Health Information – A Regional Look

The following health information maps illustrate health “hot spots” in the Hudson Valley region. Hot spots are geographic areas, represented by zip code clusters, that have higher hospitalization rates for various diseases and health conditions than the surrounding areas. Relative hotspot intensity increases as the cluster colors go from yellow to red.

Locating and focusing on “hot spots” will allow us to better coordinate local health care and provide services and supports that will improve health for all residents of the Hudson Valley.

The zip code clusters are based on SPARCS inpatient discharges totaled over 2008-2012, and the total population is from the 2010 census. For a more detailed explanation, please see Appendix A.

To learn more about the Community Needs Assessment and the data we’ve presented, contact:

Deborah Viola, PhD
Vice President
Director, Health Services Research & Data Analytics
Center for Regional Healthcare Innovation
violad@wcmc.com | 914.326.4203

Thao M. Doan, MPH
Research Analyst
Center for Regional Healthcare Innovation
doant@wcmc.com | 914.326.4207
Mood Disorders

Mood Disorders unique patients
Relative Risk

- 2.21 - 3.29
- 1.66 - 2.20
- 1.53 - 1.65
- 1.26 - 1.52
- 1.19 - 1.25

Legend:
- Study Area Counties
- Other NYS Counties
- Other NYS Zip Codes
- Surrounding States
Chronic Obstructive Pulmonary Disease (COPD)
Asthma

Asthma Relative Risk

- 2.89 - 6.33
- 1.99 - 2.88
- 1.51 - 1.98
- 1.21 - 1.50
- 1.20

- Study Area Counties
- Other NYS Counties
- Other NYS Zip Codes
- Surrounding States
Breast Cancer unique patients

Relative Risk

- 1.29 - 1.38
- 1.25 - 1.28
- 1.24

Legend:
- Study Area Counties
- Other NYS Counties
- Other NYS Zip Codes
- Surrounding States
Cervical Cancer

Cervical Cancer unique patients

Relative Risk

- 1.82 - 5.45
- 1.49 - 1.81
- 1.48

- Study Area Counties
- Other NYS Counties
- Other NYS Zip Codes
- Surrounding States

Counties:
- Sullivan
- Ulster
- Dutchess
- Orange
- Putnam
- Rockland
- Westchester
The following clusters of perinatal variables are based on NYS vital records totaled over 2010-2012.

Denominators are total live births and no data are reported for ZIP code areas with fewer than 10 births during the 3-year time period.
Premature (< 37 weeks gestation) Births, 2010-2012
Low Birth Weight (˂2500 g) Births, 2010-2012

Low Birth Weight Clusters

Relative Risk

1.74
1.35

Study Area Counties
Other NYS Counties
Other NYS Zip Codes
Surrounding States

0 Miles 10 Miles
Births associated with Late or No Prenatal Care, 2010-2012
Births covered by Public Insurance (Medicaid, etc.) or the Hospital, 2010-2012
The following clusters of C-section rates are based on Inpatient Records from 2008 through July 2013. Denominators are total births.
Total Cesarean Section Deliveries
(based on Inpatient Records from 2008 through July 2013)
APPENDIX A

The Spatial Scan Statistic, as encoded in SATSCAN

The Spatial Scan Statistic provides a methodology for identifying geo-spatial clusters of statistically elevated outcomes. This works by considering all possible clusters of spatially contiguous units. For example, consider the seven counties of the lower Hudson Valley that contain approximately 270 ZIP codes, after merging enclosed ZIPS into their respective enclosing ZIPS. Each ZIP code has data for a numerator (i.e. number of SPARCS hospital discharges for a given outcome) and a denominator, typically a census enumeration of the population that is at risk for the outcome. The algorithm then works as follows:

1. A rate of for the outcome is computed for the whole region as \( P= \text{total cases} / \text{total population for the region} \) (i.e. lower Hudson Valley)
2. For each of \( i=1, \ldots, n \) ZIP codes, the expected number of cases is computed by multiplying the ZIP code’s population by the region-wide rate. In other words, \( E_i=P \times n_i \).
3. Every ZIP code is now represented by its observed caseload, \( Y_i \) and its expected caseload, \( E_i \). Therefore the risk associated with each of \( i=1, \ldots, n \) ZIP codes = \( Y_i / E_i \). The relative risk is then computed as the ratio of the risk in the ZIP code to the risk every else in the region that is outside of the ZIP code. The data used to compute a relative risk are also used to compute a test statistic.
4. For each of \( i=1, \ldots, n \) ZIP codes, the test statistic is used to evaluate the competing hypotheses of \( H_{null} \): the ZIP code risk is no different than the risk everywhere outside of the ZIP code, versus \( H_{alt} \): the ZIP code risk is higher than the risk everywhere outside of the ZIP code.
5. For each ZIP code, after evaluating itself as a “single-ZIP” cluster, it is combined with its nearest neighbor and treated as a “two-ZIP” potential cluster. Steps 2 through 4 are then repeated to test for significance of this cluster that consists of two ZIP codes combined.
6. Step 5 is then repeated multiple times to evaluate many potential clusters, with each one growing in size (3-ZIPS, 4 ZIPS, etc.), but centered on the initial ZIP code. Some upper limit is imposed such as restricting the maximum cluster size to capture no more than 25% of the population at risk within the region.
7. Step 6 is then repeated for each of \( i=1, \ldots, n \) ZIP codes in the whole region.

The end result is that very many potential clusters are evaluated, whereas the clusters may be centered on any ZIP code and may vary in size from one ZIP to many. Each cluster is represented by 1) its relative risk and 2) statistical significance by the probability of incorrectly rejecting the null hypothesis in step 4, otherwise known as the “p-value”.

Results are geo-visualized by mapping only those clusters that are statistically significant and do not geographically overlap with other clusters. Within this subset of clusters, the cluster relative risk is used to color-code each cluster in a map. A fairly liberal \( p \)-value of \( < 0.10 \) is used for including statistically significant clusters. This allows inclusion of clusters with up to a 10% chance of concluding they have higher risk than the rest of the region when in truth they do not (incorrectly rejecting the null hypothesis in step 4).

The SATSCAN software used to apply the spatial scan statistic uses a simulation approach to hypothesis testing in a way that overcomes the well-known problem of inflated false positive rates from multiple hypotheses testing.