

TREATMENT OUTCOMES AND PERFORMANCE PILOT

STUDIES ENHANCEMENT (TOPPS II)

Steering Committee Meeting #5: The Interstate Cooperative Study (ICS) States

March 13-14, 2000

Orlando, Florida

SUMMARY REPORT

INTRODUCTION

The fifth Steering Committee (SC) meeting of the TOPPS II program was held in Orlando, Florida, on March 13-14, 2000. This program is being conducted under cooperative agreements between the Substance Abuse and Mental Health Services Administration's (SAMHSA) Center for Substance Abuse Treatment (CSAT), the Technical Assistance Center (TAC), and 19 States.

This meeting focused on the six Interstate Cooperative Study (ICS) States that are tracking client outcomes using secondary data : Maryland, Oklahoma, and Washington, who will be collecting secondary data only; and New Jersey, Texas, and Virginia, who will be collecting both primary and secondary data. Representatives from the other 13 States in the TOPPS II program, who are collecting primary data only, participated as observers. The meeting had two primary goals: (1) to discuss issues concerning secondary data analysis; and (2) to provide technical assistance to States planning on doing secondary data analysis.

BACKGROUND: Interstate Cooperative Study (ICS) Protocol Review, Amelia Arria, Ph.D.

The ICS has two objectives: develop standardized outcomes measures and standardized measures of factors that influence those outcomes; and use those standardized measures in analyses that can be replicated across participating States. The ICS will enhance the TOPPS II initiative in several ways: it will help create ongoing, sustainable outcomes monitoring systems; it will explore the viability of cross-State comparisons of treatment outcomes; and it will provide other States with a model for integrating and analyzing secondary data.

This overview of the protocol is based on decisions made during the last meeting, held October 14-15, 1999, in Baltimore, Maryland. The study is exploring two basic research questions: (1) How is completion of a treatment episode related to wages earned and/or employment status? to the probability of an arrest? to the probability of returning to substance abuse treatment? and (2) What proportion of substance abuse clients die over a specific time period compared with the general population, and does the proportion vary by treatment completion status?

An "episode" is defined as continuous care for substance abuse treatment with no more than a 30-day gap between discharge from treatment and a new admission.

Four post-treatment outcomes are being studied: (1) total wages--the sum of four quarters of wages, beginning the quarter after discharge from an episode; (2) arrest--a dichotomous variable indicating

whether a client was arrested in the year following discharge from a treatment episode; (3) readmission to treatment--a dichotomous variable indicating whether a client returned to substance abuse treatment in a specific time period; and (4) death--a dichotomous variable indicating whether a client died in a specific time period.

The independent variables include treatment status, length of stay, reason for discharge, age, gender, race, residence, pre-treatment wages, pre-treatment arrests, primary drug, age at first use, and frequency of use.

Clients are included if they entered and completed a treatment episode in FY 1997 (because the study will go back to 1996 and forward to 1998), are age 18 or older, and have a unique personal identifier available. Non-primary and methadone clients are excluded.

The study is dealing with several complexities, including the following: cross-State difference in client case-mix will contribute to variation in results; cross-State differences in macro-environmental variables (such as arrest rates and employment opportunities) will be difficult to control; and definitions of successful completion of treatment vary across States, programs, and modalities.

MEETING DECISIONS

● Decisions Concerning Subcommittees

Two subcommittees will be formed which will have the planning responsibilities:

- •The publication subcommittee will make recommendations on dissemination issues such as guidelines on the protocol and the authorship, content, and clearance process for any future publications concerning the TOPPS II Inter-State Study results.
- •The data analysis subcommittee will provide input into the overall analytic plan, work with the TAC in developing the analysis plan, delineate key policy and technical research issues in doing cross analysis, and recommend decisions on accessing and sharing data, both while the project is ongoing and after the data become public domain.

During the meeting, the following members volunteered to serve on the two subcommittees, with additional volunteers to be solicited by e-mail after the meeting, and chairs to be elected subsequently within the committees:

Publication Subcommittee

Sheila Harmison, DSW, LCSW (Project Officer)

Antoinette Krupski, PhD (SC Chair)

Technical Assistance Center (TAC) staff

Christie Dye (AZ)

Amelia Arria, PhD* (MD)

Liz Evans, MA*, CA

Irene Cramer, PhD*, MA

Data Analysis Subcommittee

Sheila Harmison, DSW, LCSW (Project Officer)

Antoinette Krupski, PhD (SC Chair)

Technical Assistance Center (TAC) staff

Amelia Arria, PhD* (MD)

Drew Hanchett, MPH (MA)

Sharon Zahorodnyj (IL)

Alfred Bidorini, MA (CT)

Robin Nelson, PhD (TX)

Anna Kline, PhD (NJ)

Michelle Jenson, MS (UT) **Subcommittee members who are not their States' principal investigators (PIs) will need written designations from their PIs.*

- **Decisions on the ICS Protocol: Bill Luchansky, Ph.D., Discussion Leader**

A discussion on the individual sections of the protocol focused on decisions regarding the following issues:

- **•Episode**--The episode definition stands as written in the protocol (continuous care for substance abuse treatment, the 30-day window between discharge and the next admission).
- **•The index episode**--This episode is defined as the last complete episode of FY 1997, with no treatment occurring in the following 30 days in the next year. It will be necessary to look backward and forward in the data to detect any clients lost from the study because of a longer length of stay. So far it seems that only about 2 percent are not captured, which, for purposes of the cross-State analysis, is a tolerable error. For longer term studies, once more data are obtained for more years, it will be possible to follow those people through time.
- **•Study populations**--The study population definition stands as written in the protocol (entered and completed a treatment episode in FY 1997, are 18 or older, and have a unique personal identifier).
- **•Time periods**--The pre- and post-treatment time periods remain as defined in the protocol (365 days prior to admission to the index episode, and 365 days after discharge from the index episode). It was suggested that, in the future, new in-treatment variables be added, such as in-treatment arrests and in-treatment employment. A study by Anglin & Hser was recommended regarding pre-, during-, and post-treatment activity.
- **•Episode completion**--The discharge code for the last SDU will be used, unless the last SDU was detox. If detox was the last service received, but not the only level of care, the discharge code will be ignored and the client will be considered a noncompleter.
- **•Comparison group**--The choices for comparison group were noncompleters, detox only, and assessment only. Detox only was ruled out because it was not a viable option for some States. A randomized delayed implementation model was proposed, using a synthetic comparison group of clients who received treatment in FY98 but not in FY97 and using their arrest and employment findings from FY97. This type of comparison group has some potential problems: policy changes; severity of client prior to treatment; and the clients who are seeking treatment. It was suggested that the ICS use noncompleters as a comparison group in a pre/post design and the synthetic group

to compare outcomes.

- **•Readmissions as an outcome**--Reentry will be moved from the outcome level to a descriptive level. It was proposed that two people (not identified) look at the research questions and report back to the group at a later date. The discussion points included several suggestions:
 - -Do not use readmission as an indicator for drug use because of the confusion over readmission as a good or bad outcome.
 - -Look at time to readmission and characteristics of clients that are readmitted to treatment.
 - -Look at predictors of time to readmission separately for detox, outpatient, and residential.
- **•Arrest**--Arrests remain an outcome variable. But the issue of whether to look at felonies only or gross misdemeanors and felonies was tabled until each State can look at their data and determine what type of arrests are recorded.
- **•Mortality**--Mortality will be moved from the outcome level to a descriptive level. The research question will be rephrased to include descriptive analyses and issues relating to death. This decision was based on the small numbers and the slow death rate among substance users. Other suggestions included using mortality to remove clients from the denominator and using mortality information to compare causes across States.
- **•Employment**--Employment remains an outcome variable in a pre/post design using noncompleters as a comparison group and the synthetic group to compare outcomes.
- **Decisions Concerning the TOPPS II Core Data Set**

Maryland, Oklahoma, and Washington grantees agreed to submit the three TEDS items and demographics to the TAC at T_1 and T_2 . The first admission between July 1 and December 31, 1999, will be submitted for T_1 . The discharge following T_1 will be submitted for T_2 .

T_3 will be the employment indicator (employed, not employed) from the administrative employment data received for the third quarter following the admission (T_1). Maryland and Washington representatives expressed concern regarding redisclosure of information and will need to go back to their respective States to determine if the T_3 indicator will be possible. Oklahoma representatives did not feel that this was redisclosure and will submit the T_3 data.

REVIEW OF INDIVIDUAL SECONDARY STATES' PROGRESS

Representatives from Oklahoma, Washington, and Maryland reviewed their States' studies and their progress in obtaining and linking data.

- **Oklahoma: Tracy Leeper, M.A.**

As a TOPPS I State, Oklahoma already had been receiving data from several sources when starting TOPPS II: Department of Public Safety AOD-related driving conviction data (fiscal years 1992-98); Department of Corrections incarceration, probation, and parole data (from the beginning of data collection through September 1999); Oklahoma Employment Security wage data (January 1996 through December 1999); Oklahoma Tax Commission tax data (calendar years 1992-98); and recently added Department of Health mortality data for 1992-97. For several years, the project has been working with

the State Bureau of Investigation's arrest database, but many technical difficulties remain. In working with the Oklahoma Health Care Authority for Medicaid and the Department of Human Services to obtain TANF, child welfare, and food stamp data, the project is trying to solve transmission issues.

The project is using data from 1996, 1997, and 1998. In trying to link the 1996 data, staff encountered denominator problems. They are therefore restructuring the dataset so each client has one record with all the outcomes together. They have tried several techniques. Currently, they are switching to transact sequel software which should process the data more efficiently.

- **Washington State: Bill Luchansky, Ph.D.**

So far, Washington has obtained several datasets: treatment data (1994-99); employment (wage and hours) data (1993-98); arrest data (1988-98); fee-for-service and managed care medical data; Department of Health hospital admissions data; and Department of Health mortality data. Department of Corrections data are still pending.

A recent report on readmissions to treatment, enclosed in the binder distributed at the meeting (under Tab 5), defined an episode as continuous care for treatment, with 30 days or less between admissions. Over a 3-year followup period, 24 percent of clients returned to either inpatient or outpatient treatment, thus refuting the notion of treatment as a revolving door. Only 17 percent returned in the second year and 12 percent by the third. A multivariate analyses after 1-year followup showed several findings: clients who had completed treatment had a lower risk of readmission than noncompleters; people receiving inpatient treatment only were more likely to be readmitted than those who received either outpatient or inpatient plus outpatient, thus arguing for a continuum of care; women were more likely than men to be readmitted; and primary alcohol users were more likely to return than primary heroin, cocaine, or amphetamine users.

Discussion Point:

Linking multiple admissions into an episode if 30 or fewer days elapse between discharge and readmission is reasonable, and it also reflects many States' reporting requirements. But further study is needed to examine the patterns of admissions and discharges within treatment services in order to better define an episode of care. Complicating issues could include service availability gaps and the possibility of a client going on a "bender" within the 30 days (which would, in fact, be a new episode of care, not a continuation of the old episode).

- **Maryland: Amelia Arria, Ph.D.**

Maryland's study aims to determine the feasibility of linking client information with information from health (Medicaid, mortality, and hospital discharge), employment, and criminal justice databases (including AOD-related driving arrests) to assess post-discharge outcomes. A substudy aims to pilot measures of client satisfaction. Another substudy will exclude methadone clients from the secondary data analysis. And an additional possible substudy will involve using mock discharge forms for methadone maintenance patients.

The study's primary question of interest is: Does successful completion of treatment objectives predict post-discharge outcomes? The design compares treatment completers with noncompleters, examining 6 months pre-admission as a predictor variable, the treatment process variables, and 6 months post-admission in the administrative databases. It will examine three additional predictor variables: decreased severity of drug problems, reduction in drug use, and number of months in treatment. Of

further interest will be examining variability or subgroup variation resulting from different modalities of treatment, client demographics, living arrangements, health coverage, mental health status, and type of drug problem.

The processes of securing cooperation from the various agencies and getting IRB approval from the university and the State have been fairly successful. The study will do probabilistic matching for the hospital discharge data because they do not include SSNs (for confidentiality purposes). For all other data, the various agencies will do the matching.

The projected sample size will be 4,000 unique admissions (Baltimore City) for Phase I and 30,000 unique admissions for Phases II and III. Nearly 10 years of treatment data are available (fiscal year 1989 through April 1999).

A correlational analysis between the last episode of treatment and all prior ones showed that more than 90 percent of people had only one episode. And of those, 92 percent had one service delivery unit (SDU). So the majority of people had one SDU and one episode--thus reducing the concern about multiple episodes and multiple SDUs.

DATA ACCESS AND CONFIDENTIALITY: ISSUES AND SOLUTIONS

Linda Graver, Dario Longhi, Ph.D., William Luchansky, Ph.D., and Steven P. Davis, Ph.D., talked about the challenges they have faced in dealing with secondary data access and confidentiality issues, and they discussed the methods they have used to resolve those issues. These individuals, along with Tracy Leeper, M.A., also co-facilitated a breakout session to further discuss the subject.

● Linda Graver, MEDSTAT Group, Santa Barbara, California

In dealing with issues related to secondary data access and confidentiality, the ICS States can draw from the experiences of the Integrated Database Project. This collaborative effort between CSAT, CMHS, and three States aims to develop a database to estimate the direct costs of services to those with alcohol and other drug abuse (AOD) and mental health (MH) disorders. It deals with confidential data--namely, any data that can be associated with or identifies an individual--because identifiers are necessary to link people across AOD, MH, and Medicaid files.

Two kinds of identifiers must remain undisclosed in order to preserve client privacy: (1) personal identifiers, which identify individuals directly (such as name, address, telephone number, and Social Security numbers [SSNs]); and (2) personally identifiable data, which identify by reference or combination of information (such as birth date with zip code or county; or county information with diagnosis).

Each State has unique confidentiality requirements. Some have formal Internal Review Boards (IRBs), which usually require extensive project descriptions; others are less formal but nevertheless guard client confidentiality. Furthermore, while the Health and Human Services 42 CFR allows using confidential AOD data for research purposes without patient consent, each State may interpret it differently. For example, some States consider the program director (who stipulates whether or not the data may be disclosed) as the State-level division head; others interpret that title at the individual clinic level--which would make it impractical to get consent.

Setting internal confidentiality standards includes several steps: sign a confidentiality agreement through a nondisclosure statement; assure the States that personal identifiers and identifying information is kept

on file for internal use only and that they will be dropped from files when no longer needed; assign new identification numbers for each client, or else encrypt the information. Other data security measures are recommended: keep a dedicated computer in a locked room; specify a data custodian; authorize specific data users; password-protect the data; secure the facility and lock the storage area; shred all output; track the data; and keep the information on a simple form.

Each State and each agency must deal with different restrictions. And interagency agreements are necessary before using data by multiple agencies.

In acquiring data, these suggestions might help: know what you need before requesting data; know which files to request (service files, client files, hospital files, provider files, or cost files); remember file types differ in each State; request files months in advance; request specific files and specific data elements. Outcome studies might require varying types information: beginning and ending treatment dates; units of service; type of services provided or program information; employment information; hospital readmission; or housing information.

Additionally, be prepared to deal with several variations: file size varies by State and agency; staff may vary in their ability to work with large files and multiple formats; file format may vary by State and agency (for example, EBCDIC versus ASCII, fixed block versus varying block, compressed versus noncompressed, and SAS versus text versus comma delineated versus relational database); and media may vary (cartridges, CDs, or diskettes).

When starting to receive secondary data, include the following in a preliminary review: (1) assess the quality (Are the data fields filled with meaningful values? Is the full range of dates represented? Are all services represented?); (2) check for missing information (such as dates, diagnosis, address, name, and detox information); and (3) check for varying levels of reporting (client, episode, visit, or service levels).

In summary, in dealing with confidentiality issues and in acquiring data, be prepared for three things: each State will have different regulations and requirements; each State will have different file types and sizes; and allow enough time to obtain requested files.

- **Dario Longhi, Ph.D., Research Manager, and William Luchansky, Ph.D., Research Supervisor; Washington State Department of Social and Health Services**

Sample data sharing agreements with four data providers (included under Tab 6 in the binders distributed at the meeting) can be used as models: (1) the Oklahoma State Bureau of Investigation; (2) the Washington State Employment Security Department; (3) the Washington State Office of Administrator for the Courts; and (4) the Washington State Center for Health Statistics.

The human element is important in obtaining such agreements. Sometimes it is necessary to use circuitous ways. For example, to obtain the Employment Security agreement, it was necessary to become an agent to do research for the legislature, thus complying with a State regulation against sharing data across agencies. Then, to obtain the employment data, in order to comply with each agency's additional confidentiality regulations and security arrangements, programmers from each agency dropped a file into the mainframe and activated a matching program. The client file contained identifiers for both substance-abusing and non-substance-abusing clients so that the identities of substance abusers could not be divulged. Once the data were accessed and matched, they were tested for reasonableness.

The health and criminal justice data were easier to obtain because they contained mostly public

information. The entire databases were obtained, so it was possible to match the data in-house, rather than rely on the other agencies to do the matching--a much more reliable, less complicated situation.

Some of the secondary data obtained go back as far as 1992, so it is possible to track millions of people across time. In the future, it will be possible to have huge datasets of people's treatment and services history.

Once all the barriers have been broken and relationships are established, then the data usually continue coming in automatically, as long as confidentiality is preserved.

- **Steven P. Davis, Ph.D.; Director, Evaluation and Data Analysis, Oklahoma State Department of Mental Health and Substance Abuse Services**

Subpart D of the 42CFR allows for disclosures without patient consent in three cases: medical emergencies, research activities, and audit and evaluation activities. The third function is relevant in this case because the project is evaluating program performance.

It sometimes takes years to develop agreements with other agencies. Many approaches can be taken: legislation, inter-agency projects, piggy-backing onto other ongoing projects, quid pro quo (trading for something), and the "good ol' boy" approach (building on personal relationships with people in other agencies). Sometimes overlapping approaches are necessary, as has been the case in Oklahoma.

A sample interagency agreement in Oklahoma includes the following components: a description of the project's purpose; the effective dates; a confidentiality section referring to the State and Federal laws applicable to both agencies; a description of data transfer and handling; modification or amendment information; appropriate signatures; and an acknowledgment statement for signature by the people who handle the data.

Discussion Points:

Question: What records are used to populate the file with "dummy" client records?

Answer: The Washington study had access to SSNs and similar data. An Iowa study using a placebo group used public records (drivers' license and arrest records) as a source of personal identifiers.

Question: Were any databases unobtainable?

Answer: Some agencies give only one dataset, and even that could take years of negotiations. And sometimes datasets don't get closed out till a year later. But for the most part, people trust researchers, they believe researchers will maintain confidentiality, and they believe performance evaluations are important.

SECONDARY DATA LINKING: ISSUES AND SOLUTIONS

Mary Jo Larson, Ph.D., and Daniel Whalen spoke about their experiences in dealing with data linking issues and challenges. These individuals also co-facilitated a breakout session to further discuss the subject.

- **Planning for Data Linking to Enhance Your Study: Mary Jo Larson, Ph.D., Senior Research Scientist, New England Research Institutes, Watertown, Massachusetts**

Plans for data linkage to enhance an outcome study will vary, depending on the study's purpose. Is it a

retrospective study, linking a State agency with a research entity or a with another State agency? Or is it a jointly planned, prospective research study? Does it involve linking a whole population, or just a specific sample, with the secondary data?

Before planning for data linkage, be sure the study question really requires it. Many questions can be answered without linking to other data. Sometimes it is easier to simply improve the survey design or methods. Linkage enhances a study only in specific cases: when it answers a question that cannot be answered otherwise; when interviews give inaccurate or incomplete data (for example, the cost of hospitalization); when interview sample attrition is a concern; when comparing across multiple payment sources (such as Medicaid and a Single State Agency).

If indeed going forth with the data linkage, set up a data sharing agreement, which typically assures the agency that: the purpose for using the data is directly related to the agency sharing the data; it is within the provisions of law; the data will be used only for their intended purpose, and ownership cannot be transferred; and the data will be returned when the research is completed. Several additional assurances might be required: the data's physical security will be protected (possibly a written protocol naming the data handlers and even an agency allowed to inspect the premises); personal data will not be released or published; the researcher will not contact individuals from the data; data will not be transferred over the Internet or through wireless transmission. Furthermore, the agency might even require assurances that: the agency reviews the work plan; the agency has the right to first review the findings; proper IRB approval is implemented; the study will use agency data definitions and guidelines in its analyses; patients cannot be prosecuted; you have a Federal and State certificate of confidentiality; violations can result in no further data or even jail.

It is possible to anticipate, and thus avoid, data sharing challenges in several ways: get patients to consent in advance to have specific records accessed later on; when getting patient consent forms and establishing data sharing agreements, establish timeframes that include possible project extensions; remember that you are at the end of the queue to obtain data; and realize that the agency sharing the data will expend time and resources to do so, so codify their knowledge in your agency, put into place a new set of conversations and data code books.

When building and filing huge databases, privacy is at risk no matter how secure the files. Two current issues are examples: in New York State, welfare officials plan to search Medicaid billings of welfare applicants to find out if they have been in treatment in the past; and in South Carolina, the U.S. Supreme Court is reviewing whether mandatory drug testing of pregnant women is constitutional if you know the data will be shared with the police or the child welfare agency. These patient privacy issues can be difficult from a research point of view.

- Data Linking Challenges: Daniel Whalen, Senior Program Analyst, The MEDSTAT Group, Santa Barbara, California

SAMHSA's Integrated Database for Mental Health and Substance Abuse Treatment Service project involves integrating Medicaid with AOD/MH data in three States: Delaware, Oklahoma, and Washington. Linking such datasets is based on identifying variables. Several techniques are available: probabilistic linking, match merging, or deterministic linking. This particular study employed probabilistic techniques, using multiple comparisons, scoring the comparisons, and obtaining a bimodal distribution of scores.

The linkage yields three groups of data: matched pairs, uncertain pairs, and nonmatches. To make the

uncertain group as small as possible, those pairs can be reviewed with either manual or automated techniques. The process involves several steps: (1) determine a variable's level of importance (scaling) (not an issue for evenly distributed identifiers, such as SSN and gender, but rather for identifiers not evenly distributed, such as last name and race); (2) block the data (limit the number of comparisons) to make the search more efficient; (3) compare the identifiers; (4) do a deterministic linking to classify a pair as linked or unlinked; (5) calculate weights by quantifying agreement between variables; and (6) do a probabilistic linking.

Results in the two States studied so far, relative to overlap from a probabilistic population estimate (Pandiani and Banks), yielded 80-86 percent linkage by probabilistic linking, 51-72 percent by match merge, and 59-76 percent by deterministic linking.

ANALYTIC CHALLENGES WITH SECONDARY DATA

Michael Finigan, Ph.D., and Bob Hubbard, Ph.D., spoke about data analysis issues. These individuals also co-facilitated a breakout session to further discuss the subject.

- **Michael Finigan, Ph.D., President, NPC Research, Portland, Oregon**

Analyzing secondary data poses several challenges:

- **•Missing data--**Don't assume you have all the data. Pilot the study to find where the missing elements are. At least 20 percent of the data fields will usually be empty and will complicate any analysis. Follow up on why the data are missing. An explanation might lead you to use proxy data instead of dropping the case.
- **•Missing or wrong dates--**Analysis depends extensively on dates because many problems can be manipulated if you know when things occurred. Missing or wrong dates are particularly problematic when dealing with large statewide systems because, unlike smaller projects, it is difficult to go back and find out information.
- **•Duration of the collection period--**The time periods while in treatment, just after treatment, and in the long term differ greatly, and they affect outcomes. For example, if you look only at a short-term period, it might seem like treatment caused costs to increase; but if you look at the longer term period, costs in fact stabilize.
- **•Anomalies--**Investigate results that appear anomalous. For example, a study showed an unexpected increase in food stamp use by individuals in residential treatment. Further investigation showed the increase to be artificial because it was the treatment provider who obtained the food stamps.
- **•Environment and history--**Investigate what else was going on during the time periods in question. Nothing happens in a vacuum. Interventions affecting clients, changes affecting programs and systems--such as Managed Care coming on board--all such factors must be disentangled during the analysis.
- **•Definitions--**In analyzing secondary data, examine how each data system defines variables, particularly arrests and treatment completion. For example, a study showed that a juvenile database defined an arrest as including multiple charges, while their recidivism database used multiple incidents (only one charge per arrest), so comparisons were invalid. Furthermore, client self-reports unexpectedly yielded higher arrest rates than administrative databases because individuals were confused about arrests versus charges. Similarly, different providers define

treatment completion differently. Furthermore, deal with the issue of multiple episodes, both for completers and for noncompleters.

Ultimately, it is important to find a balance between looking at the overall data superficially and getting bogged down in investigating what is really going on. In doing so, be mindful of costs. And start with the policy programmatic questions you want to answer.

- **Bob Hubbard, Ph.D., Director, Institute for Community-Based Research and National Development and Research Institutes, Raleigh, North Carolina**

Access to so much secondary data poses dangers. We deal with many questions and complex diverse issues. We need to address and frame those questions in the most appropriate way possible and then figure out how secondary data can help answer them.

Public health and social sciences have three approaches to answering questions: clinical trials, detailed observational studies, and secondary data. Each has limitations. Clinical trials are expensive and can create artificial situations. Observational studies can be difficult and time consuming. Secondary data give only a snapshot of what is going on. But secondary analysis can help verify the underlying principles about treatment and its effects. And it allows us to examine the effects of policy and programmatic changes over time.

In working with secondary data, consider several major issues:

- •Which of the following overarching questions do we want answered: Does treatment work? Do we want to monitor individual program effectiveness? Do we want to get data in order to help improve programs? Do we want to make comparisons across programs and States? Each of those approaches has major implications.
- •What are we measuring with the secondary data? How do the secondary data give an indication of variability in treatment? Do they capture variability at both the client and treatment levels?
- •What are the real dependent variables? In addition to the standard health, criminal justice, and drug use outcomes, do we need to access other dependent variables? Issues that are becoming increasingly important include access to treatment, barriers to treatment, selection to programs, how one becomes engaged in treatment, and the process of relapse and recovery.
- •What timeframes are being used?
- •What is the level of analysis-- Federal, State, county, within programs, within service agencies?

CROSS-STATE EVALUATION OF TREATMENT EFFECTIVENESS: Dario Longhi, Ph.D., Washington State; Bob Hubbard, Ph.D., Raleigh, North Carolina

In experimental and quasi-experimental designs, we use replication--doing the same study across different groups, across different individuals--to explain differences in treatment effectiveness (treatment-versus-comparison groups) across States. We also use rival explanations in replication across multiple cases (States or communities within States). Rival explanations in replication involves varying one thing--the measurement of "Y" (the treatment outcomes), the type of treatment "X," the types of populations ("n"s), or the control groups and time periods--to see whether treatment has an effect above and beyond all other possible reasons for people getting better.

States differ in treatment support services, such as vocational training opportunities and mental health services. Populations differ in unmeasured characteristics that affect outcomes, such as severity of

chemical dependency or physical disabilities. Differences are also found in contextual conditions, such as the labor market (employment), police department budgets (criminal justice involvement), availability of health services (use of Medicaid or health services), and eligibility requirements.

Therefore, in studying multiple cases, the design must not simply replicate across different cases; rather, cases must be chosen so that they exclude rival explanations. At least three solutions are possible: (1) case study; (2) statistical modeling with context variables; and (3) modeling within similar subgroup or context.

In a State case study, contextual factors are used as a rival explanation. Different context variables across States could include job training, labor markets, or the minority or female composition of the labor market. A State case study involves three steps: (1) Determine whether or not context variables differ from State to State; (2) Determine which ones differ; and (3) Determine whether they constitute rival explanations for the findings. This method has one problem: it doesn't show the degree of impact of any rival explanation.

But statistical modeling can help estimate the impact of contextual factors. That solution, too, involves three steps: (1) Determine whether any contextual variables also differ across communities within a given State; (2) Determine whether you can quantify each contextual variable and at what level of geography (such as "communities"); and (3) When modeling outcomes, determine whether you can estimate the degree and significance of the impact. Subsequent alternative steps could include the following: (1) Argue that impact estimates support the rival explanation in the State case study; (2) If data are available for enough States, pool the data and model across States; and (3) Go to the third possible solution--modeling within similar subgroup or context.

For example, one could model within the subgroup of urban, young men using hard drugs, or within communities with the same labor markets or the same arrest rates. This type of modeling takes two steps: (1) Determine whether any subgroups or contexts are similar across States on key variables; and (2) After rerunning the original statistical model, determine whether the results of treatment impact are similar.

In addition to looking for rival explanations, the ICS States could use case mix adjustment, nested models, and mixed models to try to account for individual and individual program differences and to find out what relationships are overarching and enduring. However, no matter how sophisticated the statistical methodology, it is important for to frame the questions from a practitioner and policymaker standpoint to prevent researchers from going down the wrong road. Logic must underlie the scientific methods.

TOPPS II PROJECT UPDATE: Sheila Harmison, D.S.W., L.C.S.W, Public Health Analyst, Division of State and Community Assistance, CSAT

It is now 1 ½ years since the project began. Nine States have begun data collection, and another three States will be starting to collect data at the beginning of April.

The STAR project, the TOPPS monitoring system for required project reporting, is now being handled by Mickey Smith, M.S.W., who recently joined CSAT. He has developed directions for submitting quarterly and summary reports via STAR, which will soon be web enabled. Eventually, a similar system will be developed for submitting Block Grant applications.

INTER-STATE STUDY DATA TRANSFER PROTOCOL: Kazi Ahmed, Ph.D., Robin Pugh Yi, Ph.D., and Fran Tepper, Ph.D., Technical Assistance Center (Johnson, Bassin & Shaw, Inc.)

Primary data collection began about 2 weeks before this meeting, when nine States began collecting T₁ (admission) and T₂ (discharge) data. The other States will be ready within the following 2-3 weeks. Most States have quality control systems in place, both software and human, to check for problems and inconsistencies. As the data come in, several additional precautions are in order: maintain quality control until the data collection process is complete; track sample size; stay in touch with providers; know your software, what it can and cannot do; back up your data, both hard and electronic copies; do error checking as soon as the data come in; train replacement interviewers completely and quickly; plan for cleaning the data while checking for errors--it is better to get missing data and correct errors right away, rather than use imputed data later on.

Two relevant items are included in the binder distributed at the meeting (under Tab 7): (1) a detailed set of instructions and due dates for submitting quarterly TOPPS II primary data, and (2) two lists of variable names (database dictionaries)--one for admission and followup (T₁ and T₂) and one for discharge (T₃). (Note: In the data dictionary, under SD_CODE (secondary drug use code), code 6 is supposed to read "non-prescription methadone," rather than "non-prescription medicine.")

Discussion Points:

- •Timing issues--The data submission timeline was developed based on an informal survey of the States to determine where they were up to. It was also developed to allow the TAC enough time to work with the data in time for the final submission date of January 1, 2001.
- •SPSS--The data will be submitted in SPSS because most States have that capability. However, the TAC can provide technical assistance to any States that need to convert data from SAS to SPSS.

The next TOPPS II Steering Committee meeting will be held sometime in October 2000. The exact date and location are still to be determined. The meeting will focus on emerging data analysis and publications issues.

C:\WP\REP7web.WPD