

May 2013

# Comparisons Among Health Behavior Surveys: Implications for the Design of Informatics Infrastructures that Support Comparative Effectiveness Research

Sunmoo Yoon

*Columbia University, sy2102@cumc.columbia.edu*

Adam Wilcox

*Columbia University*

Suzanne Bakken

*Columbia University*

Follow this and additional works at: <http://repository.academyhealth.org/egems>



Part of the [Health Services Research Commons](#)

## Recommended Citation

Yoon, Sunmoo; Wilcox, Adam; and Bakken, Suzanne (2013) "Comparisons Among Health Behavior Surveys: Implications for the Design of Informatics Infrastructures that Support Comparative Effectiveness Research," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 1: Iss. 1, Article 9.

DOI: <http://dx.doi.org/10.13063/2327-9214.1021>

Available at: <http://repository.academyhealth.org/egems/vol1/iss1/9>

This Methods Case Study is brought to you for free and open access by the the EDM Forum Products and Events at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

---

# Comparisons Among Health Behavior Surveys: Implications for the Design of Informatics Infrastructures that Support Comparative Effectiveness Research

## **Abstract**

**Introduction:** To address the electronic health data fragmentation that is a methodological limitation of comparative effectiveness research (CER), the Washington Heights Inwood Informatics Infrastructure for Comparative Effectiveness Research (WICER) project is creating a patient-centered research data warehouse (RDW) by linking electronic clinical data (ECD) from New York Presbyterian Hospital's clinical data warehouse with ECD from ambulatory care, long-term care, and home health settings and the WICER community health survey (CHS). The purposes of the research were to identify areas of overlap between the WICER CHS and two other surveys that include health behavior data (the Behavioral Risk Factor Surveillance System (BRFSS) Survey and the New York City Community Health Survey (NYC CHS)) and to identify gaps in the current WICER RDW that have the potential to affect patient-centered CER.

**Methods:** We compared items across the three surveys at the item and conceptual levels. We also compared WICER RDW (ECD and WICER CHS), BRFSS, and NYC CHS to the County Health Ranking framework.

**Results:** We found that 22 percent of WICER items were exact matches with BRFSS and that there were no exact matches between WICER CHS and NYC CHS items not also contained in BRFSS.

**Conclusions:** The results suggest that BRFSS and, to a lesser extent, NYC CHS have the potential to serve as population comparisons for WICER CHS for some health behavior– related data and thus may be particularly useful for considering the generalizability of CER study findings. Except for one measure related to health behavior (motor vehicle crash deaths), the WICER RDW's comprehensive coverage supports the mortality, morbidity, and clinical care measures specified in the County Health Ranking framework but is deficient in terms of some socioeconomic factors and descriptions of the physical environment as captured in BRFSS. Linkage of these data in the WICER RDW through geocoding can potentially facilitate patient-centered CER that integrates important socioeconomic and physical environment influences on health outcomes. The research methods and findings may be relevant to others interested in either integrating health behavior data into RDWs to support patient-centered CER or conducting population-level comparisons.

## **Acknowledgements**

Funding was provided by R01 HS019853 from the Agency for Healthcare Research and Quality; Title: Washington Heights/Inwood Informatics Infrastructure for Community-Centered Comparative Effectiveness Research (WICER).

## **Keywords**

methods, surveys

## **Disciplines**

Health Services Research

---

**Creative Commons License**



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

# Comparisons among Health Behavior Surveys: Implications for the Design of Informatics Infrastructures That Support Comparative Effectiveness Research

Sunmoo Yoon, RN, PhD; Adam B. Wilcox, PhD; Suzanne Bakken, RN, DNSc  
Columbia University

## Abstract

**Introduction:** To address the electronic health data fragmentation that is a methodological limitation of comparative effectiveness research (CER), the Washington Heights Inwood Informatics Infrastructure for Comparative Effectiveness Research (WICER) project is creating a patient-centered research data warehouse (RDW) by linking electronic clinical data (ECD) from New York Presbyterian Hospital's clinical data warehouse with ECD from ambulatory care, long-term care, and home health settings and the WICER community health survey (CHS). The purposes of the research were to identify areas of overlap between the WICER CHS and two other surveys that include health behavior data (the Behavioral Risk Factor Surveillance System (BRFSS) Survey and the New York City Community Health Survey (NYC CHS)) and to identify gaps in the current WICER RDW that have the potential to affect patient-centered CER.

**Methods:** We compared items across the three surveys at the item and conceptual levels. We also compared WICER RDW (ECD and WICER CHS), BRFSS, and NYC CHS to the County Health Ranking framework.

**Results:** We found that 22 percent of WICER items were exact matches with BRFSS and that there were no exact matches between WICER CHS and NYC CHS items not also contained in BRFSS.

**Conclusions:** The results suggest that BRFSS and, to a lesser extent, NYC CHS have the potential to serve as population comparisons for WICER CHS for some health behavior-related data and thus may be particularly useful for considering the generalizability of CER study findings. Except for one measure related to health behavior (motor vehicle crash deaths), the WICER RDW's comprehensive coverage supports the mortality, morbidity, and clinical care measures specified in the County Health Ranking framework but is deficient in terms of some socioeconomic factors and descriptions of the physical environment as captured in BRFSS. Linkage of these data in the WICER RDW through geocoding can potentially facilitate patient-centered CER that integrates important socioeconomic and physical environment influences on health outcomes. The research methods and findings may be relevant to others interested in either integrating health behavior data into RDWs to support patient-centered CER or conducting population-level comparisons.

## Introduction

The Washington Heights/Inwood Informatics Infrastructure for Comparative Effectiveness Research (WICER) project adheres to community-based participatory research principles to build an infrastructure that supports comparative effectiveness research (CER) for the purpose of improving community health in Washington Heights and Inwood, a primarily Latino community of New York City. To address the electronic health data fragmentation that is a methodological limitation of CER,<sup>1,2</sup> WICER includes a patient-centered research data warehouse (RDW) that links electronic clinical data (ECD) from New York Presbyterian Hospital's (NYP) clinical data warehouse (CDW) with ECD from ambulatory care, long-term care, and home health settings and with the WICER community health survey (CHS). The intent of RDW data integration across settings and data types is to enable the conduct of CER that can benefit from the unique contributions of a variety of data sources.

A unique aspect of WICER is the CHS, which stresses the social determinants of health and the health behaviors of key relevance to hypertension; the latter is WICER's CER focus. The purposes of this research were to identify areas of overlap between the WICER CHS and two other surveys that include health behavior data (the

Behavioral Risk Factor Surveillance System (BRFSS) Survey<sup>3</sup> and the New York City Community Health Survey<sup>4</sup> (NYC CHS)) and to identify gaps in the current WICER RDW with the potential to influence the conduct of patient-centered CER. The research posed two questions: (1) Can BRFSS and/or NYC CHS serve as national or local points of comparison for WICER CHS in terms of key determinants of health (e.g., health behaviors) or patient-reported outcomes (e.g., quality of life)? (2) How do the elements of all three surveys and WICER ECD align with data required by the County Health Rankings?

Understanding the concepts/domains of each overlapping or non-overlapping survey is an essential prerequisite to comparing the findings from appropriate domains of BRFSS and NYC CHS once population-level estimates are available from the WICER CHS. A comparison of the concepts/domains of each survey provides a quick reference for common measures of determinants of health with respect to hypertension, which may be germane to researchers interested in understanding how their CER research populations compare with national surveys on measures of behavioral health and so forth. Moreover, the alignment of WICER RDW (CHS and ECD), BRFSS, and NYC CHS with the County

**Table 1. Selected Characteristics of WICER CHS, BRFSS, and NYC CHS**

	WICER CHS	BRFSS	NYC CHS
Number of items extracted for analysis	202	381	128
Sampling	Random, cluster, and snowball sampling in community plus convenience sampling in ambulatory care clinics	Random	Random
Sample size	6,000	451,075	8,640 277 (Washington Heights/Inwood)
Frequency	Baseline and one-year follow-up	Annual	Annual
Survey administration	Face-to-face interview	Telephone	Telephone
Linkage	Linked with ECD from several care settings	Not linked	Not linked

Health Rankings model illustrates an approach that may be useful to other researchers interested in aggregating data across surveys and ECD in order to examine population health. Finally, the research addresses the context of each survey’s administration, which may affect the ability to compare measures across surveyed populations.

## Methods

### Comparisons between WICER Community Survey and BRFSS and NYC CHS

*Survey descriptions.* The WICER CHS was designed to assess the behaviors of a specific community in the context of the overall goals of the WICER project, focusing on CER studies for hypertension diagnosis, treatment, and care management. Domain experts carefully selected the WICER survey items in conjunction with community input through focus groups and early field testing. The WICER survey includes items related to social determinants of health, health behaviors, and measures of body mass index, blood pressure, and waist circumference, along with selected items from the Patient Reported Outcomes Measurement Information System (PROMIS), which is a national initiative aimed at designing highly reliable and precise standardized measures of patient-reported health status outcomes. Bilingual community health workers collect WICER data via interview (English or Spanish) in community settings, thus facilitating collection of data from hard-to-reach groups such as monolingual Spanish speakers and undocumented immigrants.

The 28-year-old BRFSS is the nation’s largest and most thorough health survey system that tracks health conditions and risk behaviors. It collects data through a random telephone survey (one adult per household). The NYC CHS is a locally adapted version of BRFSS, with data collected in a similar manner.

*Comparisons at item level.* We extracted 202 items from the WICER CHS, 381 items from BRFSS, and 128 from the NYC CHS. We excluded BRFSS and NYC CHS items that were not part of the WICER domains (e.g., disaster preparedness). First, we categorized each WICER CHS and BRFSS item according to the data

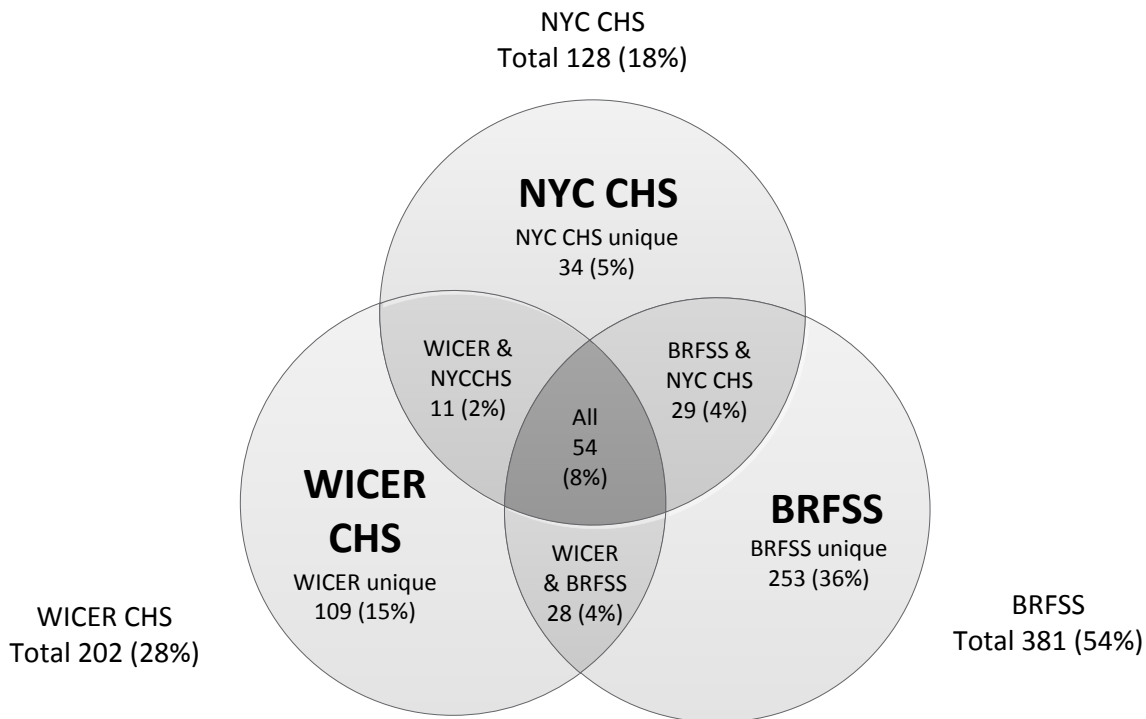
sources in which it appeared: WICER only, BRFSS only, or Both. Second, two researchers independently rated the quality of the match (exact or partial) for those items categorized as Both and discussed differences in ratings of the quality of the match until they reached consensus. Third, we examined NYC CHS items to identify items not contained in BRFSS. Fourth, we compared unique items in NYC CHS with WICER CHS by using the same methods applied to the BRFSS comparisons.

*Comparisons at conceptual level.* We compared WICER CHS, BRFSS, and NYC CHS conceptually by using Evans and Stoddart’s determinants of health framework,<sup>6</sup> which includes individual behaviors, social environment, physical environment, genetic endowment, function, disease, health care, well-being, and prosperity. One researcher mapped the items to the framework, and the other authors reviewed the mapping.

### Comparison of ECD and Survey Data Categories with County Health Ranking Framework

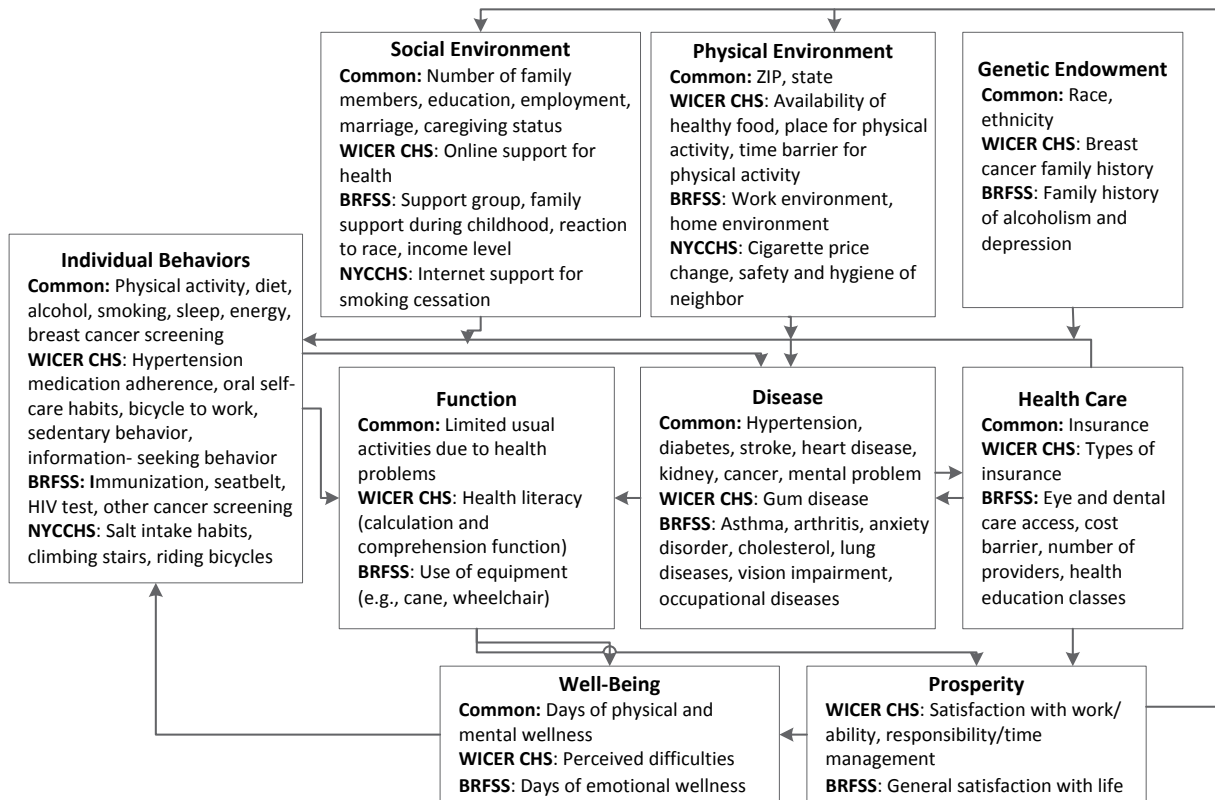
Motivated by WICER’s goal of improving the health of the community in Washington Heights and Inwood, we compared the categories of WICER RDW data (ECD plus WICER CHS) to the population health–focused County Health Ranking framework.<sup>5</sup> As compared to earlier health determinants models, the framework proposes quantification of the impact of factors affecting health outcomes: health behaviors (30 percent), clinical care (20 percent), socioeconomic factors (40 percent), and physical environment (10 percent). We also compared BRFSS and NYC CHS to the County Health Ranking framework to determine if gaps discovered in the comparison of WICER RDW with the framework could be addressed through the addition of BRFSS and/or NYC CHS to the RDW. One researcher performed the initial comparison between the framework and WICER RDW, BRFSS, and NYC CHS. The other authors then reviewed the comparison, with revisions made by consensus.

Figure 1. Degree of overlap of items in the WICER CHS, the NYC CHS, and the BRFSS.



Note: Degree of overlap of items in the WICER CHS (community), NYC CHS (city) and BRFSS (nation): 711 total items (denominator)=WICER CHS (202 items) + NYC CHS (128 items) + BRFSS (381 items)

Figure 2. Sample common and unique survey items relevant to hypertension within WICER CHS (community), BRFSS (nation), and NYC CHS (city) according to the determinants of health framework (Evans and Stoddart, 2003).



## Results

### Comparisons between WICER Community Survey and BRFSS and NYC CHS

In Table 1, we summarize selected characteristics of WICER CHS (community), BRFSS (national), and NYC CHS (city).

*Comparisons at item level.* In Figure 1, we summarize the unique items and overlap among WICER CHS, BRFSS, and NYC CHS. Of the 583 items extracted from BRFSS and WICER CHS, 31 percent were categorized as Both. The initial percentage of agreement for quality of matches was 78 percent; the data in the figure are the results after achievement of consensus. One-quarter of the matches were exact. Other matches were partial because of differences in the granularity of items. In some instances, WICER items were more detailed; in other instances, BRFSS items were more specific. The 202 WICER CHS items included 120 (59 percent) unique items, 45 (22 percent) exact matches with BRFSS items, and 37 (18 percent) partial matches with BRFSS items. The 381 BRFSS items comprised 282 (74 percent) unique items, 45 (12 percent) exact matches with WICER CHS, and 54 (14 percent) partial matches with WICER CHS. Some partially matched items were gender (BRFSS: 2 categories; WICER: 4 categories, including transgender) and alcohol (BRFSS: days of consuming any alcohol in past 30 days; WICER: types of alcohol, frequency in past six months).

Forty-five (35 percent) of 128 NYC CHS items were unique as compared to BRFSS. Of the 45 unique items, 11 items (9 percent) appeared in WICER CHS. All matches were partial because of differences in the granularity of items. In most instances, NYC CHS items were more detailed in their response choices (e.g., types of insurance, born state or country, sexual identity).

*Comparisons at conceptual level.* In Figure 2, we show selected survey items in WICER, BRFSS, and NYC CHS according to the nine domains of health determinants from Evans and Stoddart's model. In the individual behavior domain, the surveys cover physical activity, diet, alcohol use, smoking, sleep habits, and breast cancer screening behaviors. However, hypertensive medication adherence behavior, oral care behavior, sedentary behavior, and Internet information-seeking behavior are unique to WICER. Detailed information related to immunization behavior, HIV test behavior, and cancer screening behavior is unique to BRFSS. Salt intake behavior, climbing stairs behavior, and bicycling behavior are unique to NYC CHS. In the social environment domain, shared items include number of family members, level of education, employment, and marriage status. In contrast to WICER survey questions about seeking online social support and NYC CHS questions about Internet support for smoking cessation, BRFSS and includes items that address the longitudinal social environment such as family support during childhood. In terms of the health care domain, all surveys include items related to health care access. For the disease domain, while all surveys address chronic health problems, BRFSS asks about more disease conditions such as asthma, arthritis, high cholesterol, lung diseases, and occupation-related diseases. In terms of the function domain,

**Figure 3. Comparison of data elements in WICER RDW (Electronic Clinical Data-ECD, WICER Community Health Survey-CHS), BRFSS (nation) and NYC CHS (city) summarized according to the County Health Ranking framework (RWJF and UWPPI, 2012).**

	WICER RDW					
	ECD	CHS	BRFSS	NYCCHS		
<b>Health Outcomes</b>	Mortality (50%)	-Premature death	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Morbidity (50%)	-Self-report fair or poor health	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	-Poor physical health days		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	-Poor mental health days		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	-Low birthweight	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Health Behaviors (30%)	-Adult smoking	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Adult obesity	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Physical inactivity	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Excessive drinking	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Motor vehicle crash deaths	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	-Sexually transmitted infections	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	-Teen birth rate	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Clinical Care (20%)	-Uninsured	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Primary care physicians	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	-Preventable hospital stays	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	-Diabetes screening	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
-Mammography screening	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
Socioeconomic Factors (40%)	-High school graduation	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Some college	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Unemployment	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Children in poverty	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Inadequate social support	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	-In single-parent households	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
-Violent crime	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		
Physical Environment (10%)	-Indoor air quality	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	-Excessive sun exposure	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	-Access to recreational facilities	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	-Limited access to healthy foods	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
	-Fast food restaurants	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

WICER includes health literacy and numeracy questions that assess comprehension and calculation functions, but BRFSS asks about the use of devices such as a cane or wheelchair.

### Comparison of ECD and Survey Data Categories with County Health Ranking Framework

The WICER RDW (ECD and WICER CHS) includes data related to all County Health Ranking measures for mortality, morbidity, and clinical care (Figure 3). However, the WICER RDW contains data related to only 7 of 8 health behaviors, 3 of 7 socioeconomic factors, and 1 of 5 physical environment measures. Within the WICER RDW, WICER CHS uniquely contributes data related to self-report of fair or poor health, poor physical health days, poor mental health days, physical inactivity, and limited access to healthy foods. BRFSS includes data for the 9 measures not in the WICER RDW: inadequate social support, motor vehicle crash deaths, children in poverty, single-parent households, violent crime, indoor air quality, excessive sun exposure, access to recreational facilities, and fast food restaurants. NYC CHS also includes items related to the latter two measures. Although BRFSS is the most comprehensive, no single survey contains data related to all 29 measures.

## Discussion

In terms of identification of overlap between WICER CHS and BRFSS, 22 percent of WICER items were exact matches with BRFSS items. There were no exact matches between WICER CHS and NYC CHS items not also contained in BRFSS, suggesting that BRFSS and, to a lesser extent NYC CHS, have the potential to serve as population comparisons for WICER CHS for some health behavior–related data and patient-reported outcomes. Such comparisons may be particularly useful in considering the generalizability of CER study findings. Moreover, specific to the WICER project, the overlap facilitates the presentation of data to survey participants in the context of normative data from outside the Washington Heights/Inwood community. Although BRFSS captures data related to important health behaviors that are not in WICER CHS (HIV testing and cancer screening, for example), such data are available through ECD in the WICER RDW. Given the intense burden of data collection for WICER CHS and other studies of similar scope, priority for addition of items must be given to items that cannot be obtained through ECD.

The comparison of data categories of WICER RDW (ECD and WICER CHS), BRFSS, and NYC CHS to the County Health Ranking framework indicates that the WICER RDW’s comprehensive coverage can support mortality, morbidity, and clinical care measures specified in the framework and all but one measure related to health behavior (motor vehicle crash deaths). This finding is notable given that the WICER CHS sample primarily reflects individuals who are typically under-represented in surveys (e.g., Latino ethnicity, inadequate or marginal health literacy). In contrast, the WICER RDW is deficient in terms of some socioeconomic factors and descriptions of the physical environment captured in BRFSS. Linkage of these data in the WICER RDW through geocoding could facilitate patient-centered CER that integrates important socioeconomic and physical environment influences on health outcomes.

The limitations of this study include its focus on a single RDW and one project-specific CHS. Findings may differ in other contexts. Moreover, our work on population-level comparisons of data is in progress and has not yet advanced beyond the conceptual level.

## Conclusion

The research methods and findings may be relevant to others interested in integrating health behavior, selected patient-reported outcomes, and environmental data from population surveys into RDWs to support patient-centered CER. Researchers should consider how best to take advantage of existing population survey data to complement project-specific data collection and analysis and ECD data typically contained in RDWs for population-level comparisons.

## References

1. Alemayehu D. et al. “Considerations on the Use of Patient-Reported Outcomes in Comparative Effectiveness Research.” *J Manage Care Pharm*, Vol. 17, No. 9, November-December 2011, pp. S27-S33.
2. Lohr K.N. “Comparative Effectiveness Research Methods Symposium Overview and Summary.” *Med Care*, Vol. 48, No. 6, June 2010, pp. S3-S6.
3. Behavioral Risk Factor Surveillance System Survey Data. Atlanta: U.S. Department of Health and Human Services. [http://www.cdc.gov/brfss/technical\\_info-data/surveydata.htm](http://www.cdc.gov/brfss/technical_info-data/surveydata.htm), 2012.
4. New York City Community Health Survey (NYCCHS). <http://www.nyc.gov/html/doh/html/survey/chsdata.shtml>, 2012.
5. Rohan A.M. et al. “Using the Wisconsin County Health Rankings to Catalyze Community Health Improvement.” *J Public Health Manag Pract*, Vol. 15, No. 1, January-February 2009, pp. 24-32.
6. Evans R.G. and G.L. Stoddart. “Consuming Research, Producing Policy?” *Am J Public Health*, Vol. 93, No. 3, March 2003, pp. 371-9.

## Acknowledgment

The preparation of this manuscript was supported by R01HS019853 from the Agency for Healthcare Research and Quality