

Lisboa, 9-10 December 2013

Fundação Calouste Gulbenkian



INTERNATIONAL CONFERENCE

HOUSING FIRST

Ending Homelessness

Research and Evaluation

Tim Aubry

Organized by



Research and Evaluation Workshop

Tim Aubry, Ph.D., C.Psych.
Housing First Conference
Lisbon, Portugal
December 9-10, 2013



Overview of Workshop



- Introductions
- Programs & Program Evaluation
- Review of Program Logic Models
- Fidelity Assessment & Implementatin (Formative) Evaluation
- Impact Evaluation
 - RCT
 - Quasi-experimental design
- Performance Measurement



Introductions



1. Where are you from?
2. What is your involvement with HF programs?
3. What experience have you had in HF research?



Purposes of Program Evaluation (Posavec & Carey, 2007)



1. Develop and improve programs
2. Identify program outcomes
3. Report to funders
4. Help qualify for accreditation
5. Respond to requests for information



uOttawa

Centre for Research on Educational
and Community Services



Centre de recherche sur les services
éducatifs et communautaires

Program: Definition



“A more or less coherent system of inputs, activities, outputs, and outcomes that have been organized to address specific needs and resolve particular problems of a given population”





"We are neither hunters nor gatherers, we are evaluators..."

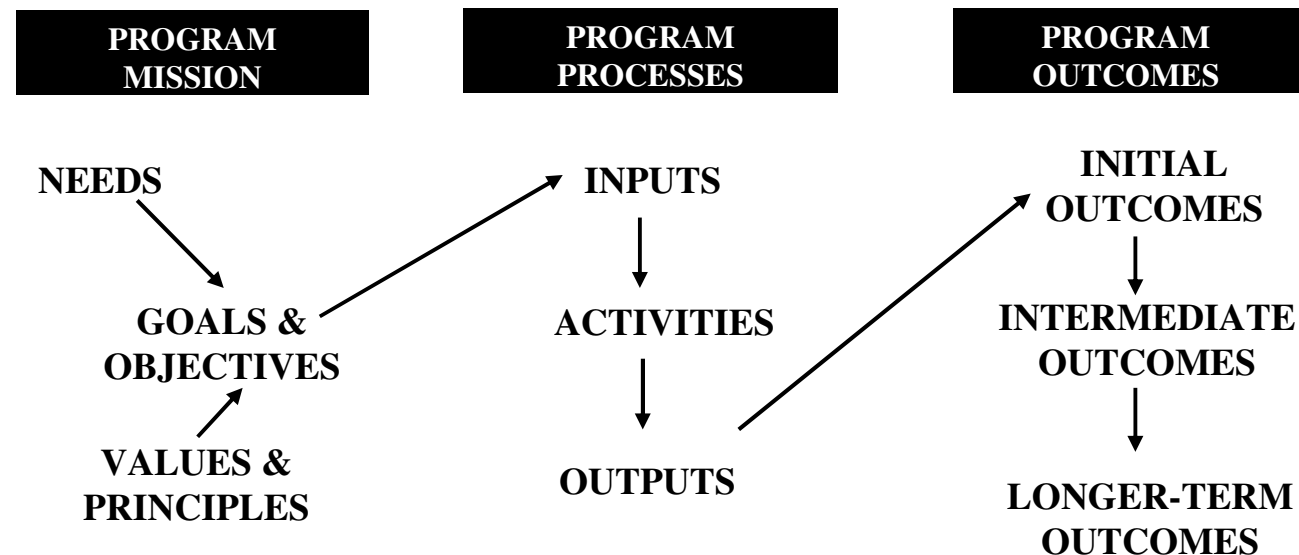
Program Evaluation: Definition



“The systematic collection of information about the activities, characteristics, and outcomes of programs to make judgements about the program, to improve program effectiveness, or to inform decisions about future programming”



Components of a Program



RELATIONSHIP TO THE THREE MAIN TYPES OF PROGRAM EVALUATION

NEEDS
ASSESSMENT

IMPLEMENTATIO
N / PROCESS
EVALUATION

IMPACT /
OUTCOME
EVALUATION

Program Theory:



“We view program theory as a relatively detailed description of the relationship between program resources, program activities, and program outcomes that shows how the program is supposed to work and, in particular how it is supposed to bring about the intended outcomes.” (Rossi, Freeman, & Lipsey, 1999)



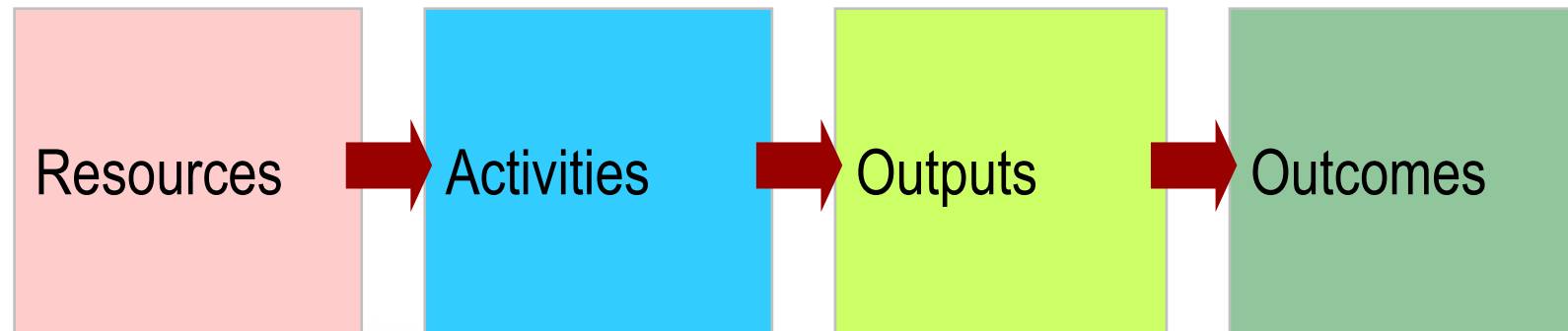
Program Logic Model: Definition



“A tool for describing program theory and for guiding program measurement, monitoring, and management”



Logic Model: If-Then Links



You need certain resources to be able to run your program.

IF you have access to those resources, **THEN** you can accomplish your activities.

IF you can accomplish these activities, **THEN** you will have delivered the services you planned.

IF you have delivered the services as planned, **THEN** there will be benefits for the clients, communities, systems, or organizations your program serves.



WORKSHEET FOR CONSTRUCTING A PROGRAM LOGIC MODEL

Agency: _____

Date: _____

Program: _____

Mission: _____

Clients Needs That Program Intends To Address	Intended Inputs	Intended Activities	Intended Outputs	Intended Shorter-Term Outcomes	Intended Longer-Term Outcomes

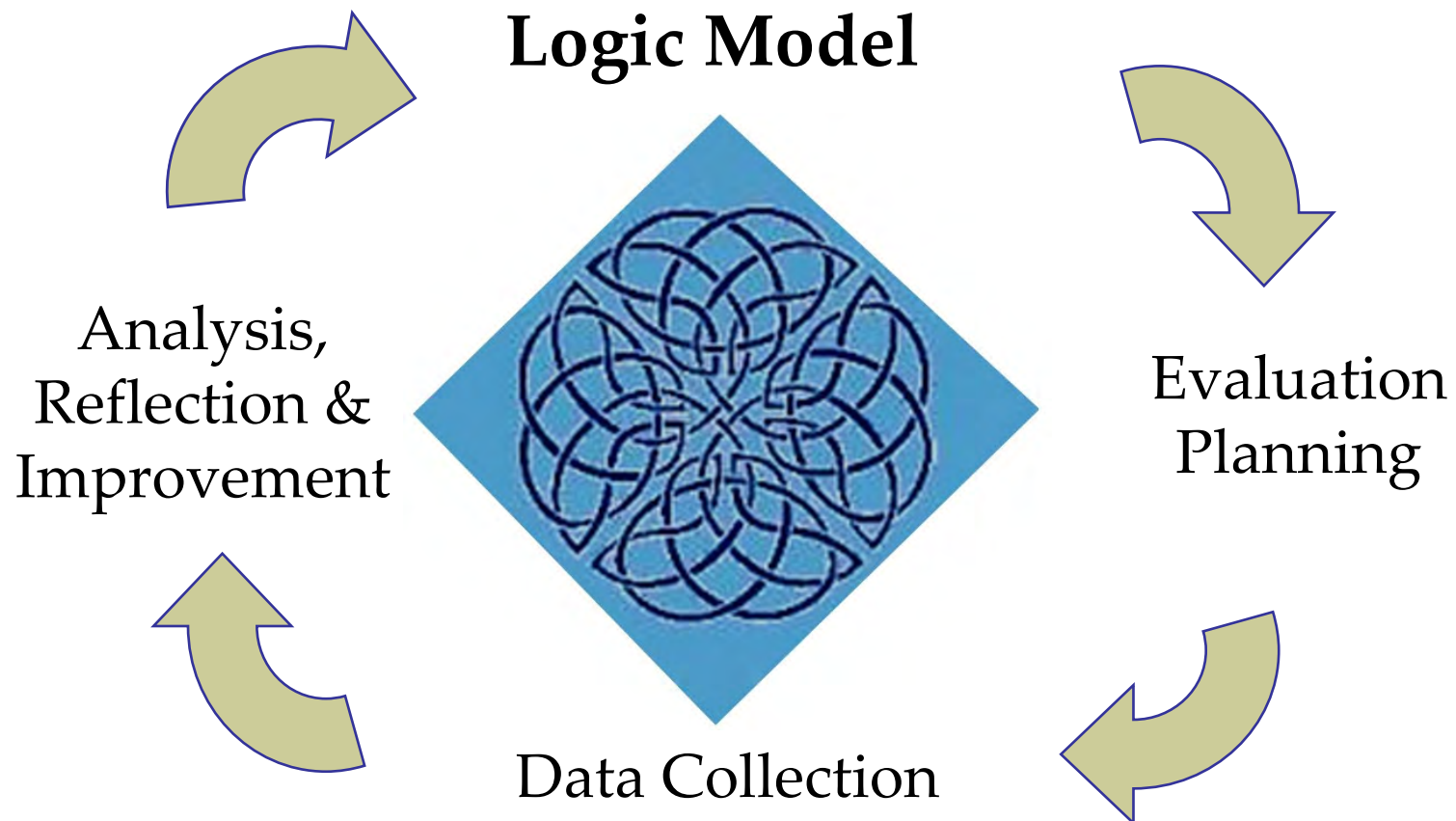
Program Logic Models

- Who has created PLMs for a HF program?
- What are the short-term outcomes of HF programs?
- What are longer-term outcomes of HF programs?

Logic Model of Housing First



Continuous Learning Cycle



Main Types of Program Evaluation



- Needs assessment
- Formative (implementation) evaluation
- Outcome (impact) evaluation
- Costing (efficiency) evaluation
- Program monitoring



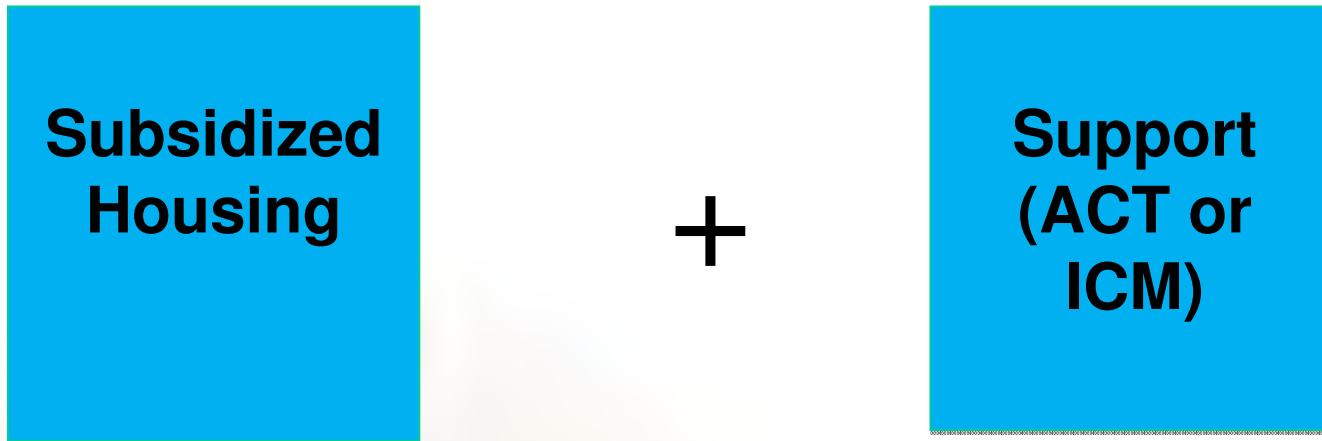
Formative Evaluation: Definition



“An assessment of whether the intended program inputs, activities, and outputs are actually being delivered to the intended recipients in the intended amount at the intended level of quality”



Housing First Approach of At Home / Chez Soi



Housing First Approach of At Home / Chez Soi



- Recovery oriented -- client choice at the centre
- Clients pay < 30% of their income for housing.
- Housing in self-contained units, mostly private sector and scattered site
- Treatment and support services are voluntary, individualized, culturally appropriate, portable
- No conditions on housing readiness. Tenancy not tied to engagement in treatment.
- Requirements: rent paid directly to the landlord, once a week visits by support team or case manager for pre-determined period.



Design of the At Home / Chez Soi Study



- Pragmatic, multi-site, randomized, mixed methods field trial in five sites across Canada (Vancouver, Winnipeg, Toronto, Montreal, & Moncton)
- Investigation of effectiveness and cost-effectiveness of **Housing First** in Canadian contexts
- Two fidelity assessments & two implementation evaluations
- Model being tested with support at two levels of intensity (high needs = ACT) moderate (ICM) vs. usual care
- Model being extended to include primary care, vocational supports, and special populations



At Home / Chez Soi: Formative Evaluation



- Evaluation questions:
 - Is the program reaching the targeted population?
 - Is the program being implemented as planned?
 - What are strengths of the program?
 - What are areas requiring further program development?
- Methods:
 - ✓ External fidelity assessment
 - ✓ Review of administrative data
 - ✓ Chart reviews
 - ✓ Key informant interviews
 - ✓ Focus groups with staff and clients



Questions?



- To what extent are the HF programs in which you are involved are reaching the right population?
- To what extent are the HF programs in which you are involved being implemented as planned?



Fidelity Standards for Housing First Approach

Fidelity Domains	Description
Housing Choice and Structure	Standards focusing on the provision of housing to consumers (e.g., housing choice, housing availability, integrated housing)
Separation of Housing and Services	Standards focusing on the relationship between housing and support provided by program (e.g., no housing readiness, standard tenant agreement, commitment to re-house)
Service Philosophy	Standards focusing on the principles and values guiding the delivery of services (e.g., service choice, harm reduction, assertive engagement, person-centered planning)
Service Array	Standards focusing on range of services available to consumers (e.g., psychiatric services, integrated substance abuse treatment, nursing services, supported employment services)
Program Structure	Standards focusing on service delivery characteristics (e.g., frequency of contact with participants, participant / staff ratio, team approach, peer specialist on staff)

Status of Current Implementation of Housing First at the Moncton Site

Scores on the Fidelity Domains for the Two Fidelity Assessments

Fidelity Domains (Total Possible Score)	August 2010	January 2012
Housing Choice and Structure – (24)	22.5	24
Separation of Housing and Services – (28)	27	28
Service Philosophy – (40)	35	36
Service Array – (32)	20	27
Program Structure – (32)	28	28

- Very good fidelity to Housing First model has been achieved and maintained.
- Program is functioning in line with Housing First recovery-oriented philosophy and practices
- Notable improvements in Service Array domain
- Still some room for improvement in three domains

Status of Implementation of Housing First in Moncton at 2 Years



1. High level of fidelity present for the Moncton program. Improved fidelity overall is apparent particularly for breadth of services in program.
2. Areas identified for further program development = (i) substance abuse treatment, (ii) person-centered planning, & (ii) addition of trained peer specialist to program.
3. Large number of program participants are experiencing stability.



Status of Implementation of Housing First in Moncton at 2 Years



4. Challenges remain for some participants who have experience multiple moves. Program has opened a 5-unit transitional apartment building.
5. Most landlords perceived the program positively; some landlords expressed dissatisfaction with program responsiveness.
6. Program staff are being transparent in communicating information about program sustainability.



Outcome (or Impact) Evaluation: Definition

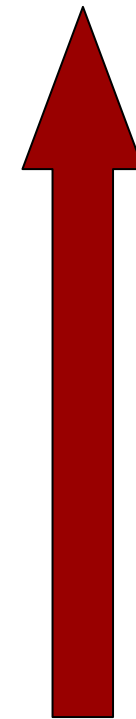


“An assessment of whether interventions have actually produced their intended effects (i.e., produced planned client benefits).”



Typology of Research Designs for Impact Assessments (Rossi, Freeman, & Lipsey, 2004)

1. RCT or True Experiments
2. Quasi-experiments:
 - Regression discontinuity
 - Matched controls
 - Statistically equated controls
 - Generic controls
3. Simple pre- post-design



Rigour

Internal Validity: Definition



“Basic minimum without which any experiment is uninterpretable: Did in fact the experimental treatments make a difference in this specific experimental instance (Campbell & Stanley, 1963, p. 97).”



Question



- What are the challenges of implementing a RCT design for investigating the effectiveness of HF?



RCT “Gold Standard” View of Evidence Hierarchy



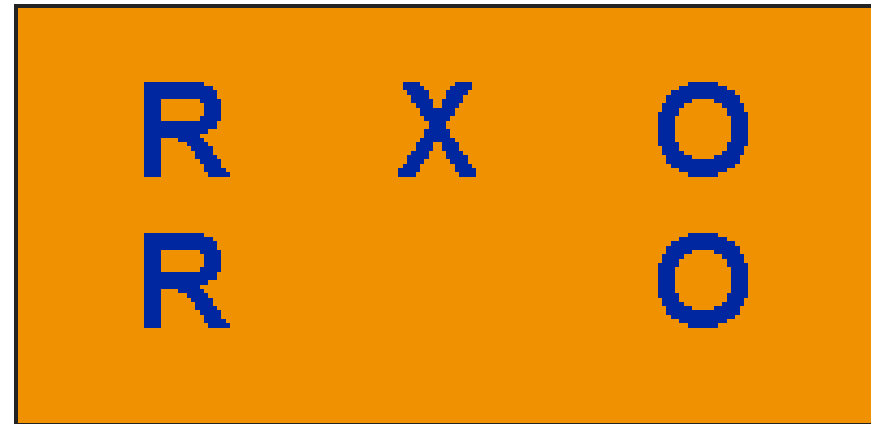
“Well-designed and implemented randomized controlled trials are considered the “*gold standard*” for evaluating an intervention’s effectiveness, in fields such as *medicine*, welfare and employment policy, and psychology.”

(U.S. D.O.E., 2003), p. 1. (emphasis added)



Research Design: (a) Threats for single group designs, (b) Threats for multiple groups, (c) Social Threats (William Trochim)

- history ✓
- maturation ✓
- testing ✓
- instrumentation ✓
- mortality ✓
- regression to the mean ✓
- selection ✓
- selection-history ✓
- selection- maturation ✓
- selection- testing ✓
- selection- instrumentation ✓
- selection- mortality ✗
- selection- regression ✓
- diffusion or imitation ✗
- compensatory equalization ✗
- compensatory rivalry ✗
- resentful demoralization ✗





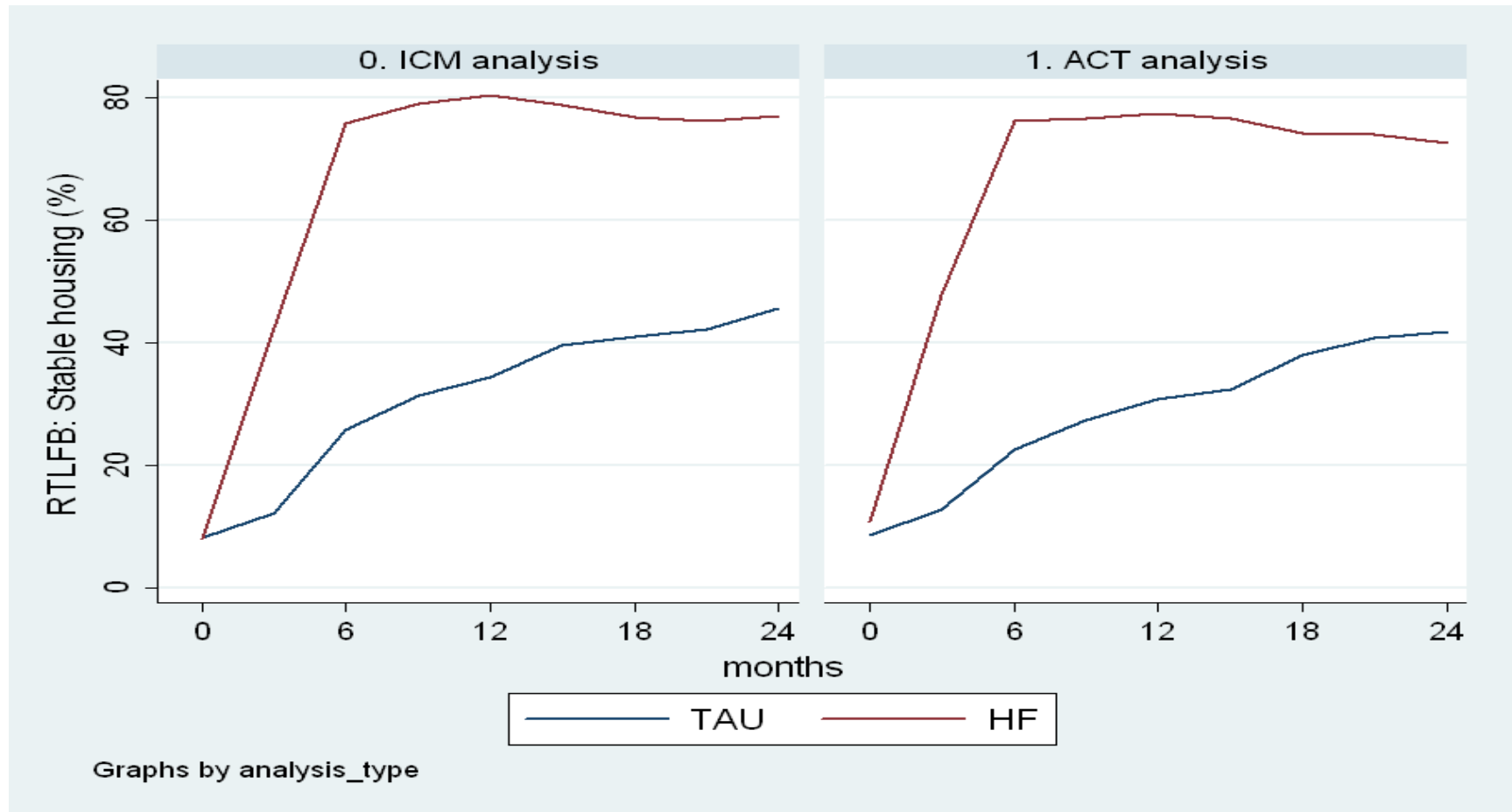
"Are you just pissing and moaning, or can you verify what you're saying with data?"





Housing: Stability – by Program

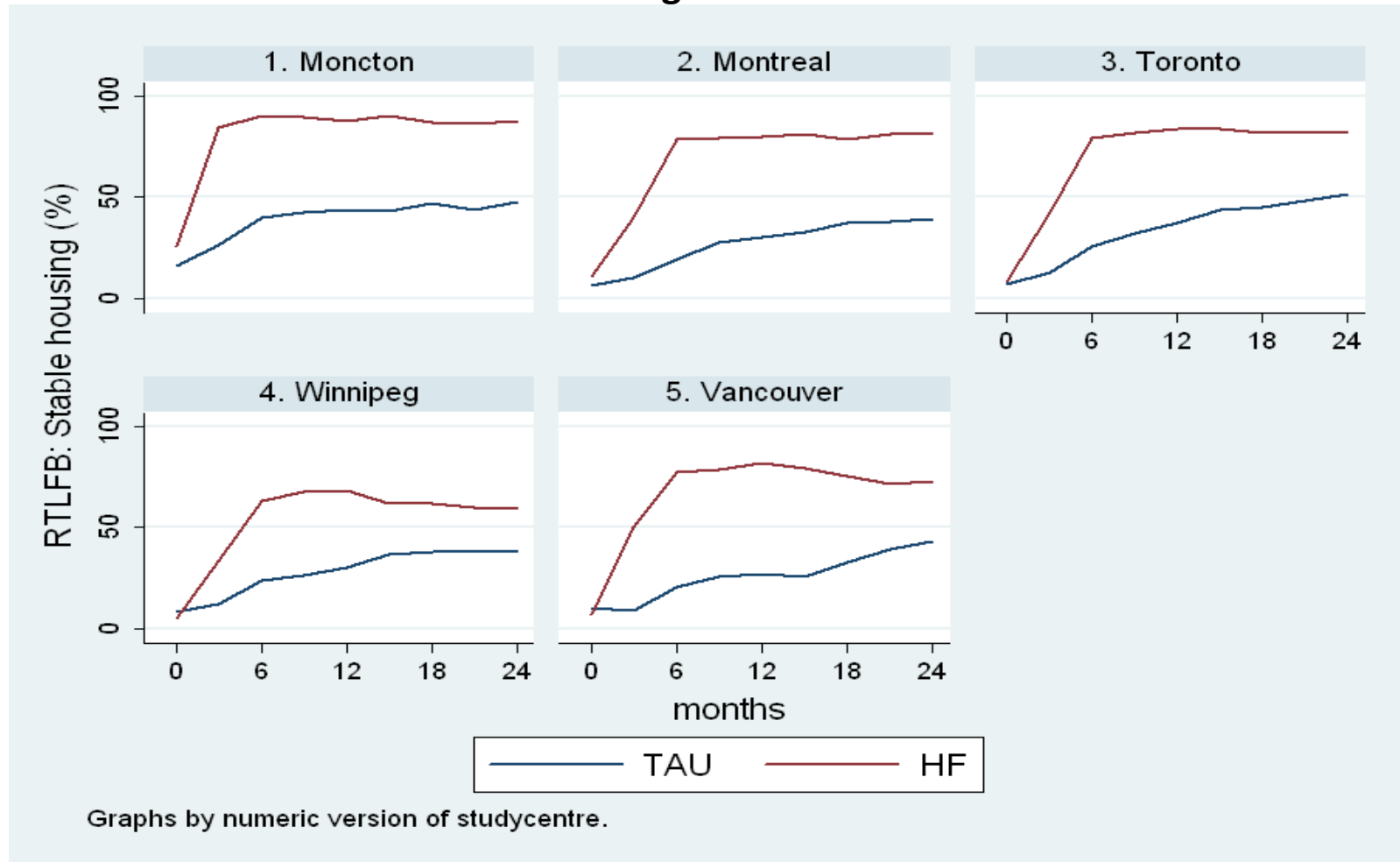
Percentage of time housed





Housing: Stability – by Site

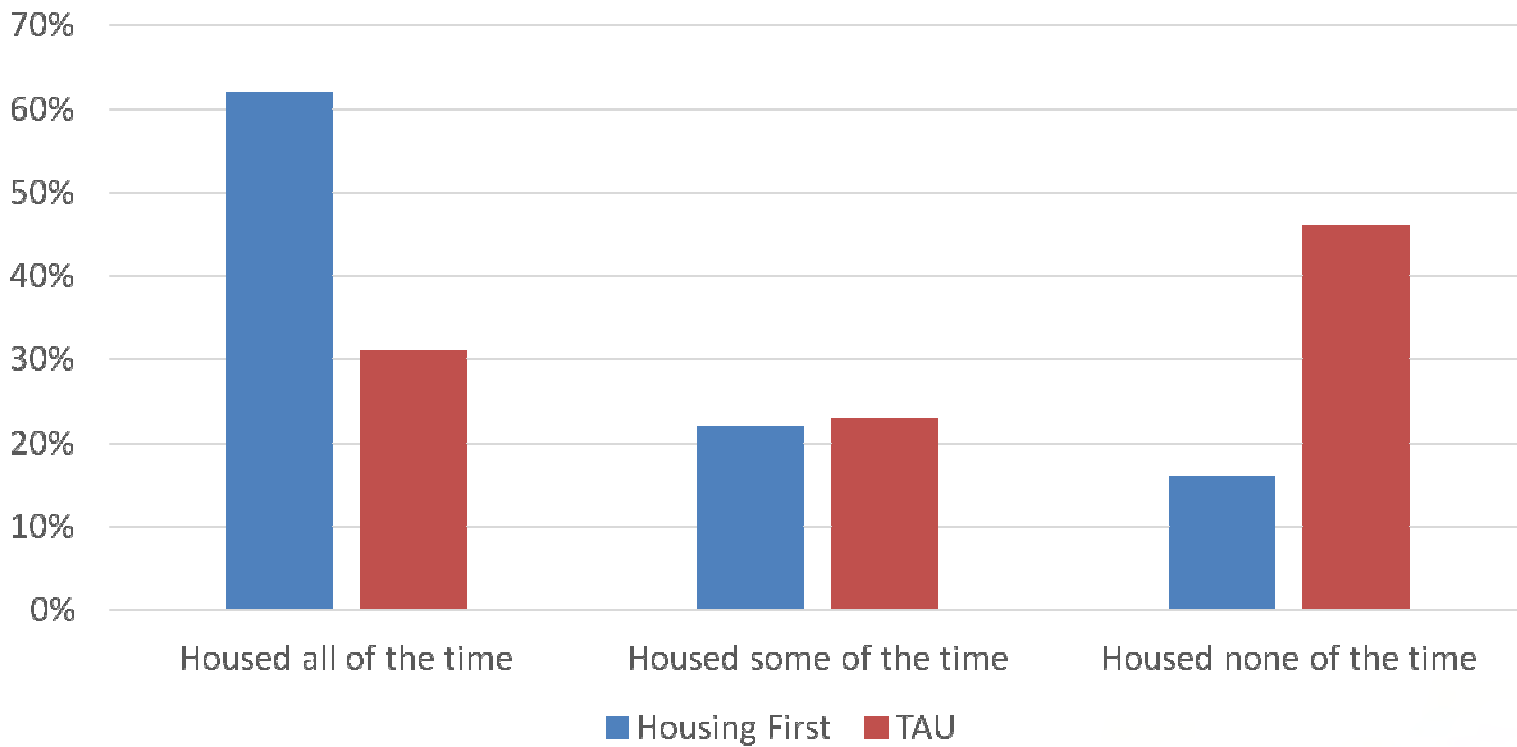
Percentage of time housed



Housing Outcomes of participants

Close to 1,000 Canadians with significant histories of homelessness were successfully housed through At Home / Chez Soi

Percentage of participants housed in the last six months of the study



Housing: Quality

- Measured housing quality through administering the Observer-Rated Housing Quality Scale in a subsample of HF and TAU residences (for those in TAU who found housing for a min. of 2 months)
- Scale measured quality based on the characteristics of the unit, building and neighborhood
- Found significantly higher quality for those in the HF intervention on unit/building scales combined than for those in the TAU group who found housing
- There is more consistency and less variability in the quality for unit/building (in 4/5 sites) of HF compared to TAU

Social Outcome Findings

HF participants in both ICM and ACT reported greater improvements than TAU participants in:

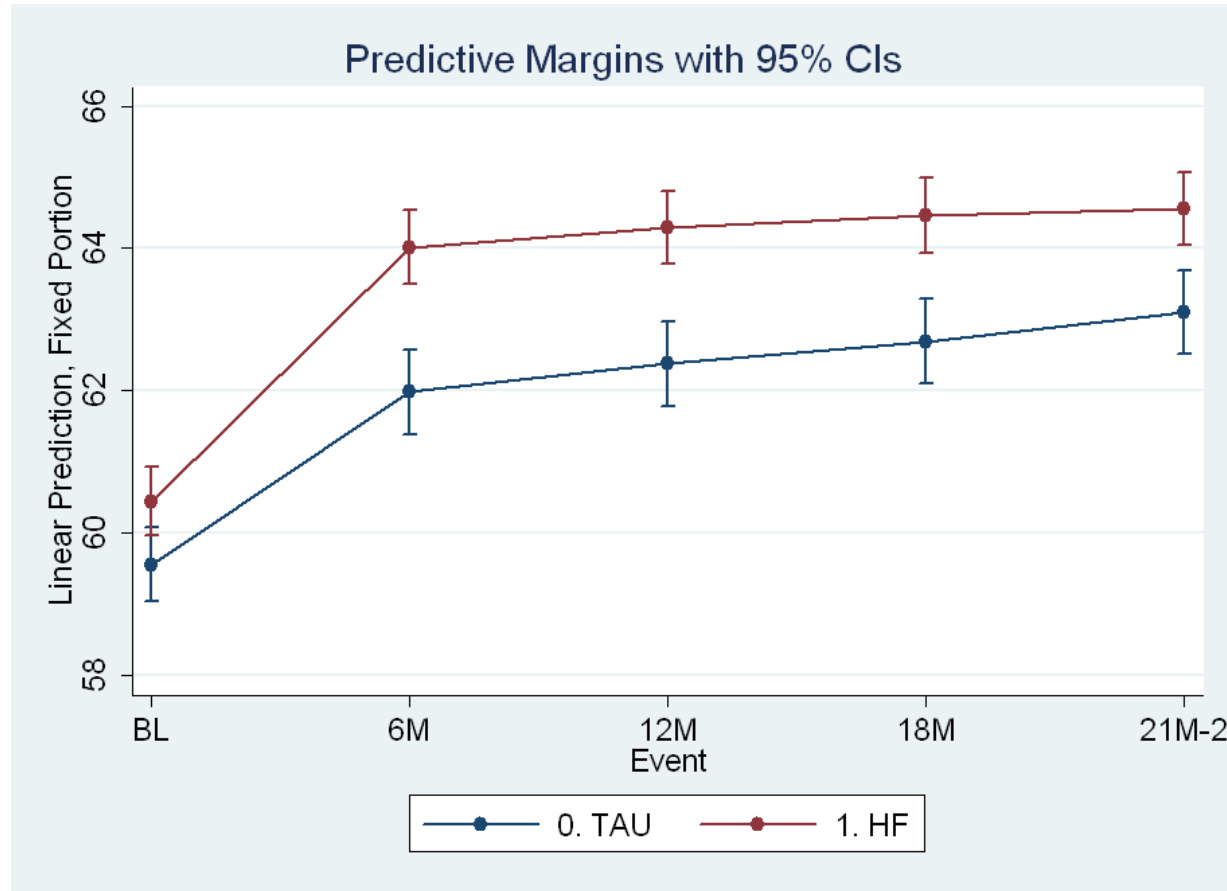
- Community Functioning
- Quality of Life



Mental Health
Commission
of Canada

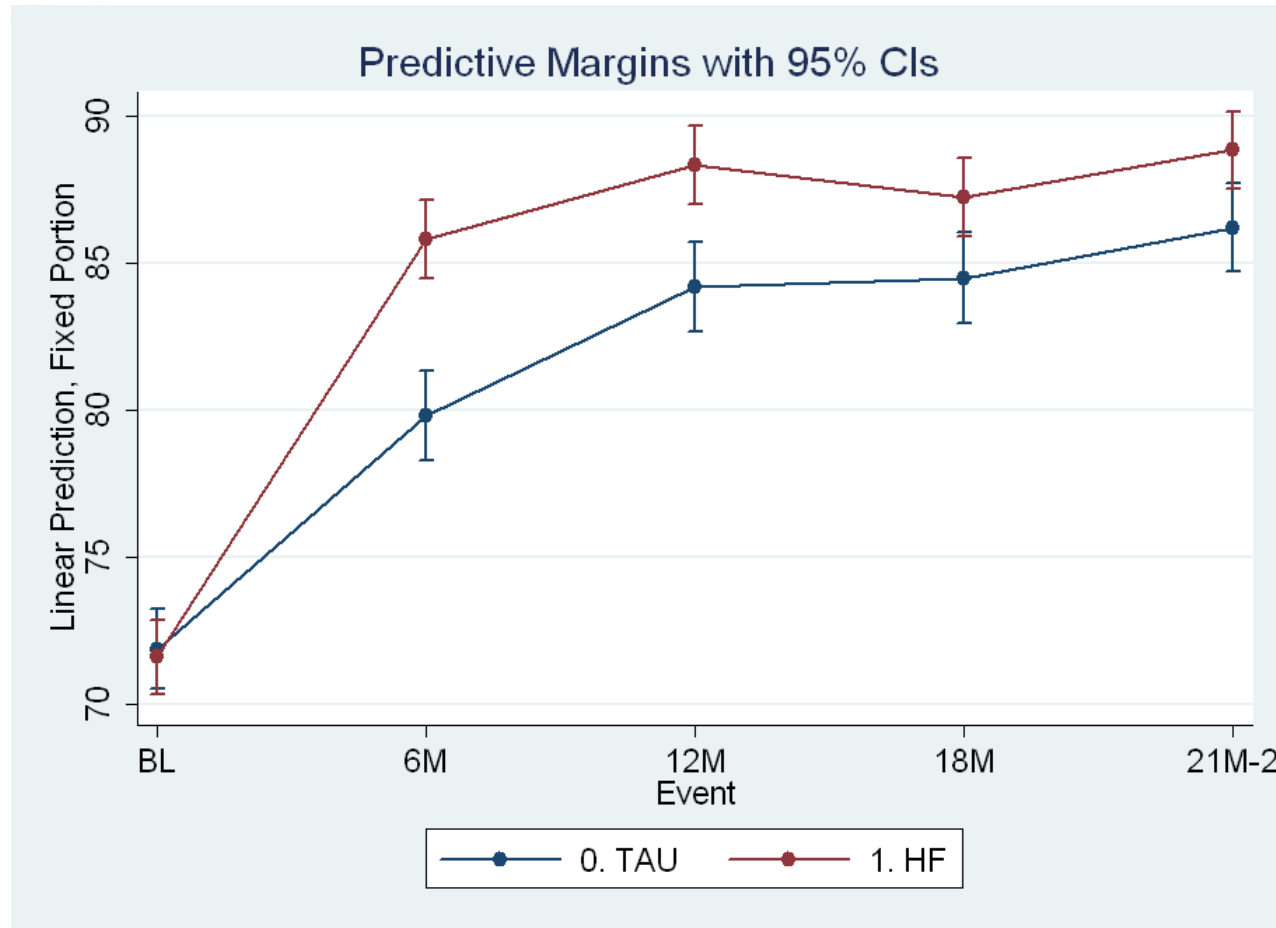
Commission de
la santé mentale
du Canada

Social Outcomes: Community Functioning



Average post BL differences are SIG
for both ICM and ACT

Social Outcomes: Quality of Life



Average post BL differences are SIG
for both ICM and ACT

Health Outcome Findings

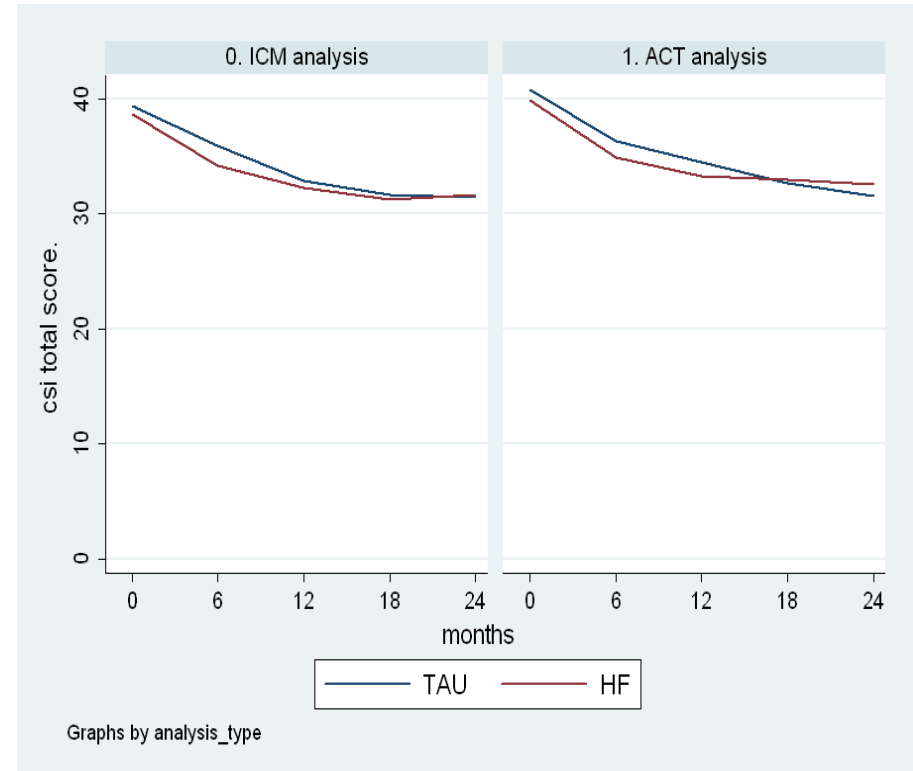
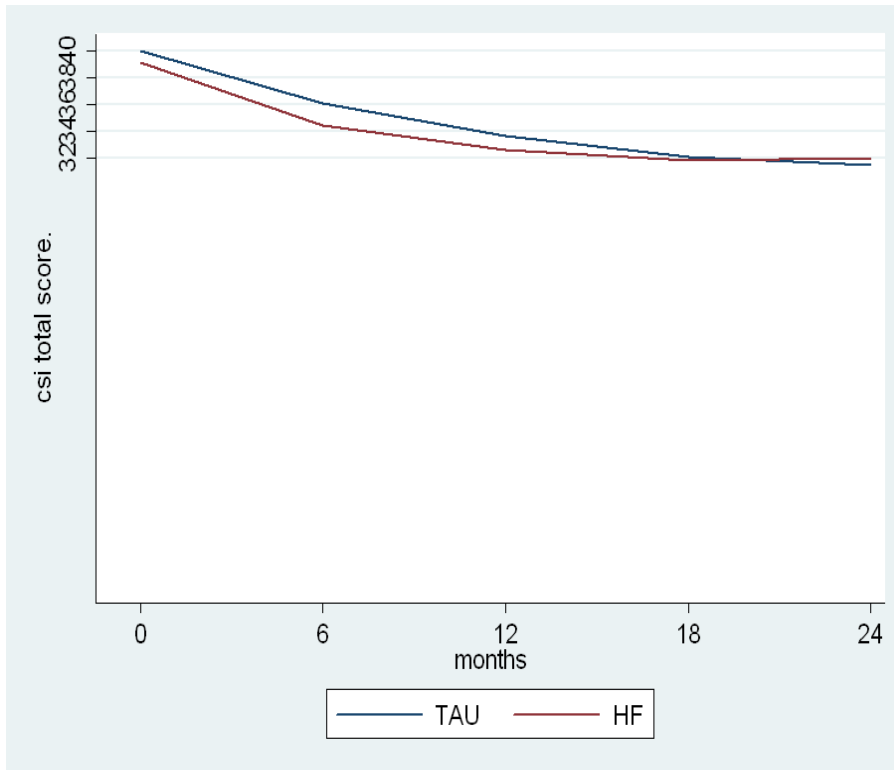
Both groups report improvements in:

- Substance use
- Mental health

Both groups maintained their physical health



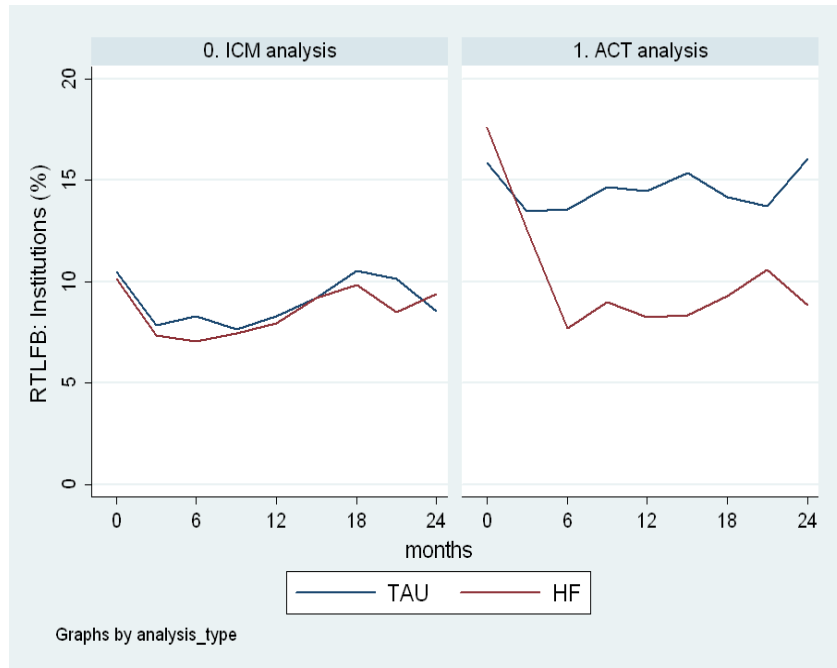
Health Outcomes: Mental Health



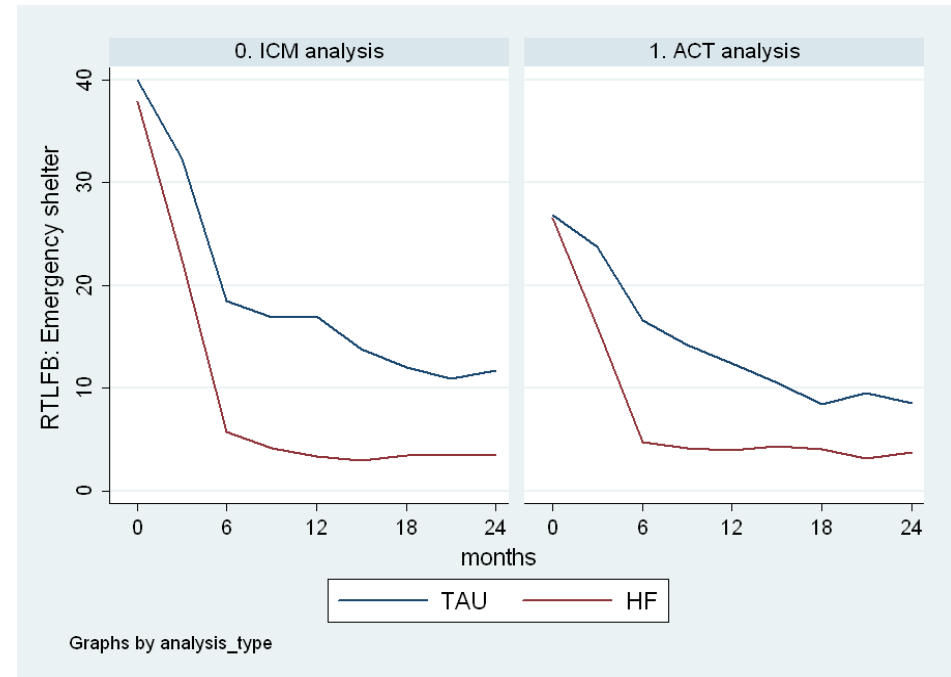


Health/Social Service Use

% Days in Institutions*



% Days in Shelters

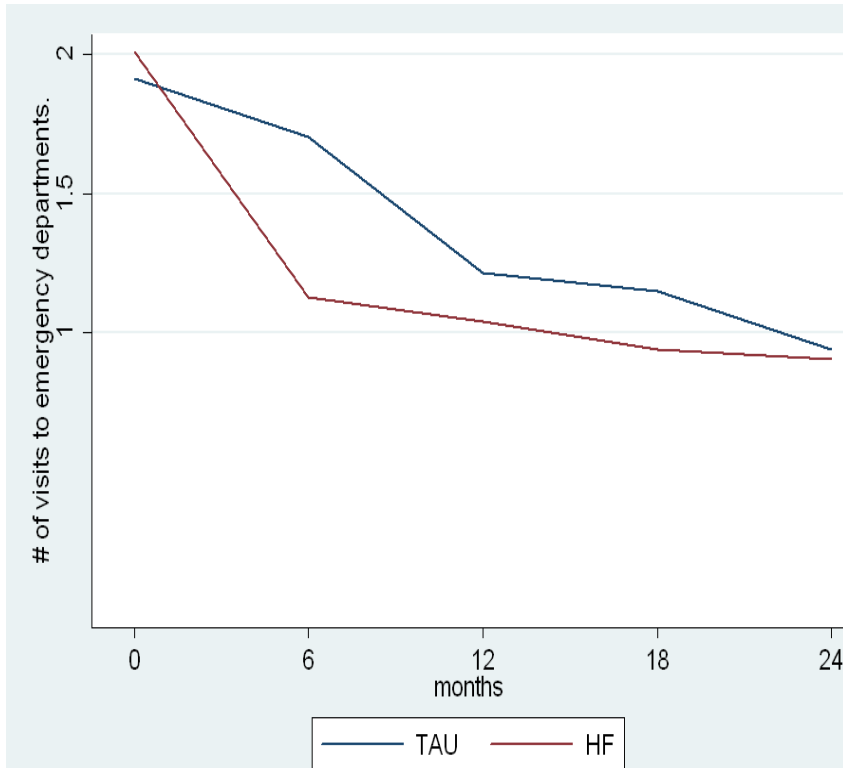


* hospitals, prisons, jails, addiction treatment facilities, etc.

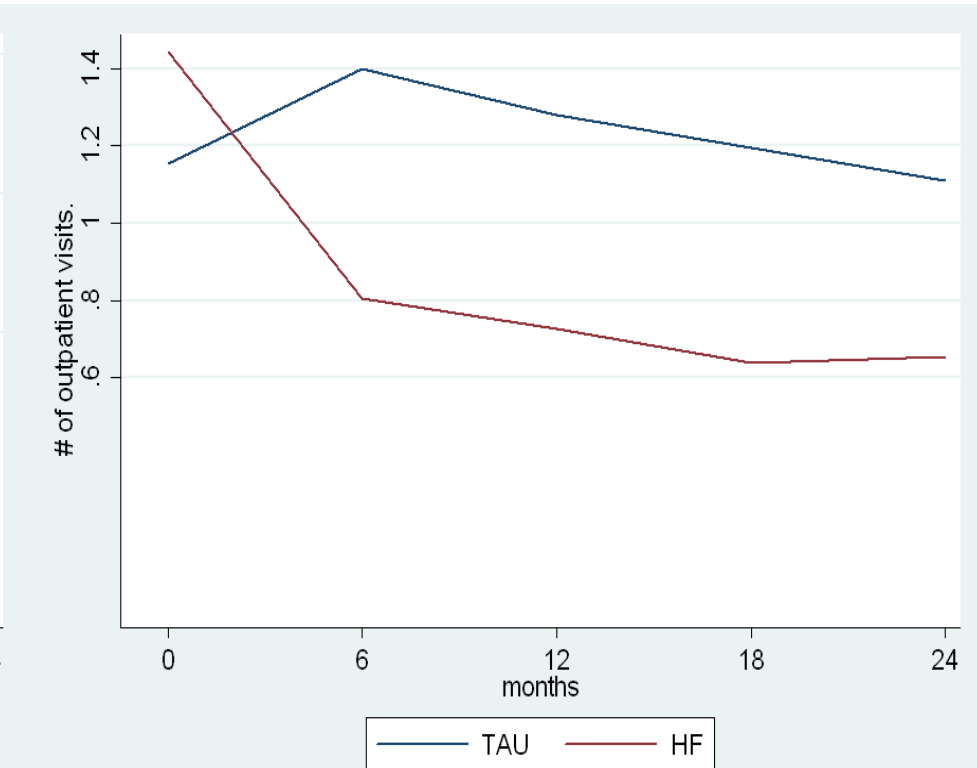


Health Service Use

ER Visits



Outpatient Visits*

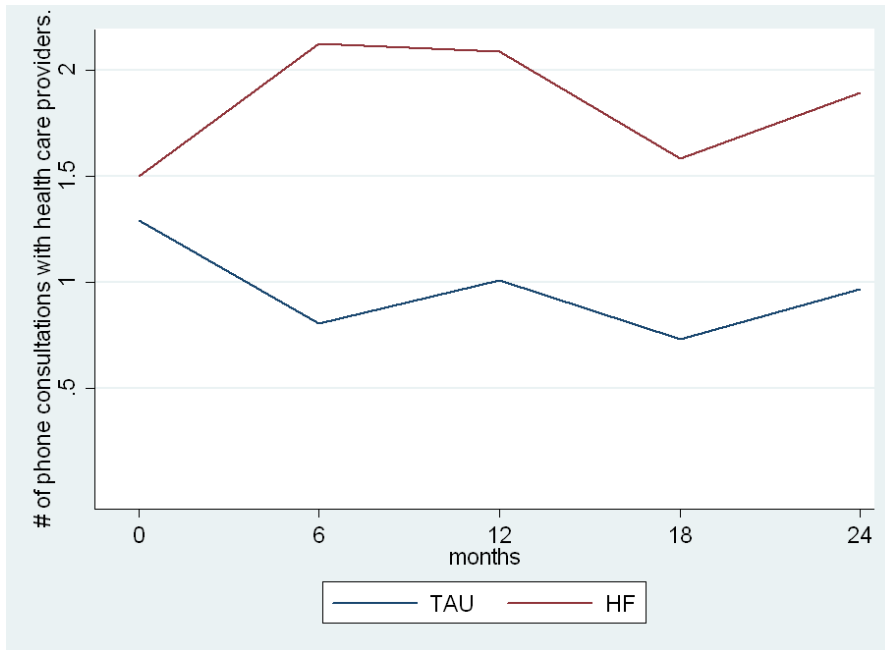


* Services at a hospital on an OP basis incl. day hospitals not incl. not lab or diagnostic test visits

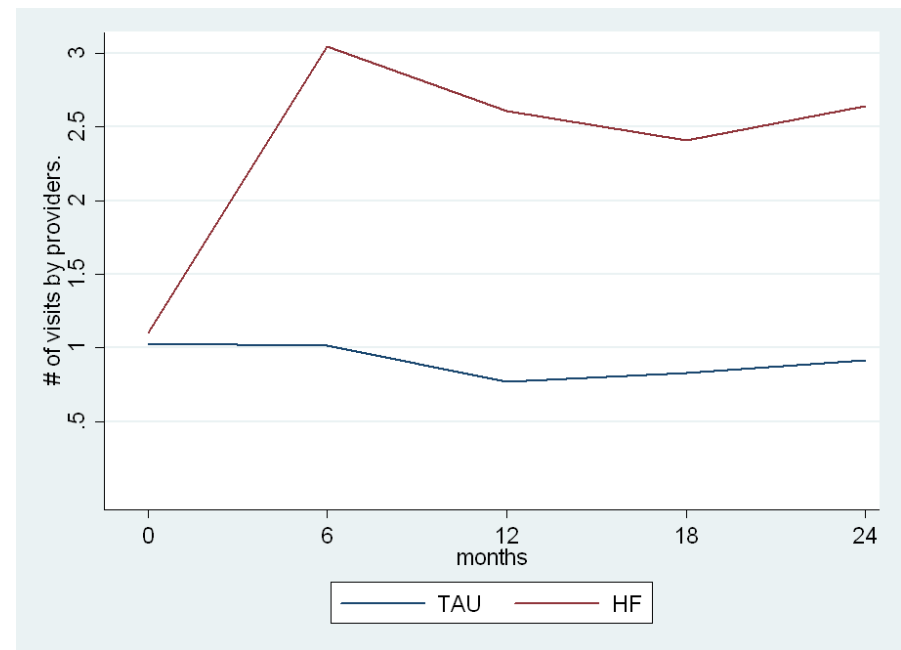


Health Service Use

Phone Calls w/ HSS Providers



Visits from HSS Providers



Pluralist Pragmatist View of Evidence Hierarchies



- Randomized control group trials (RCTs) are not the only studies capable of generating understandings of causality.
- RCTs are not always best for determining causality and can be misleading.
- RCTs should sometimes be ruled out for reasons of ethics.
- In some cases, data sources are insufficient for RCTs. Pilot, experimental, and exploratory education, health, and social programs are often small enough in scale to preclude use of RCTs as an evaluation methodology, however important it may be to examine causality prior to wider implementation.
- Actual practice and many published examples demonstrate that alternative and mixed methods are rigorous and scientific.

American Evaluation Association Response To U. S. Department of Education Notice of proposed priority, *Federal Register* RIN 1890-ZA00, November 4, 2003 "Scientifically Based Evaluation Methods."

<http://www.eval.org/doestatement.htm>

Questions Regarding Use of Quasi-Experimental Design?

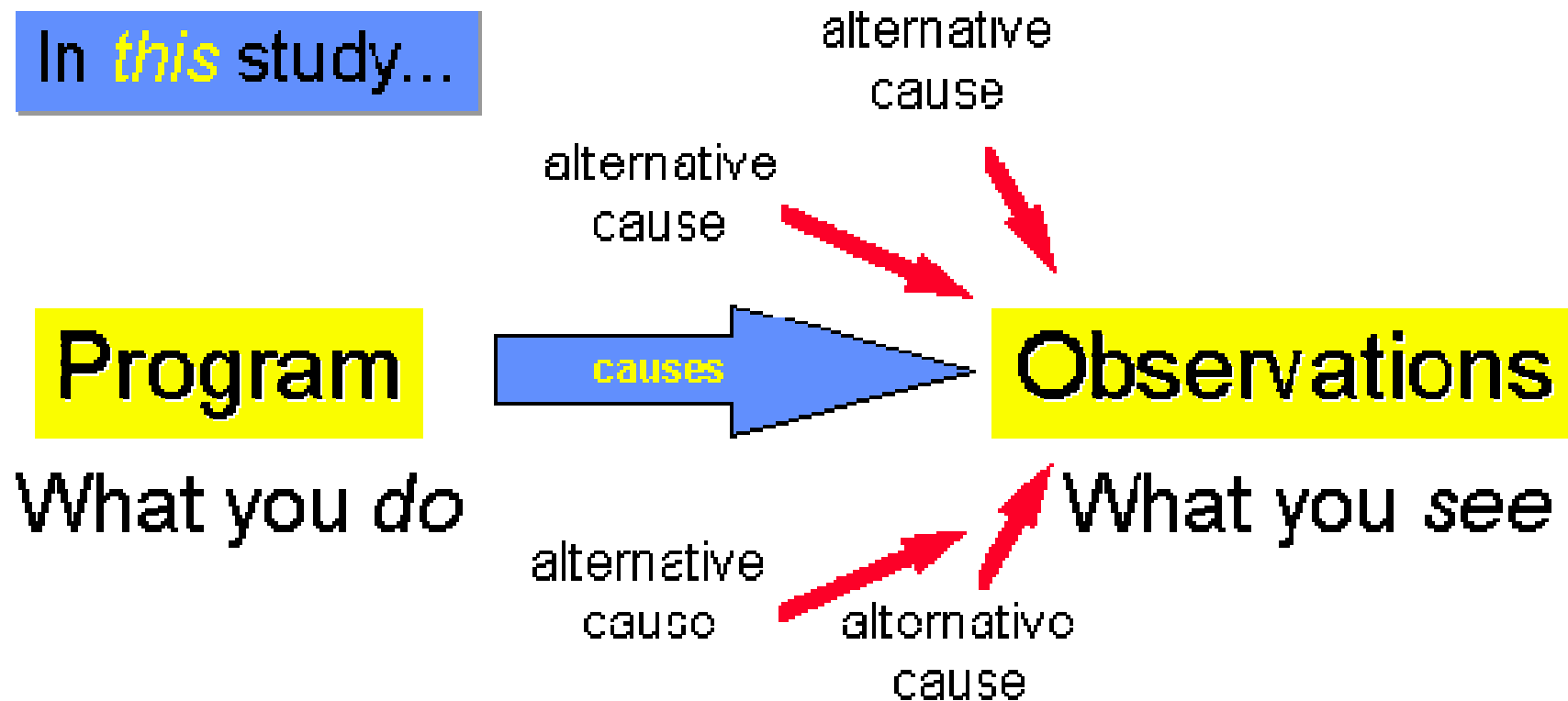


- Have you ever been involved in a study with a quasi-experimental design that includes a non-equivalent comparison group?
- What are the benefits of this design for conducting impact research on HF programs?
- What are the limitations of this design?



The problem with quasi-experimental designs:
identifying the true cause(s) of an outcome
(i.e., the program vs. other possible causes)

In *this* study...



Most frequently used quasi-experimental design:
The non-equivalent groups design (NEGD)
(see Wholey et al., chap. 5; & google Wm. Trochim's
website, *Research Methods Knowledge Base*)

N O X O

N O O

Hypothetical results from NEGD (Trochim):
(X = program [intervention] group;
O = comparison group)

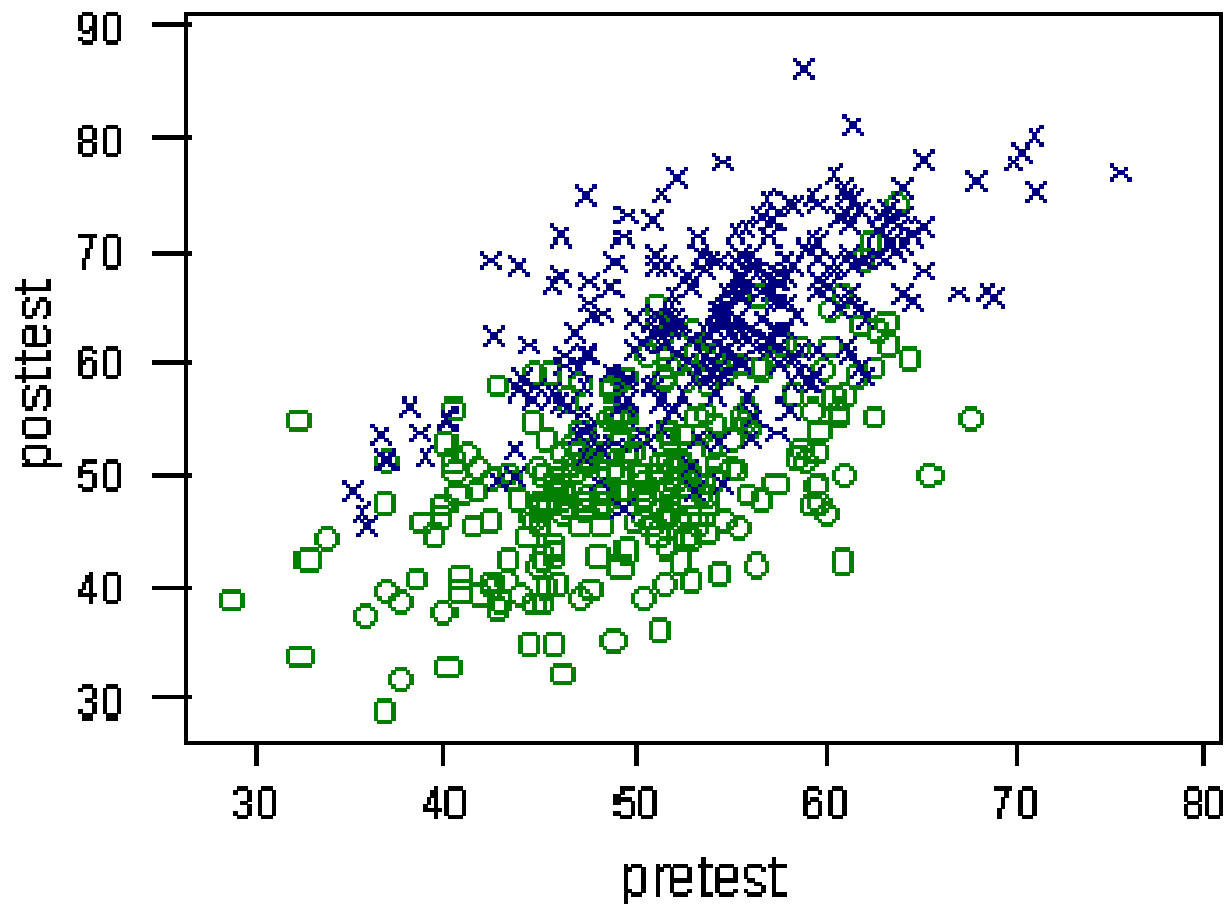
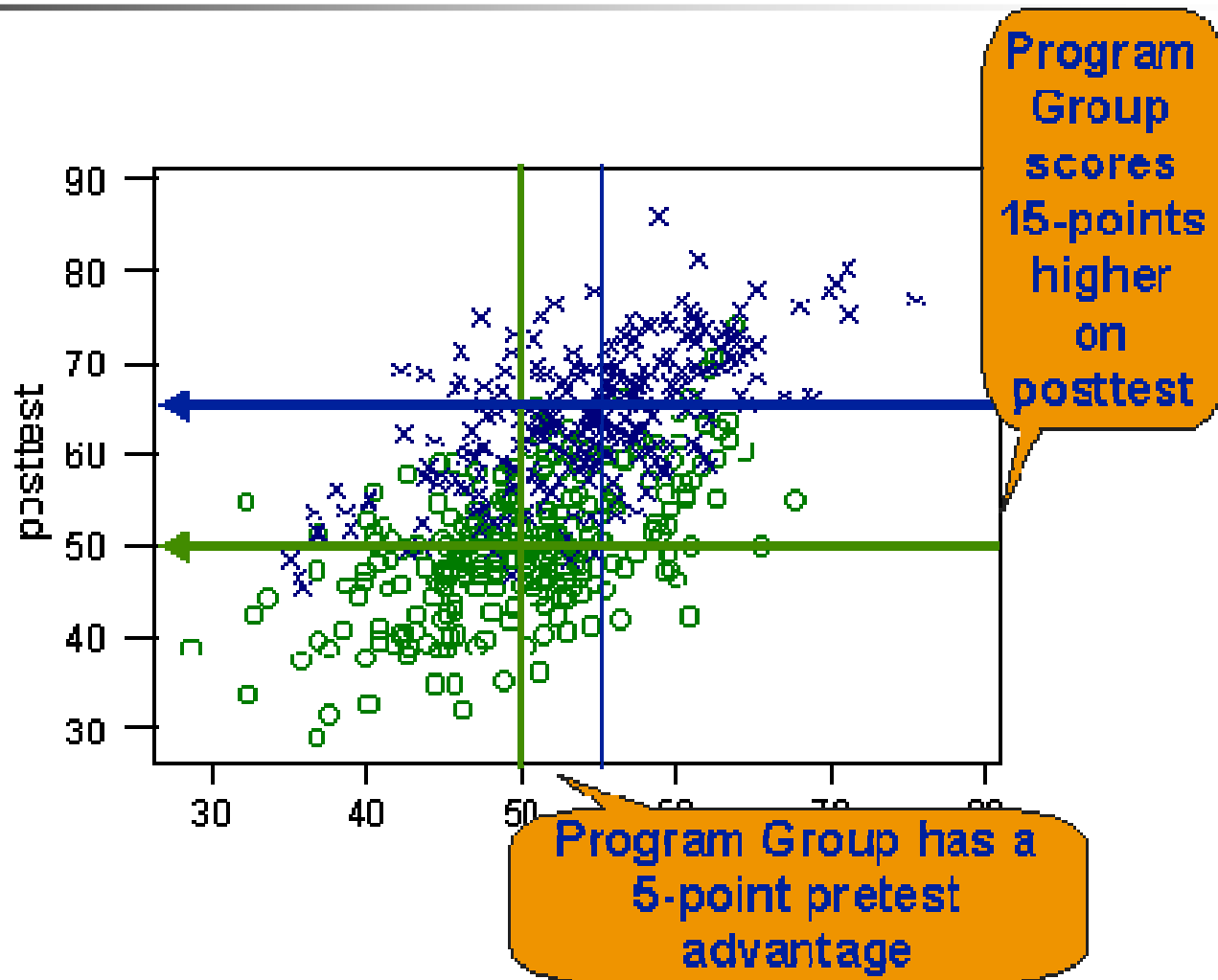


Illustration of *selection bias* in the NEGD: the groups were different before the program. Selection bias is the biggest threat to internal validity in the NEGD.



Performance Measurement (= M no. 1)



Definition: The ongoing collection and analysis of information on indicators of how well programs are performing, especially with regard to the delivery of service (outputs) and the achievement of results (outcomes)

Involves translating program logic models into practical tools for performance monitoring and management

(Rossi, Freeman, & Lipsey, 1999, p. 190)



Performance Measurement (continued)



- Impetus: GPRA (US); federal, provincial, and municipal governments (Canada); United Way (US & Canada)
- Sometimes also called “performance monitoring,” “outcome monitoring,” or “results-based management”
- Especially focused on program outcomes, but also on program outputs



Performance Measurement (continued)



- Program outputs = the services (or “units of service”) delivered to participants
- Indicators or measures of outputs: might include number of clients served, per-child cost of foster care, rapidity and quality of child protection services provided, etc.



Performance Measurement (continued)



- Program outcomes = the benefits derived by recipients of a service
- Indicators or measures of outcomes: might include increased child safety, better family functioning, improved parenting, etc.



Performance Monitoring (= M no. 2)



- Definition: The reporting and interpretation of information on performance indicators, in order to evaluate how well a program is functioning with respect to the delivery of service (outputs), the achievement of results (outcomes), or both
- Sometimes also called “program monitoring” (Rossi et al., 1999, p. 190)



Performance Monitoring (continued)



Involves monitoring three relatively distinct domains of program performance:

- Service utilization: Do the intended participants (i.e., members of the target population) actually receive services?
- Program implementation: Do participants receive the intended services?
- Program outcomes: Do participants derive the intended benefits?



Monitoring of Outputs vs. Outcomes



- Outputs: are easy to measure and monitor and largely under control of staff and management
- Outcomes: are harder to measure and monitor and less controllable, but more central to program mission and more critical to program success



Performance Management (= M no. 3)



Is the ultimate purpose of most performance measurement and monitoring

- Definition: In light of performance measurement and monitoring information, the ongoing taking of corrective action to improve program outputs and/or outcomes
- Examples: increasing the number of families served (if targets are not being met), or decreasing caseloads (if heavier than planned)



Selecting Performance Measures



To select indicators of program outputs and outcomes, imagine that others were looking at your program. Ask yourself:

- What would they look for as evidence of success?
- What would indicate to them that the program was working effectively?
- What would indicate that the program was running efficiently?



Selecting Performance Measures: Criteria



- **Validity:** Does the indicator measure what it is supposed to?
- **Reliability:** Does the indicator measure provide consistent data?
- **Clarity:** Is the indicator understandable to a lay person?
- **Timeliness:** Can the data be gathered in time to be useful?
- **Actionable:** Is the indicator within a program's sphere of influence?
- **Usefulness:** Can management and programs use the data for program improvement purposes?



Types of Outcome Measures



- Standardized self-report measures
- Simple counts of client status
- Collateral rating scales
- Consumer satisfaction measures



Questions Regarding Outputs and Outcomes for HF Programs



- What are outputs of HF programs on which it would be useful and informative to collect data?
- What are outcomes of HF programs on which it would be useful and informative to collect data?



Recommended Outcome Measures of a Minimum Data Set for Monitoring HF Program Outcomes



Primary Outcome Measures	Secondary Outcomes Measures
Residential Timeline Follow Back Calendar (6 mos.)	Outcome Questionnaire - SMI (each contact)
Vocational Timeline Follow Back Calendar (6 mos.)	Gain-SPS (12 mos.)
Multnomah Community Ability Scale (12 mos.)	Health, Social, Justice Service Use (6 mos.)
Alcohol Use Scale-Revised (12 mos.)	
Drug Use Scale – Revised (12 mos.)	
Service Satisfaction Scale (12 mos.)	

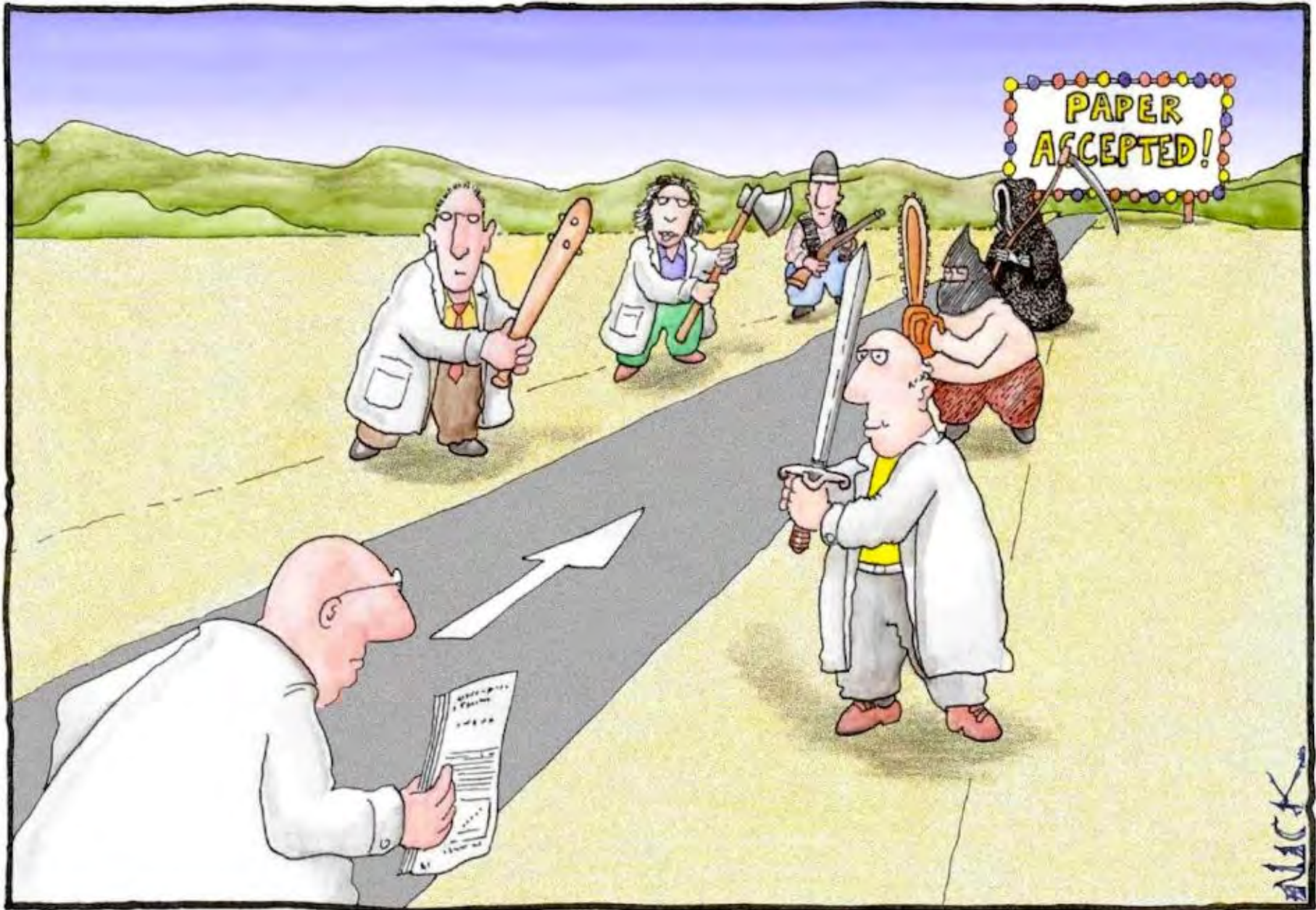


Suggestions for Successful Performance Monitoring and Management



- Integrate data collection into service delivery
- Provide regular training of service providers on data collection
- Keep burden on service providers to a minimum
- Provide regular feedback on collected data
- Keep analysis simple
- Make program improvements based on collected data





Most scientists regarded the new streamlined peer-review process as "quite an improvement."