

Evaluation Dialogue

Between OMB Staff and
Federal Evaluation Leaders

Digging a
Bit Deeper into
Evaluation
Science

April 2005



Why are we here today – How can we benefit from this dialogue?

- Obtain clarity for evaluation community on what approaches appropriate for PART/BPI
- Encourage understanding of evaluation approaches & products generally accepted by evaluators
- Ultimately, we all aim to improve Federal programs, solving problems to increase effectiveness

Format of Dialogue

- Session I: Brief overview of evaluation approaches
- Session II: Examples of evaluation approaches and discussion

What is program evaluation?

- A systematic assessment of how well a program is working
- Consists of various activities:
 - Needs assessment
 - Design assessment
 - Process/ Implementation
 - Evaluability assessment
 - Outcome and Impact
 - (Formative vs Summative)

How are evaluation questions and types relevant to the PART?

- Needs assessments and process studies
 - Relevant to PART Sections 1, 2, & 3
 - PART question 2.6
- Outcomes and impact studies
 - Relevant to PART Section 4
 - PART question 4.5

Why should we conduct evaluations?

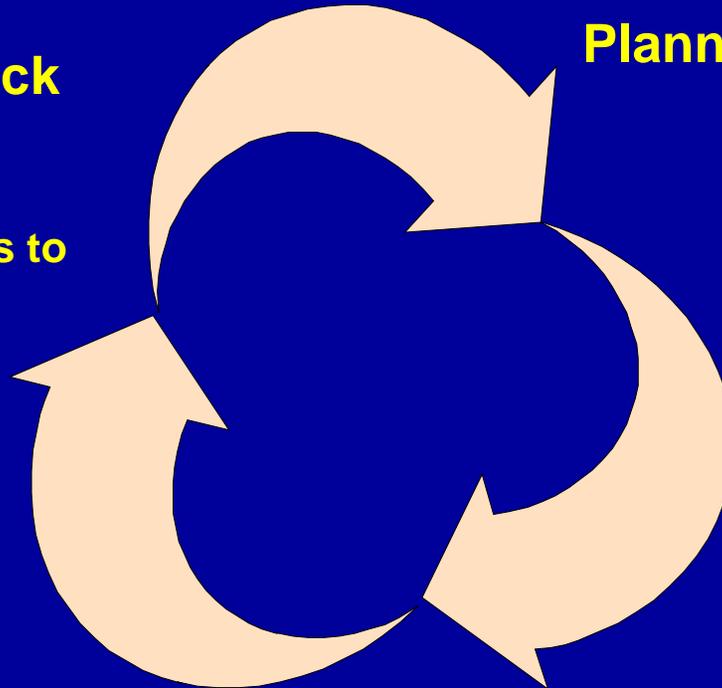
Provide feedback for program improvement and external accountability

- Answer evaluation questions about results and the processes that managers directly control to achieve results
- Document effectiveness and value added to society

Evaluation / Management Cycle

Evaluation Feedback

- Feedback of Evaluation Findings to Managers
- Refinement of Program



Planning/Decision Making

- Identify needs, problems, solutions, refinements
- Conceptualization of Program
- Formulation of Evaluation Questions and Design

Implementation

- Actualize the Program Plan
- Collection of Evaluation Data
- Analysis of Data

Who conducts evaluations?

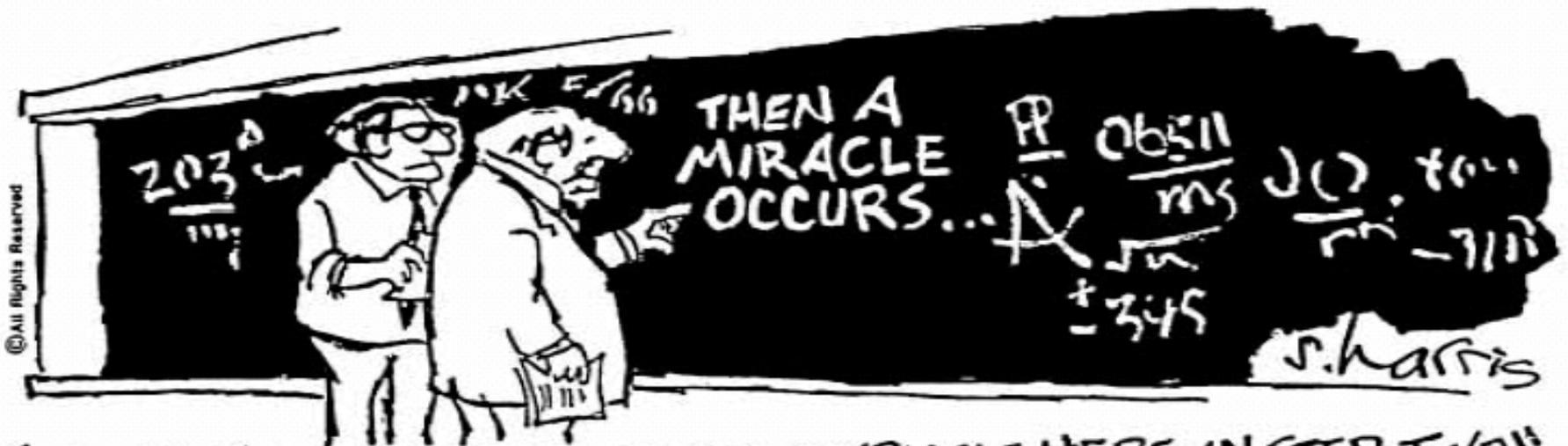
- Professionals blend a wealth of scientific approaches and perspectives
- Within federal agencies, evaluators found in a variety of offices
- Field is supported by professional organizations and degree programs
- (See evaluation information resources handout)

What steps do evaluators use?

- 1. Conceptualize the program**
- 2. Develop relevant and useful evaluation questions**
- 3. Select appropriate evaluation approaches for each evaluation question**
.....
- 4. Collect data to answer evaluation questions**
- 5. Analyze the data and draw conclusions**
- 6. Communicate results and recommendations**

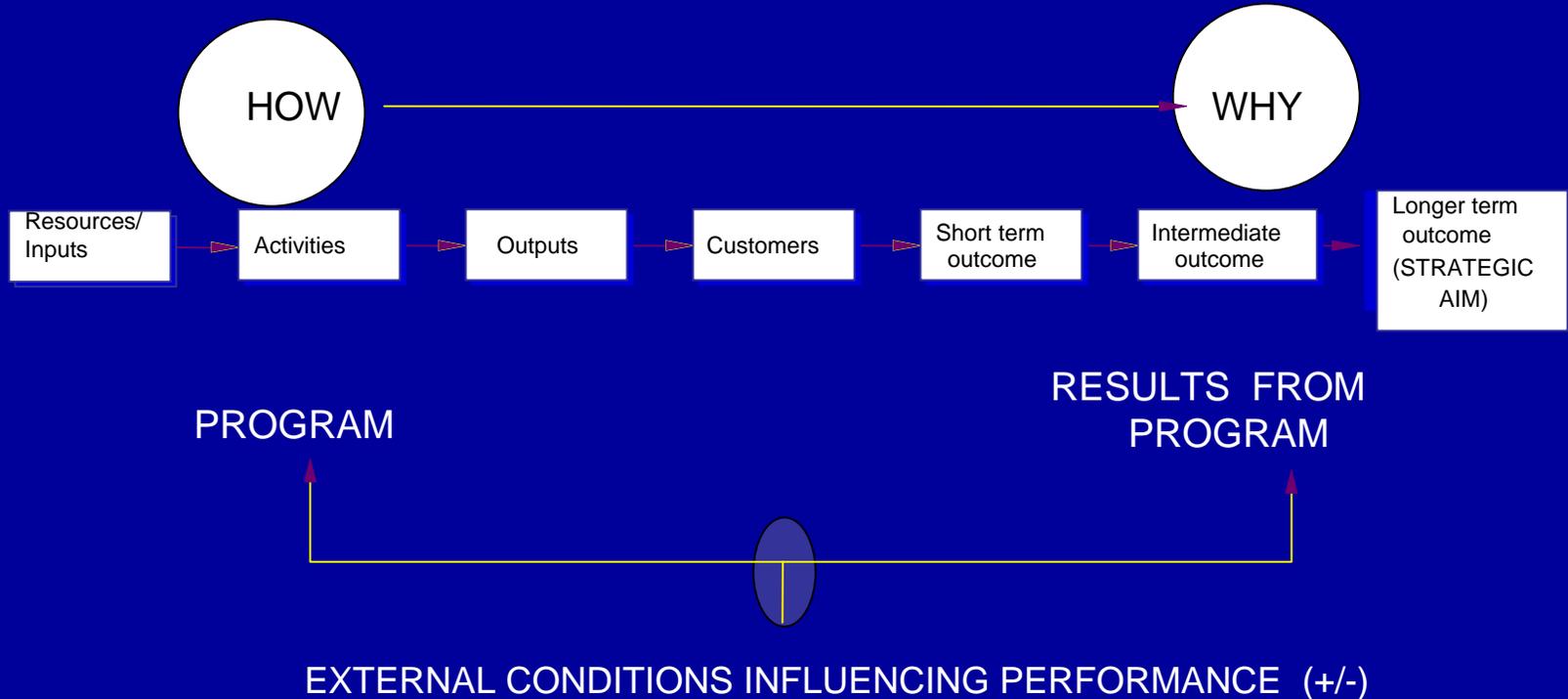
Step 1. Conceptualize the Program by showing simple flow of logic

The Logic Model

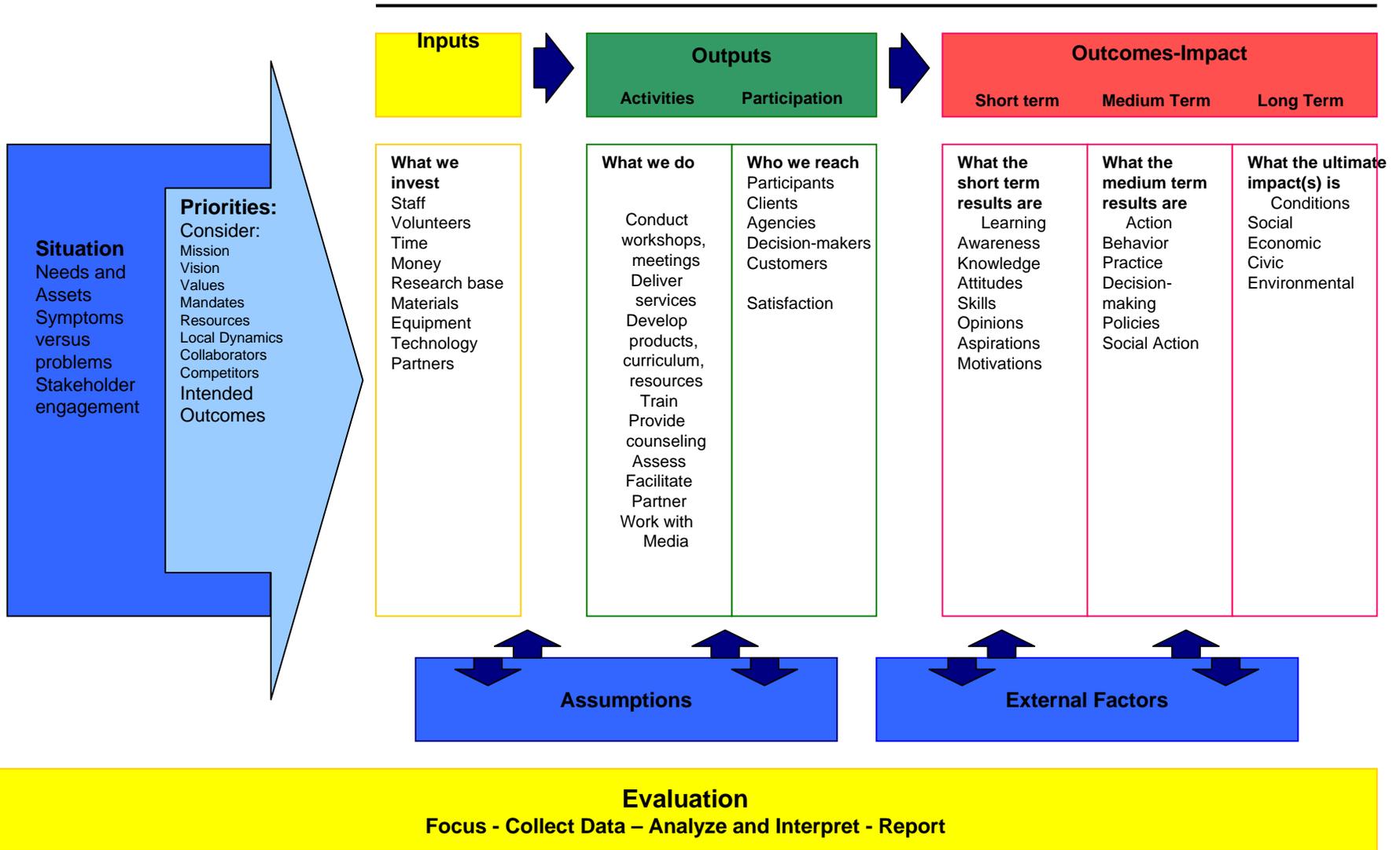


"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Logic models illustrate the causal relationships among program elements and define program success



Generic Logic Model Worksheet



What uses does the Logic Model serve?

- Planning tool
- Communication tool
- Implementation tool
- Evaluation design, data identification and selection tool

Step 2. Develop relevant and useful evaluation questions

Why are good questions important?

- Articulate the issues and concerns of stakeholders
- Posit how the program is expected to work and its intended achievements
- Frame the scope of the assessment
- Drive the evaluation design

Types of evaluation questions as they fit into the logic model

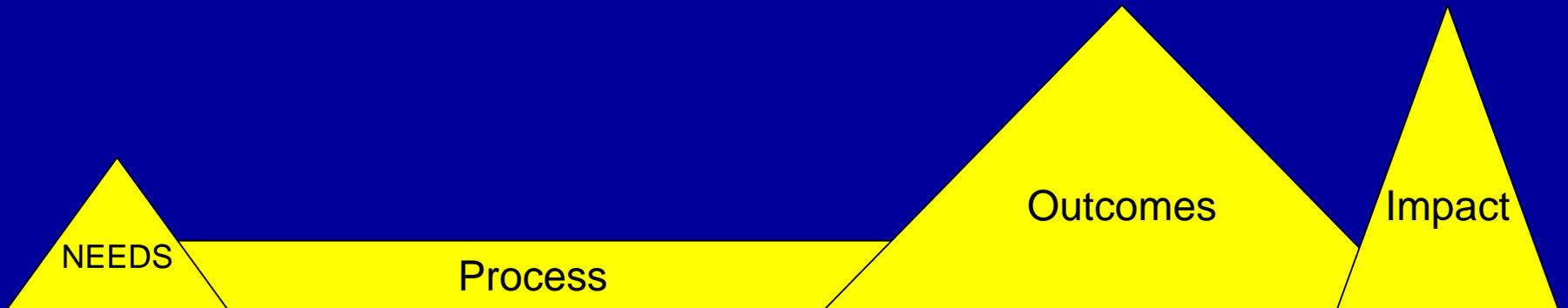
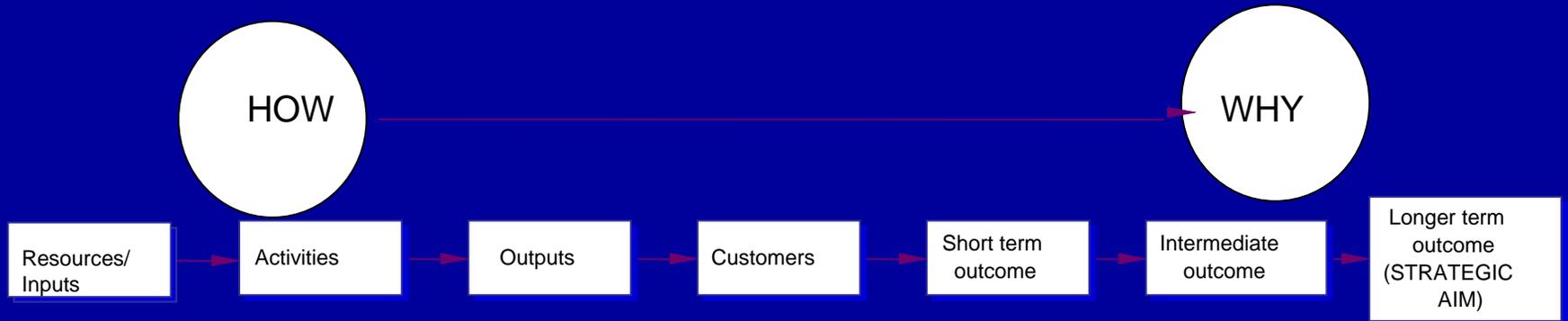


Table 1: Common Evaluation Questions Asked at Different Stages of Program Development

Program Stage	Type of Activity	Common Evaluation Questions
Program design	1. Needs assessment	<ul style="list-style-type: none"> •What are the dimensions of the problem and the resources available to address it?
	2. Design assessment	<ul style="list-style-type: none"> •Is the design of the program well formulated, feasible, and likely to achieve the intended goals?
Early stage of program or new initiative within a program	Process evaluation or implementation assessment	<ul style="list-style-type: none"> •Is the program being delivered as intended to the targeted recipients? •Is the program well managed? •What progress has been made in implementing new provisions?
Mature, stable program with well-defined program model	1. Evaluability assessment	<ul style="list-style-type: none"> •Is the program ready for an outcome or impact evaluation?
	2. Outcome monitoring or evaluation	<ul style="list-style-type: none"> •Are desired program outcomes obtained? •Did the program produce unintended side-effects?
	3. Process evaluation	<ul style="list-style-type: none"> •Why is a program no longer obtaining desired outcomes?
	4. Net impact evaluation	<ul style="list-style-type: none"> •Did the program cause the desired impact? •Is one approach more effective than another in obtaining the desired outcomes?

Step 3. Select appropriate evaluation approaches to answer evaluation questions

How do we control for alternative explanations of effects?

- Ensure conditions necessary for establishing causality
- Use design elements that control for alternative explanations
- Use multiple indicators
- Build strong argument

How do study designs use comparisons to tie observed changes to the program goals?

Compare differences in program entity or condition to differences in:

- Entity or condition without the program
- Itself using data collected over many points in time
- A standard

What are the criteria to select an evaluation design?

- Matches evaluation question
- Fits available resources
 - Time and Funds
- Data availability / Acquisition of data
- Appropriate to the program type
 - Regulatory, Research, Service Delivery

What are good fits of evaluation designs to evaluation questions and types of programs?

- Process and outcome monitoring
- Quasi-experimental – Single group
- Quasi-experimental – Comparison groups
- Randomized experiments

Table 2: Common Evaluation Approaches For Assessing Program Effectiveness

Typical designs used to assess program effectiveness	Design features that help control for alternative explanations	Best suited for (typical examples)
Process and outcome monitoring or evaluation	<p>Compares performance to pre-existing goal or standard. For example:</p> <ul style="list-style-type: none"> • OMB R&D criteria of relevance, quality and performance. • Productivity, cost effectiveness and efficiency standards 	<p>Research, enforcement, information and statistical programs, and business-like enterprises with few, if any, alternative explanations for observed outcomes.</p> <ul style="list-style-type: none"> • ongoing programs producing goods and services • complete national coverage
Quasi-experiments – Single Group	<p>Compare outcomes for program participants or entities before and after the intervention.</p> <ul style="list-style-type: none"> • Multiple data points over time are necessary. • Control for alternative explanations by statistical adjustments and analyses such as modeling. 	<p>Regulatory and other programs with</p> <ul style="list-style-type: none"> • clearly defined interventions with distinct starting times • complete national coverage • random assignment of participants or entities to groups is NOT feasible, practical, or ethical.
Quasi-experiments – Comparison Groups	<p>Compares outcomes for program participants or entities with outcomes for a comparison group selected to closely match the “treatment” group on key characteristics.</p> <ul style="list-style-type: none"> • Key characteristics are plausible alternative explanations for the outcome. • Measure outcomes before and after intervention (pretest, posttest). 	<p>Service and other programs with</p> <ul style="list-style-type: none"> • clearly defined interventions that can be standardized and controlled • limited national coverage • random assignment of participants or entities to groups is NOT feasible, practical, or ethical.
Randomized experiments	<p>Compares outcomes for randomly assigned program (treatment) participants or entities with outcomes for a randomly assigned “control” group prior to intervention.</p> <ul style="list-style-type: none"> • Measure outcomes before and after intervention (pretest, posttest). 	<p>Service and other programs with</p> <ul style="list-style-type: none"> • clearly defined interventions that can be standardized and controlled • limited national coverage • random assignment of participants or entities to groups is feasible and ethical.

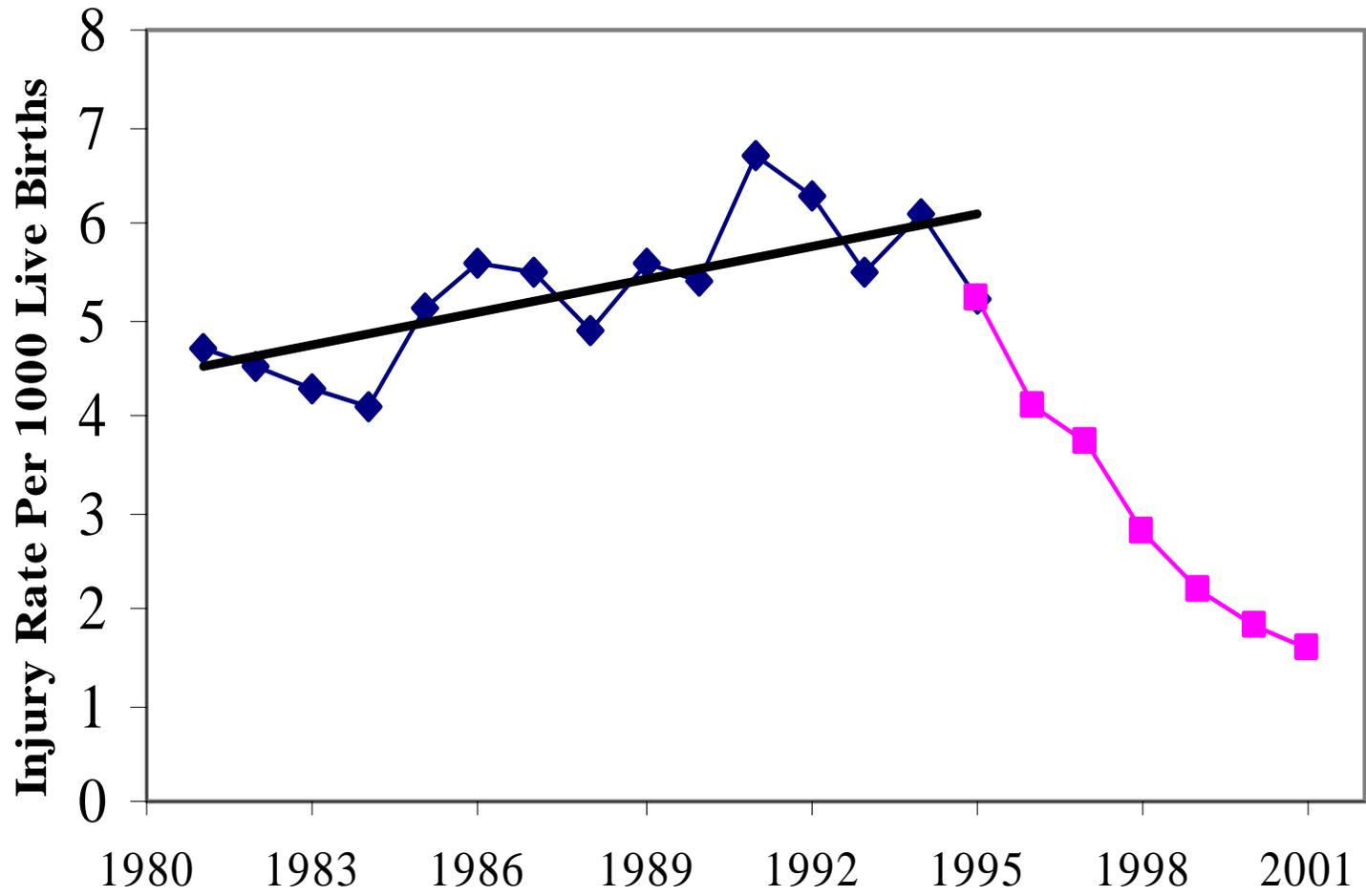
Example of Outcome Monitoring: Medfly Program

- Is the program controlling the Medfly population at an appropriate target level?
- Weekly monitoring of the Medfly population level and dispersion to detect outbreaks
- Review program policies, practices, and resources to identify causes of outbreaks

Example of Quasi Experimental Single Group Design: Baby Walker

- Has the safety standard been effective in reducing injuries?
- Interrupted time series compared program pre and post standard introduction.
- Controlled for alternative explanations by measurements and logical elimination of identified possible alternative explanations

Baby Walker-Related Injury Rate: 1981 to 2001



Example of Quasi Experimental Comparison Group Design: GI Bill

- Did educational assistance meet needs of beneficiaries (veterans)?
- Compared program users with non users on education achievement, income attainment, & career goals
- Statistically controlled for demographic, educational level & military rank characteristics

Example of Randomized Design: Upward Bound

- Does the program help low income, academically high-risk students complete high school and attend college?
- Applicants were randomly selected to the program and compared to non-selected applicants
- Random assignment controlled for many alternative explanations, such as demographics and motivation level

How do we determine the quality of an evaluation?

- Evaluation questions have been answered fully
- Findings support conclusions
- Conclusions portray strong causal arguments
- Meets professional evaluation standards
 - Utility, Feasibility, Propriety, and Accuracy

Checklist of Questions for Assessing the Quality and Usefulness of a Program Evaluation

Are the study's objectives stated?

Were the objectives appropriate with respect to the development stage of the program?

Is the study design clear?

Was the design appropriate given the study objectives?

Was the indicated design in fact executed?

Did the variables measured relate to and adequately translate to the study objectives and are they appropriate to the study objectives and are they appropriate for answering the client's questions?

Are sampling procedures and the study sample sufficiently described? Were they adequate?

Are sampling procedures such that policymakers can generalize to other persons, settings, and times of interest to them?

Is an analysis plan presented and is it appropriate?

Were data-collector selected and training adequate?

Were there procedures to ensure reliability across data collectors?

Were there any inadequacies in data collection procedures?

Were problems encountered during data collection that affect data quality?

Are the statistical procedures well specified and appropriate to the task?

Are the conclusions supported by the data and the analysis?

Are study limitations identified?

What possibly confounds the interpretation of the study findings?

How can we work together to ensure the best evaluations?

- Work to understand the program concept via logic model and/or the strategic plan
- Develop good evaluation questions
- Select appropriate evaluation study designs to answer questions
- Use conceptualization to identify needed performance measures

Federal Evaluation Leaders

Working with OMB to dig up the best evaluation information possible!



Contributor Acknowledgements

- LCdr Eric Bernholz, USCG
- Alan Ginsburg, ED
- Marcelle Habibion, VA
- John Heffelfinger, EPA
- David Introcaso, HHS
- Cheryl Oros, USDA
- N.J. Scheers, CPSC
- Stephanie Shipman, GAO
- Linda Stinson, DOL
- Bill Valdez, DOE