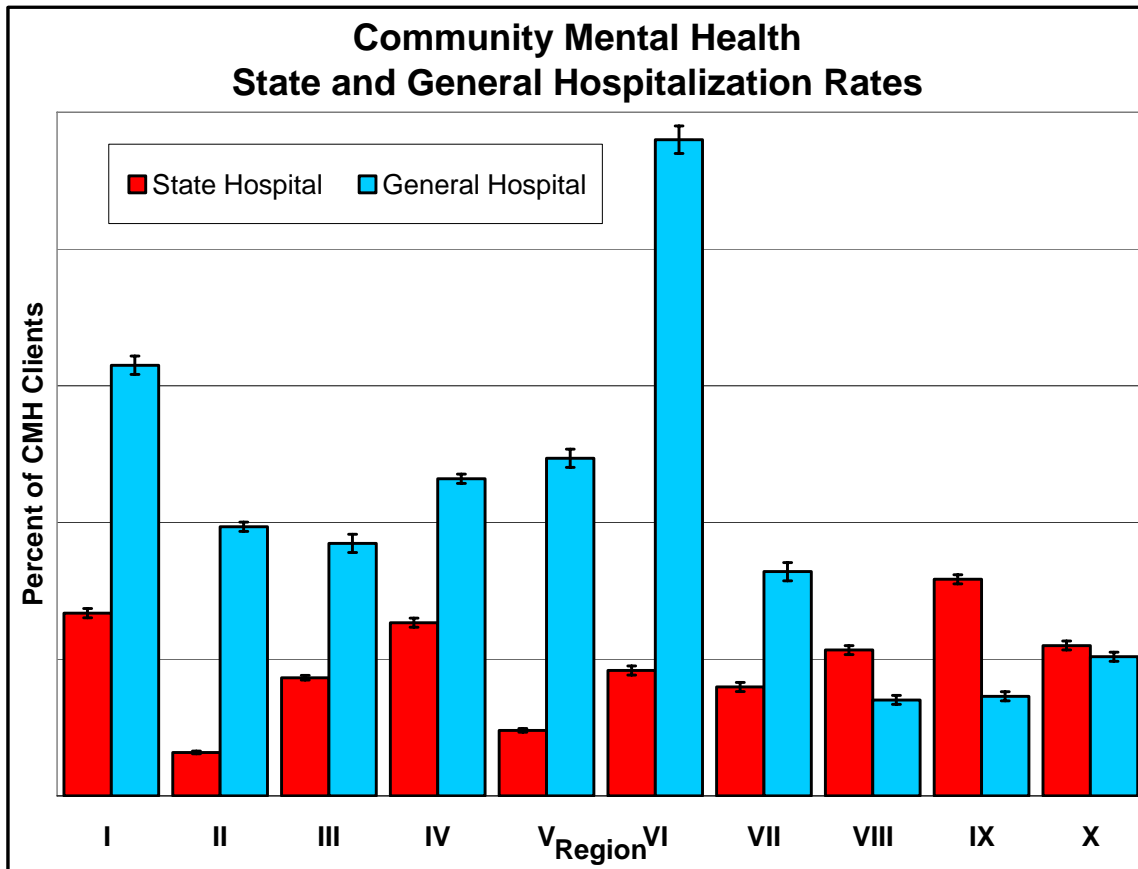


New Hampshire Population Overlap Estimation Project

Using Existing Databases to Meet State Information Needs and Federal Reporting Requirements

Six-Month Status Report



New Hampshire Bureau of Behavioral Health
Heidi Johnson

hjohnson@dhhs.state.nh.us
603-271-5031

The Bristol Observatory
John A. Pandiani, Ph.D. and Karen L. Danyow
bristob@together.net 802-453-7070

NASMHPD Research Institute
18th Annual Conference on State Mental Health Agency Services Research,
February 11-13 2008, The Westin Arlington Gateway Hotel, Arlington, Virginia

BACKGROUND

The New Hampshire Bureau of Behavioral Health recently implemented a new data reporting system designed to have minimal impact on community programs and to provide maximum protection of consumer privacy. This system has limited ability to monitor system functioning or to produce URS and NOMS data. To address this deficit, BBH issued an RFP for a contractor to provide valid and reliable measures of service system performance in ten key areas using only existing administrative data.

APPROACH

The Bristol Observatory (TBO) was selected by a team of independent reviewers to provide these measures. TBO works with the Bureau of Behavioral Health to obtain anonymous HIPAA-compliant extracts from other state agency databases, determines the numbers of people shared by BBH and other databases, provides formal reports of findings, and consults with program managers regarding the interpretation of findings. Estimates of caseload size and overlap are based on the statistical technology of Probabilistic Population Estimation.

Access to Other State Agency Data

At project mid-point, data for six of the ten key areas have been obtained and analyzed. Access to other state agency data is currently being negotiated.

One major facilitator of data access has been a series of formal briefings for program and data managers from other state agencies. Informal connections to staff of other state agencies have also been very valuable for opening doors and facilitating discussions of the project with relevant staff. Most important, the data required for this project primarily consist of HIPAA-compliant limited data sets that include only date of birth and gender of individuals served.

This project has also been very willing to discuss alternative approaches and to adapt to the special needs of different programs. These alternatives include direct linkage while preserving anonymity of service recipients, the potential for only requiring aggregate data from other participating state agencies, and the potential for on-site data analysis where no personal information leaves the other state agency site.

METHODOLOGY

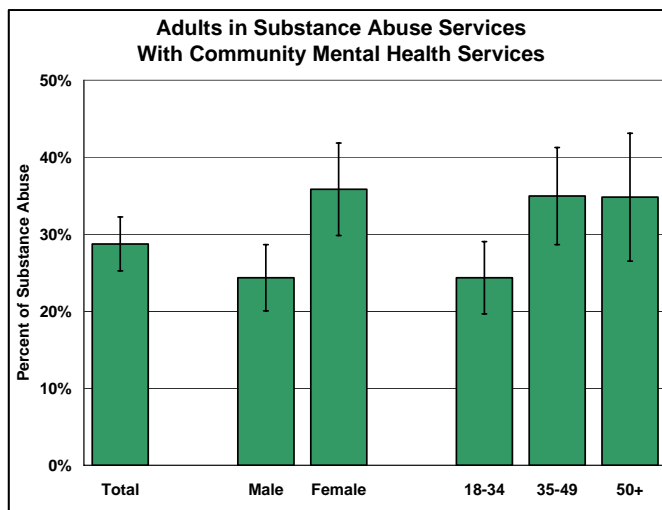
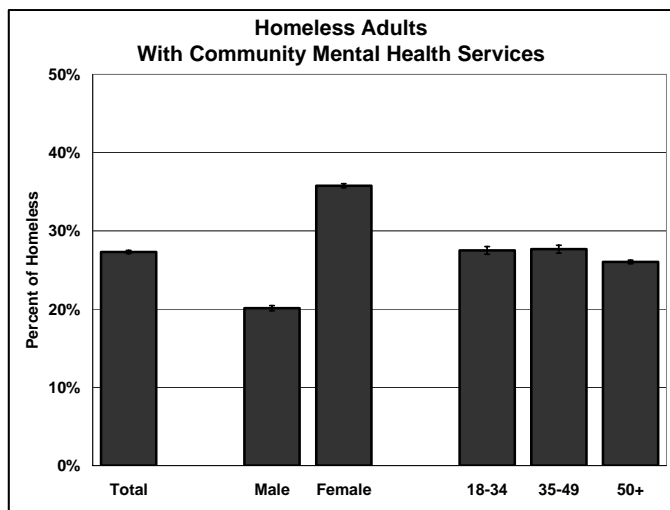
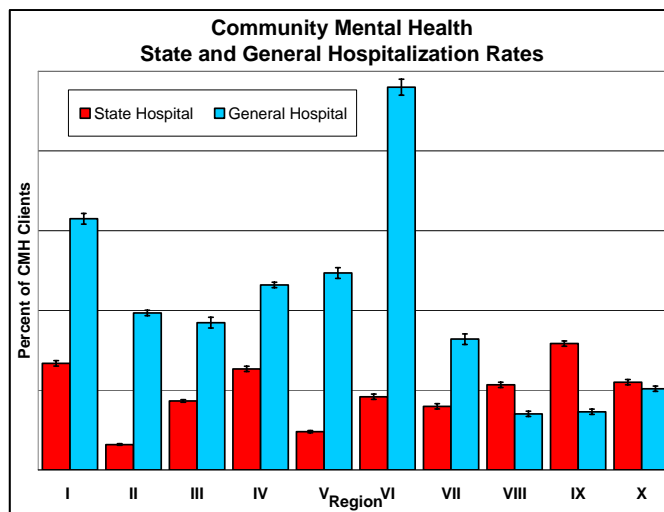
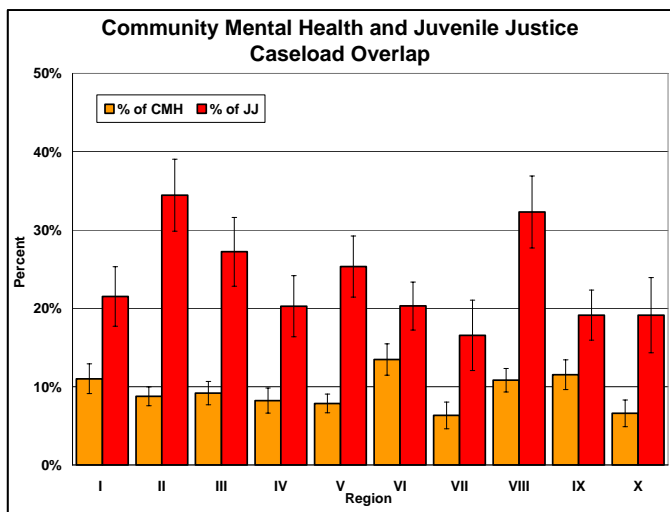
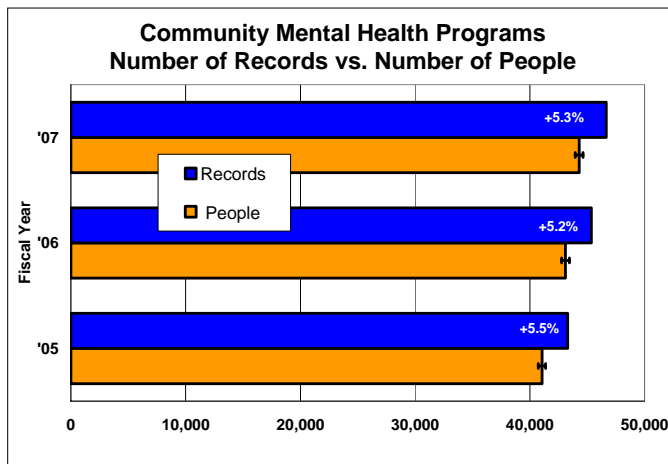
Probabilistic Population Estimation (PPE), the primary analytical tool used in this project, is a statistical procedure that provides valid and reliable estimates of the number of individuals shared by anonymous data sets. A brief technical overview is provided as an appendix to the document.

THE BRISTOL OBSERVATORY

The Bristol Observatory was incorporated in 1997 to support rational, data-based decision making at the state and national levels by providing valid and reliable measures of the functioning of systems of care using anonymous extracts from existing administrative databases using PPE. IN this way, TBO seeks to address the growing tension between two core American values: the value of personal privacy and the value of rational management and public accountability of human services programs.

FINDINGS

The first round of analysis focused on community mental health caseload size and overlap with other state agency caseloads within a single time period. Subsequent rounds of analysis will focus on change in caseload size and overlap over time and on treatment outcomes, including hospitalization and criminal and juvenile justice involvement after community mental health and inpatient treatment.



RELATED READINGS

Un-duplicated Client Counts

Probabilistic Population Estimation of the Size and Overlap of Data Sets Based on Date of Birth. *Statistics in Medicine*, May 2001. (Banks and Pandiani)

Personal Privacy vs. Public Accountability: A Technological Solution to an Ethical dilemma. *Journal of Behavioral Health Services and Research*, November 1998. (Pandiani, Banks, and Schacht)

Inpatient Psychiatric Services

The Utilization of State and General Hospitals for Inpatient Psychiatric Care. *American Journal of Public Health*, 88: 448-451, March 1998. (Banks and Pandiani)

Practice Patterns and Hospitalization Rates: A Statewide Program Evaluation. *Administration and Policy in Mental Health*, 26: 33-44, September 1998. (Banks, Pandiani, Gauvin, Reardon, Schacht, and Zovistoski)

Mortality

Mortality of Mental Health Service Recipients in Vermont and Oklahoma. *Psychiatric Services*, 53: 1025-1027 August 2002. (Pandiani, Banks, Bramley, Moore)

Age and Mortality Among Problem Drinkers. *Addiction*, 95: 1249-1254, August 2000. (Banks, Pandiani, Schacht, and Gauvin)

Criminal Justice

Crime Victims and Criminal Offenders Among Adults With Serious Mental Illness. *Psychiatric Services* 58: 11, 1483 - 1485 November 2007. (Pandiani, Banks, Carroll, Schlueter)

Elevated Risk of Being Charged with a Crime for People with a Severe and Persistent Mental Illness. *Justice Research and Policy* 2: 19-36 Fall 2000. (Pandiani, Banks, Clements, and Schacht)

Employment

Impact of Multi-Agency Employment Services on Employment Rates. *Community Mental Health Journal* 40: 333-345 August 2004. (Pandiani, Tracy, Simon, Banks)

Children's Services

Access to Child and Adolescent Mental Health Services. *Journal of Child and Family Studies* 14: 3, 431-441 September 2005. (Pandiani, Banks, Simon, Van Vleck, Pomeroy)

After Children's Services: A Longitudinal Study of Significant Life Events. *Journal of Emotional and Behavioral Disorders* 9: 131-138 Summer 2001. (Banks, Pandiani, and Schacht)

For a complete list of related publications visit:

www.thebristolobservatory.com

METHODOLOGICAL NOTE PROBABILISTIC POPULATION ESTIMATION

Probabilistic Population Estimation is a statistical procedure that determines the number of people (with known confidence intervals) who are represented in data sets that do not contain unique person identifiers. Probabilistic Population Estimation uses information on the distribution of birth dates in a data set to determine the number of people represented in the data set. The number of people necessary to produce the number of birthdays observed in a single birth year cohort, for instance, would be calculated using the following formula:

$$P_j(l_j) = \sum_{i=1}^{l_j} \frac{365}{365-i}$$

where “P_j” is the number of people and “l_j” is the number of birth dates observed. Similar logic is used to determine the number of people who appear in more than one data set. The table below provides illustrative results of Probabilistic Population Estimation for populations of specified size.

Population Estimates for Specified Numbers of Birth Dates within a Year

Birth Dates	Number of People	Birth Dates	Number of People
1	1.003 ± 0.103	180	249 ± 20
10	10.15 ± 0.776	250	423 ± 38
20	20.6 ± 1.54	300	632 ± 64
50	54 ± 4	330	860 ± 101
100	117 ± 9	360	1603 ± 325

POPULATION OVERLAP

In order to probabilistically determine the number of people shared across data sets that do not include a common person identifier, the sizes of three populations are determined and the results are compared. The number of people in each of the original data sets are the first two populations. The number of people in a data set that is formed by combining the two original data sets is the third data set.

The number of people who are shared by the two data sets is the difference between the sum of the numbers of people represented in the two original data sets and the number of people represented in the combined data set. This occurs because the sum of the number of people represented in the two original data sets includes a double count of every person who is represented in both data sets. The number of people represented in the combined data set does not include this duplication. The difference between these two numbers is the size of the duplication between the two original data sets, the size of the caseload overlap. In terms of mathematical set theory, the intersection of two sets is the difference between the sum of the sizes of the two sets (A+B) and the union of the two sets (A∪B):

$$(A \cap B) = (A + B) - (A \cup B).$$

A video of Dr. Steven Banks explaining PPE is available for viewing at
www.thebristolobservatory.com

The Bristol Observatory

The Bristol Observatory (TBO) is a contract research firm that specializes in human services program and service system research and evaluation with a special emphasis on treatment outcomes. We specialize in providing knowledge derived from the analysis of large databases while protecting the personal privacy of the individuals represented in those databases. One of our primary tools in this work is Probabilistic Population Estimation, a statistical data mining tool that uses information on the distribution of dates of birth in anonymous data sets to produce the information that is increasingly demanded of service system administrators.



New from TBO

PPE ^{Plus}

PPE ^{Plus} is a procedure for calculating measures of caseload overlap for narrowly prescribed time periods (e.g. 30 days, 180 days). The utility of administrative databases for service system research and evaluation is substantially enhanced by the use of PPE ^{Plus}. In conjunction with anonymous extracts from a variety of data sets, PPE ^{Plus} can provide information regarding significant events after state hospital discharge such as participation in community mental health services or readmission to other hospitals. PPE ^{Plus} can also determine rates of criminal justice involvement during specified time periods before and after substance abuse treatment as required by the SAMHSA NOMS reporting system.

Web-based PPE Processing

Web-based processing of PPE analyses allow states and other customers to determine cross-agency caseload size and overlap without disclosing any person-level information. This approach begins with local processing of person-level or event-level data using computer programs provided by TBO. Aggregate (non-person level) data resulting from this local processing is up-loaded to TBO's on-line PPE processor for determination of caseload size and overlap. Results of this analysis will be returned to the user in a convenient Excel format. These results will include year-of-birth and gender specific estimates of caseload size and overlap as well as instructions for further analysis of these data at the local level.

Web-based PPE processing substantially reduces the cost of determining caseload overlap and the time between data submission and receipt of results. Web-based processing also eliminates the need for HIPAA "Data Use Agreements" or "Business Associate Agreements" because no personal information is shared with TBO.

Web-based PPE processing (U.S. Patent # 7,139,675 B2) is currently available for Beta testing on a limited basis.

Full service PPE analysis and reporting continues to be available from TBO. These services include expert consultation, research design, data analysis, and report preparation.

For more information:

e-mail bristob@together.net

visit www.thebristolobservatory.com

call 802-453-7070

RELATED READINGS

Un-duplicated Client Counts

Probabilistic Population Estimation of the Size and Overlap of Data Sets Based on Date of Birth. *Statistics in Medicine*, May 2001. (Banks and Pandiani)

Personal Privacy vs. Public Accountability: A Technological Solution to an Ethical dilemma. *Journal of Behavioral Health Services and Research*, November 1998. (Pandiani, Banks, and Schacht)

Inpatient Psychiatric Services

The Utilization of State and General Hospitals for Inpatient Psychiatric Care. *American Journal of Public Health*, 88: 448-451, March 1998. (Banks and Pandiani)

Practice Patterns and Hospitalization Rates: A Statewide Program Evaluation. *Administration and Policy in Mental Health*, 26: 33-44, September 1998. (Banks, Pandiani, Gauvin, Reardon, Schacht, and Zovistoski)

Mortality

Mortality of Mental Health Service Recipients in Vermont and Oklahoma. *Psychiatric Services*, 53: 1025-1027 August 2002. (Pandiani, Banks, Bramley, Moore)

Age and Mortality Among Problem Drinkers. *Addiction*, 95: 1249-1254, August 2000. (Banks, Pandiani, Schacht, and Gauvin)

Criminal Justice

Crime Victims and Criminal Offenders Among Adults With Serious Mental Illness. *Psychiatric Services* 58: 11, 1483 - 1485 November 2007. (Pandiani, Banks, Carroll, Schlueter)

Elevated Risk of Being Charged with a Crime for People with a Severe and Persistent Mental Illness. *Justice Research and Policy* 2: 19-36 Fall 2000. (Pandiani, Banks, Clements, and Schacht)

Employment

Impact of Multi-Agency Employment Services on Employment Rates. *Community Mental Health Journal* 40: 333-345 August 2004. (Pandiani, Tracy, Simon, Banks)

Children's Services

Access to Child and Adolescent Mental Health Services. *Journal of Child and Family Studies* 14: 3, 431-441 September 2005. (Pandiani, Banks, Simon, Van Vleck, Pomeroy)

After Children's Services: A Longitudinal Study of Significant Life Events. *Journal of Emotional and Behavioral Disorders* 9: 131-138 Summer 2001. (Banks, Pandiani, and Schacht)

For a complete list of related publications visit:

www.thebristolobservatory.com