than 40:

where

 =

 *n* = sample size

 *α* = level of significance

 = *Z* score for the level of significance

The margin of error for population estimates at the ZCTA level for small age groups (e.g., children) can be large and result in rates with extremely wide confidence intervals, particularly for sparsely populated ZCTAs. Thus, rates were not computed for any ZCTA where the population count was missing or <50 or the APCD counts exceeded the population total for a given behavioral health outcome. The number of suppressed rates varied by outcome based on the population denominator used:

SED: 161 (18.2%) of all ZCTAs

SMI: 46 (5.2%) of all ZCTAs

SUD: 17 (1.9%) of all ZCTAs

MULTI: 17 (1.9%) of all ZCTAs

***Community Service Board (CSB) Prevalence Rates***

To compute CSB-level rates, ZCTAs had to be linked to Virginia counties and independent cities. Because ZCTA boundaries can cut across multiple counties, this was not an exact match. Assignment of a county/independent city to each ZCTA was performed using the procedure *Geocorr 2018*[[8]](#footnote-8)  from the Missouri Census Data Center’s (MCDC) Geographic Correspondence Engine, which is based on a county allocation factor derived from the 2010 Census. The allocation factor provides the proportion of source population (e.g., ZCTA) in the target locality (e.g., county). Based on the MCDC data, each ZCTA was assigned to the county/independent city that contained the largest proportion of the ZCTA’s population. ZCTAs were contained entirely within a county/independent city boundary for the majority of ZCTAs (618 of 883 = 70.0%) and had at least 75% of their population in a single county/independent city for 89.3% (789 of 883) of ZCTAs.

CSB-level rates per 100,000 were computed for SED, SMI, SID, and MULTI by summing outcome counts and population totals for ZCTAs within each CSB’s county/independent city service area provided by the Virginia Department of Behavioral Health & Developmental Services (VDBHDS).[[9]](#footnote-9)  Using the methods described above, 95% confidence limits were computed for each rate.

***Computing Behavioral Health Indexes (BHI)***

***Indicators***

Twenty-five ZCTA-level Indicators from the 2018 American Community Survey (ACS) 5-year data were downloaded[[10]](#footnote-10) and used in the BHI modeling. Most indicators were presented as a numerator divided by the population denominator as a percent (numerator / denominator), with exceptions being indicators representing income levels or prices. Indicators were converted into z-scores by subtracting the mean value and dividing by the standard deviation. These z-scores were also aggregated to the community service board (CSB) level by taking population-weighted averages of those ZCTA-level z-scores that map into particular CSBs. A detailed list of indicators can be found in the Excel sheet entitled “BHI2\_SupplementalTables.xlsx”.

*Missing Data*

Missing data occurred in three ways: a ZCTA-level indicator was missing, the number of outcomes levels in a ZCTA was below X, or the underlying population of a ZCTA was below Y. These data were imputed using *k*-nearest neighbor (KNN).

***Weighted Quantile Sum Regression***

All indicators were coerced to be positively associated with the outcome in the same direction by multiplying some indicators by -1 (see Table 1 for list of bivariate correlations between indicators and outcomes). Some of the bivariate correlations may appear counterintuitive, but several indicators were likely markers for other factors. For example, the bivariate correlation between each indicator and the degree of rurality is shown in Table 2 – many indicators (e.g., access to public transportation, bachelor’s degree or higher education) had a strong negative association with rurality. It is important to note that the purpose of this analysis was to generate the best place-based model of each of the behavioral health outcomes and not to provide causal associations.

Weighted quantile sum regression was then performed at the ZCTA level, separately for each of the four outcomes, using four quantiles and with the specification that the overall index is positively associated with the outcome, indicating that a higher BHI score relates to an increased rate of SED, SMI, SUD or the composite MULTI outcome. This least-squares regression model was weighted and constrained, assigning weights to each indicator such that the maximum amount of variation in the outcome was explained; indicators that explained more variation in the outcome were assigned a higher weight; and the sum of all the indicators’ weights included in the model summed to one. Because all weights summed to one, the proportion of importance of an indicator variable included in the model is represented by its weight.

 This analysis produced the following results at the ZCTA level:

,,

where n is the number of ZCTAs, the is the z-score for the jth indicator for the ith ZCTA, and is the weight for the jth indicator.

Indicator weights for each of the Behavioral Health Indexes (SED, SMI, SUD, and MULTI) are shown in Table 3.

***Index Creation***

To allow index creation at the CSB level, the z-scores for the ZCTA-level indicators were aggregated to the CSB level via population weighted averaging. To predict the rate of SED, SMI, SUD, and MULTI for each CSB, these standardized CSB level indicators were multiplied by the respective standardized obtained from the ZCTA-level analysis and summed across indicators:

where H is the number of CSBs, and is the aggregated z-score for the jth indicator in the hth CSB. These calculated values were then converted into the Behavioral Health Index (BHI) by first transforming the predicted value for each CSB to fall within a range of 0 (worst) to 100 (best):

CSBs with higher BHI scores have better place-based conditions for a given outcome relative to those with lower BHI scores. The final BHI scores for each outcome (SED, SMI, SUD, and MULTI) are presented in Table 4.

***Table 1.*** *Bivariate correlations between and Behavioral Health Index (BHI) indicators and outcomes.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Indicator** | **SED** | **SMI** | **SUD** | **MULTI** |
| Median annual household income | -0.45 | -0.38 | -0.63 | -0.45 |
| Median home value of owner-occupied units | -0.37 | -0.39 | -0.59 | -0.43 |
| % with a high school diploma or higher (ages 25yrs and older) | -0.31 | -0.22 | -0.38 | -0.26 |
| % with bachelor's degree or higher (ages 25yrs and older) | -0.30 | -0.28 | -0.45 | -0.32 |
| % of population 15 yrs+ now married (excluding those who are separated) | -0.25 | 0.07 | -0.13 | 0.03 |
| % of workers 16 yrs and older who take public transport (bus, train, subway) to work | -0.21 | -0.41 | -0.39 | -0.42 |
| % of households with more than one occupant per room | -0.15 | -0.31 | -0.27 | -0.32 |
| Average travel time to work (minutes) | -0.11 | 0.00 | -0.19 | -0.02 |
| % of children who are uninsured | -0.09 | -0.19 | -0.13 | -0.19 |
| % of workers 16 yrs and older who walk or bike to work | -0.05 | -0.28 | -0.20 | -0.27 |
| % of households paying more than30% of income on housing | -0.03 | -0.16 | -0.12 | -0.17 |
| % of properties with 1 or more poor housing conditions (e.g., no plumbing, no kitchen, overcrowded, cost burdened) | -0.02 | -0.17 | -0.11 | -0.17 |
| % of population aged 25-64 who are unemployed | 0.03 | -0.09 | 0.03 | -0.08 |
| % of households paying more than 50% of income on housing | 0.05 | -0.11 | -0.03 | -0.10 |
| % of population in the same residencewithin the past 12 months | 0.08 | 0.30 | 0.26 | 0.31 |
| % of civilian noninstitutionalized population ages 19-64 who are uninsured | 0.15 | 0.02 | 0.20 | 0.05 |
| % of 3- and 4-year-olds not enrolled in school | 0.18 | 0.12 | 0.18 | 0.14 |
| % of workers 16 yrs and older who take a car, taxi, or motorcycle to work | 0.18 | 0.31 | 0.33 | 0.31 |
| % of households with no access to a vehicle | 0.19 | 0.01 | 0.17 | 0.05 |
| % of population under 18 yrs living below the poverty level | 0.21 | 0.10 | 0.25 | 0.13 |
| % of children living in households headed by a single parent | 0.25 | 0.16 | 0.29 | 0.20 |
| % of housing units built prior to 1950 | 0.29 | 0.15 | 0.33 | 0.20 |
| % of housing units that are vacant | 0.30 | 0.33 | 0.45 | 0.36 |
| % of population ages 18 to 64 yrs with household incomes below the poverty level | 0.32 | 0.17 | 0.37 | 0.21 |
| % of households receiving public assistance (cash public assistance or Food Stamps/SNAP) | 0.33 | 0.12 | 0.31 | 0.15 |

***Table 2.*** *Bivariate correlations between rurality and Behavioral Health Index (BHI) indicators.*

|  |  |
| --- | --- |
| **Indicator** | **Correlation with Rurality1** |
| % of housing units that are vacant | 0.55 |
| % of population in the same residence within the past 12 months | 0.52 |
| % of workers 16 yrs and older who take a car, taxi, or motorcycle to work | 0.31 |
| % of housing units built prior to 1950 | 0.24 |
| Average travel time to work (minutes) | 0.20 |
| % of 3- and 4-year-olds not enrolled in school | 0.20 |
| % of households receiving public assistance (cash public assistance or Food Stamps/SNAP) | 0.18 |
| % of population ages 18 to 64 yrs with household incomes below the poverty level | 0.15 |
| % of civilian noninstitutionalized population ages 19-64 who are uninsured | 0.14 |
| % of population 15 yrs+ now married (excluding those who are separated) | 0.14 |
| % of population under 18 yrs living below the poverty level | 0.04 |
| % of children living in households headed by a single parent | 0.02 |
| % of households with no access to a vehicle | -0.11 |
| % of population aged 25-64 who are unemployed | -0.12 |
| % of children who are uninsured | -0.18 |
| % of households paying more than 50% of income on housing | -0.26 |
| % of properties with 1 or more poor housing conditions (e.g., no plumbing, no kitchen, overcrowded, cost burdened) | -0.32 |
| % of workers 16 yrs and older who walk or bike to work | -0.34 |
| % of households with more than one occupant per room | -0.34 |
| % of households paying more than 30% of income on housing | -0.35 |
| % with a high school diploma or higher (ages 25yrs and older) | -0.35 |
| Median annual household income | -0.42 |
| Median home value of owner-occupied units | -0.44 |
| % with bachelor's degree or higher (ages 25yrs and older) | -0.50 |
| % of workers 16 yrs and older who take public transport (bus, train, subway) to work | -0.57 |

*12018 Education Demographic and Geographic Estimates (EDGE) Program locale codes (12 categories) used as the rurality measure. See* [*https://nces.ed.gov/programs/edge/docs/EDGE\_NCES\_LOCALE.pdf*](https://nces.ed.gov/programs/edge/docs/EDGE_NCES_LOCALE.pdf) *for more details.*

***Table 3.*** *Indicator**weights for Severe Emotional Disturbance (SED), Serious Mental Illness (SMI), Substance Use Disorder (SUD) and SED/SMI/SUD (MULTI).*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Indicator | SED Weight | SMI Weight | SUD Weight | MULTI Weight |
| % of population ages 18 to 64 yrs with household incomes below the poverty level | ≈ 0.0 | ≈ 0.0 | 0.062 | ≈ 0.0 |
| % of households paying more than30% of income on housing | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 |
| % of households paying more than 50% of income on housing | 0.105 | ≈ 0.0 | 0.064 | ≈ 0.0 |
| % with bachelor's degree or higher (ages 25yrs and older) | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 |
| % of population under 18 yrs living below the poverty level | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 | 0.002 |
| % of workers 16 yrs and older who walk or bike to work | 0.096 | 0.131 | 0.142 | 0.146 |
| % of workers 16 yrs and older who take a car, taxi, or motorcycle to work | 0.041 | ≈ 0.0 | 0.039 | ≈ 0.0 |
| % of workers 16 yrs and older who take public transport (bus, train, subway) to work | 0.129 | 0.144 | ≈ 0.0 | 0.132 |
| Average travel time to work (minutes) | 0.054 | 0.047 | 0.044 | 0.042 |
| % of households with more than one occupant per room | ≈ 0.0 | 0.089 | 0.109 | 0.081 |
| % with a high school diploma or higher (ages 25yrs and older) | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 |
| % of population in the same residencewithin the past 12 months | 0.075 | 0.059 | 0.113 | 0.087 |
| % of population 15 yrs+ now married (excluding those who are separated) | ≈ 0.0 | 0.112 | ≈ 0.0 | 0.104 |
| Median home value of owner-occupied units | ≈ 0.0 | 0.001 | 0.019 | ≈ 0.0 |
| Median annual household income  | 0.042 | 0.083 | 0.004 | 0.069 |
| % of 3- and 4-year-olds not enrolled in school | 0.063 | 0.005 | ≈ 0.0 | 0.015 |
| % of households with no access to a vehicle | 0.089 | 0.033 | ≈ 0.0 | ≈ 0.0 |
| % of housing units built prior to 1950 | 0.107 | 0.026 | 0.146 | 0.046 |
| % of households receiving public assistance (cash public assistance or Food Stamps/SNAP) | ≈ 0.0 | 0.015 | 0.101 | ≈ 0.0 |
| % of properties with 1 or more poor housing conditions (e.g., no plumbing, no kitchen, overcrowded, cost burdened) | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 | ≈ 0.0 |
| % of children living in households headed by a single parent | 0.194 | 0.105 | ≈ 0.0 | 0.120 |
| % of population aged 25-64 who are unemployed | ≈ 0.0 | 0.114 | ≈ 0.0 | 0.122 |
| % of civilian noninstitutionalized population ages 19-64 who are uninsured | ≈ 0.0 | ≈ 0.0 | 0.003 | ≈ 0.0 |
| % of children who are uninsured | ≈ 0.0 | ≈ 0.0 | 0.155 | 0.006 |
| % of housing units that are vacant | 0.003 | 0.036 | ≈ 0.0 | 0.029 |

***Table 4.*** *Community Service Board (CSB)-level index scores1 from weighted quantile sum regression.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Community Service Board (CSB) | SED Index Score | SMI Index Score | SUD Index Score | MULTI Index Score |
| Alexandria CSB | 88 | 81 | 100 | 84 |
| Alleghany Highlands Community Services | 16 | 2 | 0 | 3 |
| Arlington CSB | 100 | 100 | 88 | 100 |
| Blue Ridge Behavioral Healthcare | 29 | 13 | 35 | 16 |
| Chesapeake CSB | 34 | 15 | 62 | 21 |
| Chesterfield CSB | 50 | 16 | 66 | 21 |
| Colonial Behavioral Health | 56 | 19 | 74 | 26 |
| Crossroads Services Board | 28 | 13 | 39 | 16 |
| Cumberland Mountain Community Services | 25 | 5 | 19 | 9 |
| Danville-Pittsylvania Community Services | 15 | 6 | 17 | 9 |
| Dickenson County Behavioral Health Services | 13 | 4 | 4 | 7 |
| District 19 CSB | 18 | 16 | 38 | 21 |
| Eastern Shore Community Services | 4 | 0 | 26 | 0 |
| Fairfax-Falls Church CSB | 87 | 56 | 97 | 61 |
| Goochland-Powhatan Community Services | 52 | 15 | 77 | 18 |
| Hampton-Newport News CSB | 23 | 24 | 52 | 29 |
| Hanover County CSB | 57 | 12 | 61 | 17 |
| Harrisonburg-Rockingham CSB | 36 | 22 | 69 | 27 |
| Henrico Area MH & Developmental Services | 36 | 17 | 51 | 21 |
| Highlands Community Services | 34 | 7 | 30 | 13 |
| Horizon Behavioral Health | 36 | 11 | 47 | 15 |
| Loudoun County CSB | 84 | 39 | 90 | 44 |
| Middle Peninsula-Northern Neck CSB | 36 | 7 | 47 | 11 |
| Mt. Rogers Community MH & MR Services Board | 25 | 2 | 30 | 5 |
| New River Valley Community Services | 43 | 27 | 48 | 33 |
| Norfolk CSB | 12 | 39 | 44 | 43 |
| Northwestern Community Services | 37 | 15 | 53 | 18 |
| Piedmont Community Services | 27 | 7 | 35 | 12 |
| Planning District 1 Behavioral Health Services | 19 | 8 | 8 | 12 |
| Portsmouth Dept. of Behavioral Healthcare Services | 7 | 20 | 26 | 22 |
| Prince William County Community Services Bd. | 72 | 41 | 93 | 47 |
| Rappahannock Area CSB | 64 | 29 | 78 | 35 |
| Rappahannock-Rapidan CSB | 53 | 22 | 68 | 26 |
| Region Ten CSB | 49 | 28 | 70 | 33 |
| Richmond Behavioral Health Authority | 0 | 35 | 24 | 36 |
| Rockbridge Area Community Services | 28 | 13 | 37 | 17 |
| Southside CSB | 11 | 0 | 17 | 2 |
| Valley CSB | 31 | 8 | 42 | 12 |
| Virginia Beach Department of Human Services | 48 | 18 | 71 | 25 |
| Western Tidewater CSB | 33 | 13 | 42 | 18 |

*1Index scores were rescaled to range from 0 (worst) to 100 (best).*

1. APCD included claims from inpatient, outpatient, and emergency department settings. APCD does not include cash-for-service, indigent care, or Veteran’s Administration care. [↑](#footnote-ref-1)
2. Zip Code Tabulation Areas (ZCTAs) are created by the U.S. Census Bureau based on census blocks and were designed to correspond to Zip Code areas. Over 70% of ZCTAs share at least 80% of their area with their corresponding Zip Codes. [↑](#footnote-ref-2)
3. Data were downloaded from the American FactFinder website, which was decommissioned on March 31, 2020. These American Community Survey data (table B01001) are now available for download at <https://data.census.gov/cedsci/>. [↑](#footnote-ref-3)
4. Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical science*, 101-117. [↑](#footnote-ref-4)
5. Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. The American Statistician, 52(2), 119-126. [↑](#footnote-ref-5)
6. Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404-413. [↑](#footnote-ref-6)
7. Thulin, M. (2014). The cost of using exact confidence intervals for a binomial proportion. *Electronic Journal of Statistics*, *8*(1), 817-840. [↑](#footnote-ref-7)
8. MCDC Data Applications Website. <https://mcdc.missouri.edu/applications/geocorr2018.html>. Accessed March 1, 2021. [↑](#footnote-ref-8)
9. Virginia Association of Community Services Boards, Inc. Website. <https://dbhds.virginia.gov/assets/doc/BH/oss/CSBOverviewMar2018.pdf>. Accessed March 15, 2021. [↑](#footnote-ref-9)
10. Data were downloaded from the American FactFinder website, which was decommissioned on March 31, 2020. These American Community Survey data are now available for download at <https://data.census.gov/cedsci/>. [↑](#footnote-ref-10)