Enmeshed or Engaged?
Program Evaluation Strategies in Cross-System Programs

The National Conference on Substance Abuse, Child Welfare and the Courts
September 15, 2011

Ruth Huebner, Ph.D.,
Child Welfare Researcher,
Eastern Kentucky University

Robert Walker, M.S.W., L.C.S.W.,
Behavioral Health Researcher,
University of Kentucky

Disclaimers and Caveats

- We are evaluation researchers:
  - Using naturalistic or observational designs

- We are engaged, but sometimes worry about being enmeshed scientific advisors using a partnership model with practitioners.

- Both of us are deeply embedded in state government funded projects.

- We want to open a conversation with you regarding the evaluator role and implementation of program evaluation.

- Our premise: The demands of the current services grants exceed traditional role definitions between program and evaluation professionals in terms of:
  - complexity of service delivery;
  - complexity of selection and use of evidence-based practices; and
  - how to represent what services actually do.
Objectives

- To explore emerging roles for program evaluators in projects with inter-agency collaborations.
- To demonstrate how practitioners and administrators can take a more active role in owning the evaluative processes and outcomes.
- To explore strategies of merging provider-collected with administrative data from multiple data sets.
- To understand analysis and model building with complex nested data.

The Role of the Evaluator in Program Evaluation

ENGAGED PARTNER, SCIENTIFIC ADVISOR, OR; STRICT INDEPENDENCE FOR PURE “SCIENCE”?

The Research Challenge: Can you Control the Environment?

- Different Funding Sources and Jurisdiction
- Different Restrictions and Statutes
- Different Data Systems
- Different Practitioners in Each Agency
- The Same Families in all Three Agencies
- Different Missions and Important Outcomes

Research Design Challenges

Randomized Control Trial
- Discrete research question or problem:
  - Often short term
  - Readily measured and simple, often singular
  - Focus on discrete disorders
  - Controlled conditions
- Discrete Intervention
  - Medication
  - Simple comparison of intervention v. TAU

Cross Agency Studies
- Complex Problems
  - Substance abuse
  - Multiple co-morbidities
  - Families (whatever they are)
  - Communities
  - Multiple agencies
- Complex Intervention
  - Multiple strategies
  - Many staff and turn over
  - Many individual exceptions
  - Complex outcomes
Problems with “Pure” Science

- While clinical trials are the gold standard, they need close scrutiny in application to real world clients.
- They focus on single, discrete disorders and exclude subjects with co-morbidities.
- They also rarely account for subtle cultural or inequality effects among the sample by comparison to ‘real’ clients.
- They generally rely on well trained providers whose skill level may exceed community practice.

Another worry about Evidence-Based Practices

- Many practitioners are unprepared to read or read critically the methodology and analytical sections of research papers.
- Governmental requirements to use evidence-based practices diminishes the science supporting individually applied interventions in favor of program adoption of a favored practice.
- Dogma? Or science? Can’t be both.

- Most everyone fails to note that the “change” reported in clinical trials in behavioral health is simply subjects no longer meet all the diagnostic criteria for a disorder.
- The dropping of one DSM criterion can mark the change from disordered to recovered.

- The clinical trial outcomes almost never consider actual functionality such as improved overall well-being, improved employment, relationship satisfaction, social interactions, etc.
- They tend to focus on very simple, concrete outcomes with an assumption that not meeting DSM criteria mean overall improvement.
- They cannot be tried with all key ethnicities or cultural groups due to funding limitations.
Challenges of Random Assignment

- Seen as the ‘gold standard’ of program evaluation but is increasingly difficult when dealing with people in naturalistic settings. Discussion by Guo & Fraser (2010)
- Randomization may not exclude hidden selection bias and is influenced by processes such as the Hawthorne effect and potential biases in assignment to conditions
- Average TX effects mask individual results that may stem from multiple and important factors.

Positive Outcomes Fade Over Time

  - Even studies of drug effects when replicated over time tend to show diminished results. This is surprising, given that drugs are supposedly highly consistent. Other results also diminish.
  - Positive results are not published.
  - Studies with animals and humans are highly susceptible to many sorts of perception and selection biases.

A Variety of Methods

- Requiring randomization as the sole proof of intervention effectiveness will exclude many potentially effective and worthwhile practices.
- Random assignment is not practical or even possible for some studies – particularly when court and legal proceedings are involved.

Programs Change Over Time

- Klein (July 18, 2011). Head Start Doesn’t Work. Time Magazine
  - Proven in the 1960's to improve school performance. Taken to scale
  - Head Start Study by DHHS found that the effects were minimal and vanished by 1st grade.
  - Regression toward the mean or mediocrity.
Randomized Control Trials of Parachutes

- Did you know that there have never been ANY RTCs of parachute use as a prevention of death when falling from heights?
- The parachute companies are making huge profits with only observational evidence.


The Current Reality of Program Evaluation

- The data for program evaluation must come from practitioners who are often unsophisticated about scientific data collection methods, and anxious or uninterested in evaluation results.
- Approaches like random assignment are generally viewed with resistance. Research clarity may fall victim to routine clinical habits.

Emerging Role?

**Traditional**
- External Expert
- Practitioners are Dependent
- Independent Judgment
- Report to Prove

**Scientific Advisor**
- Internal Coach
- Practitioners are Partners
- Collaboration
- Reports to Improve

Scientific Advisory Roles

- What is meant by ‘scientific advisor?’
- We think there is need for someone to be a bridge between practice and science - both the science of selecting/implementing practices and the science of examining outcomes.
- Examples:
  - Guiding providers about a range of practices for the given client population;
  - Suggesting instruments to measure key clinical or behavioral characteristics for both clinical and research needs;
  - Suggesting ways to improve fidelity or at least describable service characteristics.
Continuing Advisor Role

- Check for adherence or fidelity to program procedures and serve as a catalyst for growth and increased potency for intervention. Make adjustments to program; test the reality of program assertions.

- Improve program accountability for achieving results rather than completing process or producing outputs.

- Examine the impact of strategies on results to infuse this idea: “Does what we do make a difference in what we want to achieve?”

- Guide decisions on continuing programs and showing the value – costs/benefits.

E.g. Program Strategy: Quick Access to the First Treatment

- Adults that Flee
  (closed case opened >=6months; N=251)

Discussion

How have you struggled with your role?

What do you think your role should be?

Is there room for variation in this among different programs and evaluators?
Engaging Practitioners In Evaluation Research

Overcoming Terror Among Practitioners and Learning to Learn Together

Program Evaluation and Practice: Goals and Fears

- Articulate and then monitor clear outcome expectations ("but we will know if it is working and it is way too hard to show it!")
- Improve program delivery and outcomes ("but the program is based on solid documented principles and EBPs, so we know what we are doing, and this will be so boring and take so long!")

Program Evaluation and Practice: Goals and Fears

- Show effects of an innovative program ("but what if we find out it isn't working - what then?")
- Demonstrate benefits in terms of a contract, funding agency, public, customer, or government ("but the results are statistically significant at the .05 level!") (Research myth)

Things to Remember in this Partnership

- Evaluators may carve out various roles in publicly funded programs, but...
  - It is likely that the data belong to the program, not the evaluator;
  - Some contracts call for program roles in publications based on program data;
  - Negotiating terms for all communications to the public about what can and cannot be said about the results is critical.
**Engaging Practitioners**

Practitioners own the data \textit{and} the results unless otherwise agreed.

- If you want providers onboard when you land, make sure they are onboard when you take off.
- Ask them “How will you know if it is working?”
- Teach them about the uselessness of mere anecdotes.

**Engaging Practitioners**

Evaluation can be a force for change and improvement.

- Program first rather than evaluation first.

Evaluation is a process not a report.

- Explore options and ideas, not jump to conclusions.
- Work together to understand program processes and outcomes

**Engaging Practitioners**

Providers can tell the story too.

- Present findings in graphs, diagrams that they understand.

Overcoming terror and learning to learn together.

- Present snippets of data at every meeting and generate reports?

Ask “What does this mean for families”

- Translate results in the numbers of children or families.

**Discussion**

How have you engaged practitioners?

What challenges do you face?

Contractual or interpersonal barriers?
Nested and Integrated Data

The Challenges of Cross System Data Complexity

Making Data Work Together

An Idea whose time has come!

- Provider-Collected data
- Census Data by County
- Child Welfare Data
  - Child based
  - Family based
- Court Data

An Idea that is very hard to implement!

Provider-Collected Data

- Be prepared for the limitations that go with these data.
- Reliability and validity need to be described carefully.
- Provider data in electronic form may be part of MIS financials and may not capture salient clinical variables as needed.
- Extraction of data from records is extraordinarily complex and compromised by reliability problems.

Provider Data

- Don’t assume that terms mean what everyone says or thinks they mean.
- Your evaluator use of terms may vary greatly from provider use.
- Even ‘simple’ terms can be a source of confusion - what is ‘admission’, ‘case management’, - how is discharge defined, etc.
Census Data

- Be sure to look into other secondary sources of census data and analyses of these data.
- University of Wisconsin has county-level data for every county in the U.S. with many other variables added to the census facts.

Child Welfare Data

- Check to see how these data are configured and how these tables can or cannot relate to other state data sets.
- How does the child protective service data system preserve identities across varying family configurations?
  - i.e., one child who now has a new stepparent and new step siblings, but the other child of the original family does not and is separated.

Court Data

- Again, check carefully for formatting, how identifiers work across docketed cases.
- Do the data allow for following persons or just cases?
- Are these data already matched or matchable to police or corrections data?

Matching

- Programs serving families with maltreatment histories often must match:
  - Child welfare data (NCANDS)
  - Court data
  - Police and arrest data
  - Behavioral health provider data
  - Treatment Episode Data Set (TEDS)
  - Vital statistics data (birth, death)
  - Research-specific data
**Constraints**

- Various confidentiality and data security provisions often limit client identifier uses.
- HIPAA rules can interfere with the use of identifiers.
- State IT support for doing anything other than the quotidian is generally tepid.
- Many state data sets are in antique formats with out-of-date data dictionaries, etc.
- “Confidentiality” is also a way for agencies to avoid the work involved in doing something with the data tables.

**Tips for Matching**

- Whenever possible, collect data only once.
- Try to reduce duplication not only for program efficiency, but two measures of the ‘same thing’ often are not the ‘same thing.’
- Be sure to develop clear anchor dates for needs matching.
  - Most state and agency datasets contain multiple year data – watch for contamination from previous year data.

**Check carefully for agreement of variable definitions across datasets.**

- E.g., many datasets will include a field for ‘substance use’.
  - Was this a lifetime measure, past 30 day, past 12 month, past 90 days?
  - Is it specific to type of drug or alcohol?
  - Does the field incorporate or signify multiple drugs?
- Decisions about which substance use variable to use may depend on which dataset has the most robust measure.

**What is a Service?**

- How are services defined by the different systems?
- What events are recorded? Are events things that happen to the case or are they what an agency does? Both?
- How up-to-date are the data you are using?
Provider Opinion Data

- What are the variations in opinions that color the client data set?
  - E.g., around client compliance coding?
  - Progress?
  - Termination status?
- Are there provider biases that are affecting these measures?

What is Admission? Discharge?

- What constitutes case opening in each of the datasets?
- What constitutes case closing? How do you handle “continued generally?”
- What definitions must you use for your reporting to the RPG and for the local evaluation?
  - Are these the same?

What to do with Conflicting Data

- How are you going to handle conflicts between data sets?
  - E.g., the child welfare dataset shows the client employed.
  - The provider shows unemployed, disabled.
  - The court data shows client stole from wife’s employer.
- Are your rules for RPG the same as for the local evaluation?

Talk through your Queries

- The development of grounded queries may take several runs and trials.
- There may be many preliminary steps to prepare a dataset before matching begins.
- For example, the algorithm for obtaining birth events data in Kentucky takes 13 steps just to prepare the table and identify the potential sample.
Bin Bag or Managed Data Tables?

- Some state datasets simply admit anything submitted as long as field definitions are correct.
- What data cleaning is done by the state IT staff? What procedures are used for this?
- Could data cleaning effect what you are looking for?
- Consultation is critical to avoid mistaken understanding of tables and variable definitions.

Our Challenges in Using These Data

- Timing and sequence of events: When is recurrence of child abuse really recurrence?
- Cutoff points for any strategy: What is ‘quick’ access to treatment?
- Messy data or incomplete data: How to avoid bias in correcting, discarding?

Steps in Analysis for Paired Datasets

- Descriptive statistics
  - Code for group membership
    - Naturally occurring groups or propensity scores
    - Groups based on sequence of events
    - Groups based on low or high dose
  - Comparative statistics control for case risk, age of the child, or other confounds as needed.

Nested Data from Several Sources

- Costs/ROI
- Community
- Adults
- AOD/MH Outcomes
- Substances, treatment
- Co-occurring disorders
- Child Welfare Outcome
- Case or Family
- Family Mentor Activity
- Family Functioning
  - NCFAS
- Children
- Repeated Progress Measures
Map into logic models

Building models – logistic regression, survival analysis, SEM to examine multiple relationships.

Translate back into ‘What does this mean for families, for children, for communities’?

Present in graph or picture.

---

Keep Vigilant for ‘Hidden’ Confounds

In examining birth outcomes for a sample of pregnant women with substance use disorders, examining other demographics may be important.

Payer source is of growing concern as part of health outcomes.

---

Source of Pay for Birth of Baby***

In this case, the payer may play an important role in the analysis.

![Source of Pay for Birth of Baby Chart](chart.png)

- **KIDS NOW Plus (n = 1,616)**
- **Comparison group (n = 1,616)**
- **General population (n = 1,616)**

- Medicaid: 63.2%, 75.1%
- Private insurance: 35.6%, 23.5%
- Other: 14.0%, 5.0%
- Champus/Tricare: 14.0%, 0.9%
- Self-pay: 2.2%

*** p < .001, ** p < .01, * p < .05

---

Discussion

What challenges do you face when merging and analyzing data?

What are your unresolved issues and challenges?
Thank You and Contact Information

Ruth A. Huebner, PhD
Child Welfare Researcher
Kentucky Department for Community Based Services
rutha.huebner@ky.gov

Robert Walker, M.S.W., L.C.S.W., M.Div.
University of Kentucky
333 Waller Avenue, Suite 480
Lexington, KY 40504
(859) 257-6623
Robert.Walker@uky.edu
jrwalk0@email.uky.edu