AMERICAN EVALUATION ASSOCIATION
www.eval.org

Research, Technology, & Development
Topical Interest Group

Evaluating Outcomes of Publicly Funded Research, Technology and Development Programs: Recommendations for Improving Current Practice

Version 1.0 Draft Final

By the Research, Technology and Development Topical Interest Group of the American Evaluation Association (AEA)

Presented at

American Evaluation Association Conference

October 17, 2014

Gretchen Jordan, RTD TIG Co-Chair

Lunch provided courtesy of

Alberta Innovates – Health Solutions

&

The American Evaluation Association

RTD TIG

Presentation of the RTD TIG Paper Outline

- Purpose, scope, need
- Evaluation context
- Recommendations: Evaluation planning
- Generic logic models and indicators
- Recommendations: Methods
- Recommendations: all
- Summary, next steps

Purpose, Approach

- The purpose of this paper is engage RTD evaluators, program managers, and policy makers in a dialogue about a current RTD evaluation practice and how it might be improved.
- The end goal is consensus on a common RTD evaluation language and practice that is then broadly implemented.
- This is needed because the diversity in RTD programs leads to evaluation without enough consideration of context.

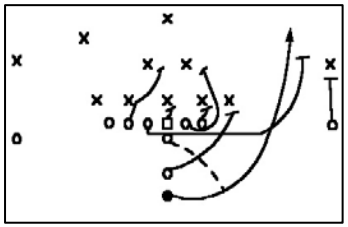
Approached through

  - Review of US government, national academy guidance and other literature,
  - Our years of practical experience, and
  - Expert review (written and in workshops)

Scope is Broad But Not Comprehensive

- Publicly funded
- Program level
- All aspects: research, technology, development and deployment
- Including innovation, defined as a new product, process or organizational practice that is entering the "market"
- Outcomes before, during and after (life cycle)
- Program contribution to outcomes
- Purpose: both accountability and learning

Need to Connect the Parts, Think "Contributing to a Causal Package"



Use
Effects of Use

Context: Micro, Meso, Macro

Players and Plays

Deliver

Produce

Operate, Maintain

Adopt a Process

Put Into Practice

Public Action

Incentivize, Restrict

Manufacturing Research

Commercialization Research

Policy Action

Advocacy

Discovery Research

Applied Research

Development Research

Research to Inform Debate

Build Knowledge, Skills, Abilities

Build/Maintain Infrastructure

Research Education Technical Connectedness Government Business

The Need For Common Frameworks and Practice and Comparable Studies

- Ideally there would be sufficient data and theory to enable policy makers to better target interventions, even to the point of comparing the cost, size and speed of pay off among alternatives.
- That is not possible because
  - There is no grand theory that connects types of interventions, contexts.
  - Current knowledge is not always used to inform RTD evaluation.
- To build data and theory for the innovation process and system, there will need to be multiple studies and synthesis across those.
- Synthesis is easier if studies use similar terminology, good research design, and make clear the full context of an intervention.

Current Context for RTD Evaluation in U.S.

- GPRA Modernization Act of 2010 (GPRAMA 2010), Office of Management and Budget (OMB) Circular A-11, and OMB/OSTP Annual Memo on Budget Priorities
  - require performance planning, measurement and evaluation;
  - see evaluation as an important tool
- GPRAMA has increased emphasis on cross-organization collaboration and government-wide priority setting.
- Progress has been made in RTD evaluation since 1993.
  - National Academy and studies and agency-sponsored guidelines/ frameworks
  - For all types of RTD programs
  - For science, economic, and social impacts
- Contention exists over appropriate evaluation design and methods.

Relationship to AEA "Evaluation Roadmap for Effective Government"

While we endorse all of the 17 recommendations, we have singled out two of them to expand upon for RTD programs:

1. Build into each new program and major policy initiative an appropriate evaluation framework to guide the program or initiative throughout its life.
2. Promote the use and further development of appropriate methods for designing programs and policies, monitoring program performance, improving program, operations, and assessing program effectiveness and cost.

Recommendations for Planning and Implementing Evaluation in RTD Programs

- Recognize evaluation as a management tool to be used across the program life cycle
- Use different types of evaluations to answer different questions
- Plan evaluations around a logical framework

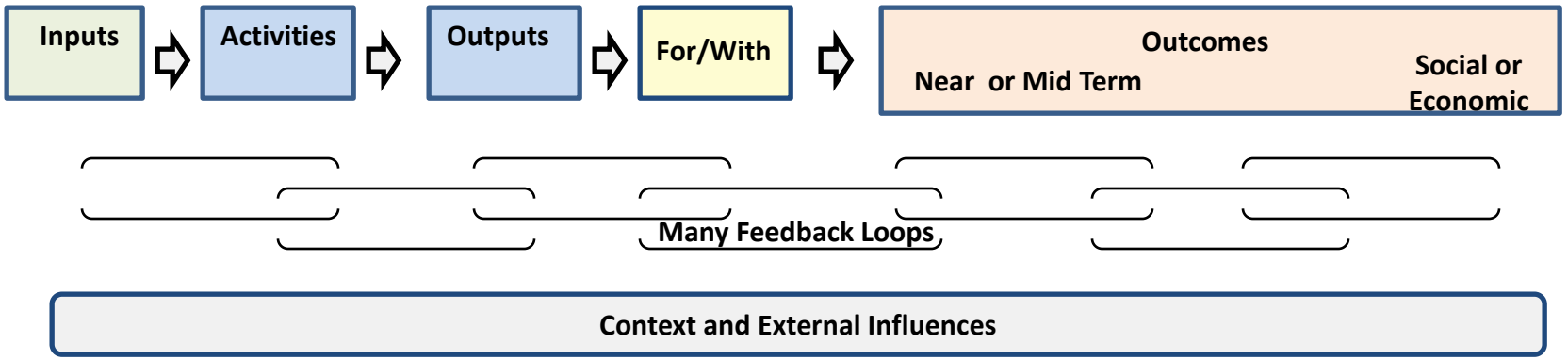
Recognize evaluation as a management tool to be used across the program life cycle

Stage in the Program Life Cycle	Question Simply Stated	Evaluation "Criteria"
Planning	What will the program do, when and why?	• Program implementation design • Evaluation plan exists
	Are we doing the right thing?	• Relevance
Early/Mid Implementation	Are we doing it the right way?	• Economy • Efficiency • Quality • Performance (early)
Mid/End of Implementation	What has been the outcome/impact?	• Effectiveness • Performance • Value For money
Learning/ Redesign	What do we do next?	• Use of evaluation findings

Use Different Types of Evaluations to Answer Different Questions

- Prospective outcome evaluation
- Monitoring outputs
- Process evaluation with short term outcomes
- Retrospective outcome evaluation

Plan Evaluations Around a Logical Framework

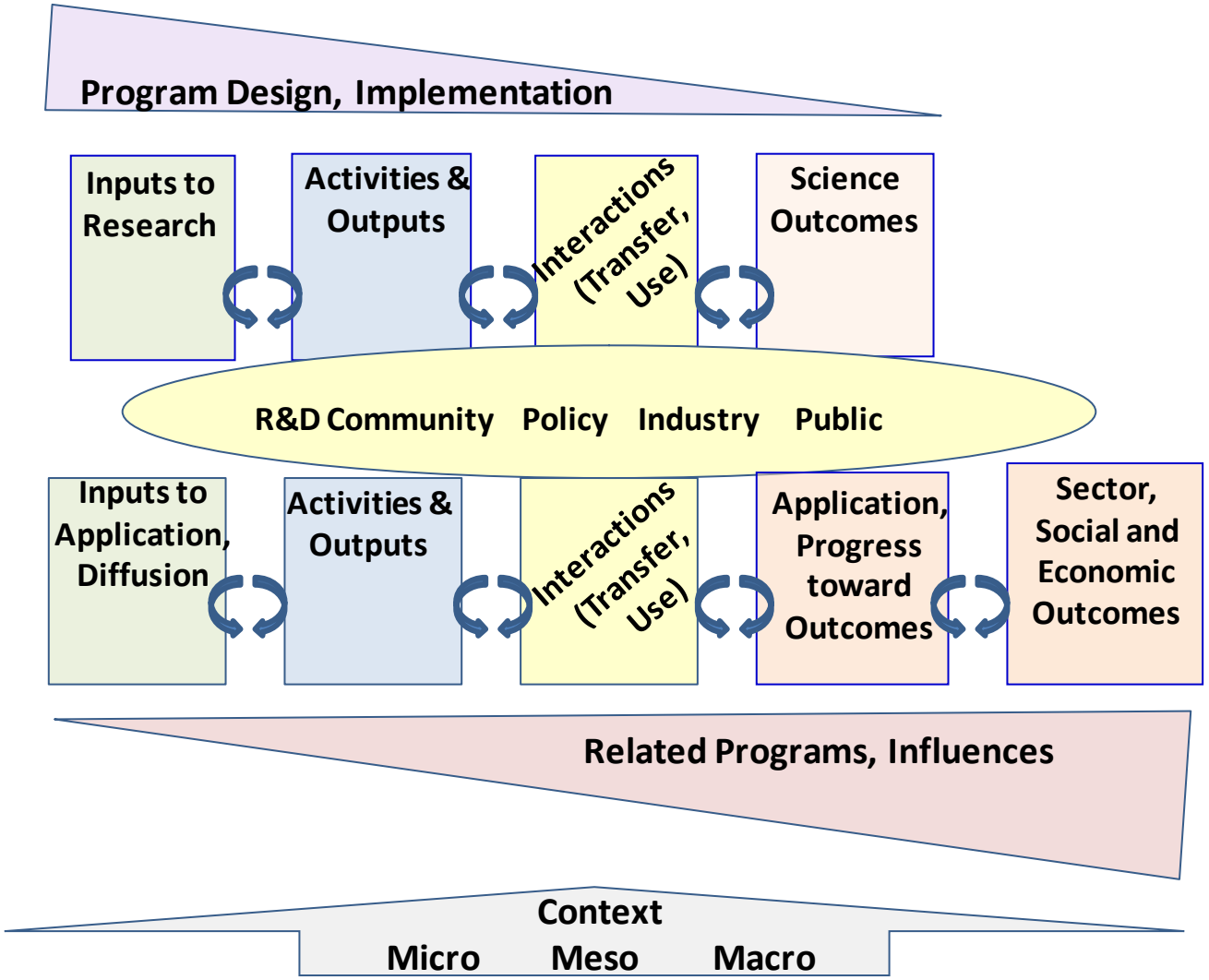


Indicators for						
Inputs	Activities	Outputs	Interactions	Near term Outcomes	Mid term Outcomes	Social or Economic Outcomes
Characteristics of likely differentiating factors; External influences on achievement						
Micro			Meso/Sector			Macro

A Proposed Generic Framework – With Context To Describe the Diversity in RTD Programs

- Separates science outcomes from application and end outcomes.
  - to distinguish science questions from impact and policy questions;
  - end outcomes of current work are not under the direct influence of the program;
  - important to measure dissemination and take up.
- Technology and development activities may or may not draw on science outcomes.
- For any new innovation there is an "application and progress" stage before end outcomes.
- Context must characterize 3 levels for systems evaluation – micro, meso (or sector) and macro.

A Proposed Generic Logic Model and Context To Outline the Diversity in RTD Programs



We Will Need a Framework of Frameworks to Describe Major Archetypes

A set of more detailed generic logic models and frameworks would help characterize

- Outcomes and pathways to outcomes for various sectors (e.g., health, energy)
- Pathways to outcomes for combinations of characteristics,
  - Type and context of research (e.g. applied research in area where RTD networks already exist), and
  - Context for adoption of new product (e.g., supportiveness of current technical, business and government infrastructure, consumer demand)
- Detail on commonly used mechanisms such as strategic clinical networks in health research, or Engineering Research Centers

A Menu of Indicators For the Generic Logic Model

- Each element of the logic model is described by the listing of indicators.
- This results in a menu of contextual indicators and many outcomes of RTD that can be measured, depending on
  - the type of RTD and its desired objectives,
  - target audiences for the application of the RTD, and
  - timing of the evaluation relative to the time passed since the activities took place.
- The list, while not comprehensive, reflects outcomes identified in numerous evaluation frameworks and literature reviews.

Table 2. Examples of Indicators and Outcomes Across the Scope of RTD Programs -1

Program Design, Implementation:

- Efficiency, effectiveness of planning, implementing, evaluating; Stakeholder involvement
- Robustness of program partnerships, other delivery infrastructure
- Progress in required areas (e.g., e-government)

Contextual Influences:

- Characteristics of researchers (team size, diversity)
- Nature of RTD problem (type, scope, radicalness)
- Characteristics of interactions (continuity, diversity, etc.)
- Nature of research application (breadth, depth, timing, radicalness of change; sector absorptive capacity)
- Characteristics of macro environment (availability of capital, capabilities; ease of coordination)

Inputs/Resources for Research:

- Expenditures on research
- Expenditures on research support activities, such as database development, research planning and priority setting
- Depth, breadth of knowledge base and skill set of researchers and technologists, teams, organizations
- Capabilities of research equipment, facilities, methods that are available
- Vitality of the research environment (management, organizational rules, etc.)

Table 2. Examples of Indicators and Outcomes Across the Scope of RTD Programs -2

Activities (the Research Process) and Outputs:

- Plan, select, fund, researchers, research projects, programs
- Quality, relevance, novelty, of selected researchers, projects, programs
- New knowledge advances (publications, patents, technical challenges overcome)
- Quality and volume of other outputs (grants made, projects completed, number of reports, people trained, etc.);

Interactions (Includes Transfer and Use):

- Research collaborations, partnerships formed; preparation for transition to application
- Dissemination, exchange of research outputs (publications, inclusion in curricula, etc.)
- Industry engagement, co-funding, follow on funding for the research
- Public engagement, awareness of outputs (participation, media mentions)

Science Near-Term Outcomes:

- Publication citations; patent applications, patents
- Awards, recognition, professional positions
- Expansion of Knowledge base in terms of technical leadership and absorptive capacity
- Advances in research/technical infrastructure (new research tools, scientific user facilities, testing facilities)
- People educated in RTD area and research methods
- Linkages/communities of practice/networks
- Technical base (technology standards, research tools, databases, models, generic technologies)
- Commercialization/utilization support base (manufacturing extension programs, supportive codes, etc.)

Table 2. Examples of Indicators and Outcomes Across the Scope of RTD Programs -3

More RTD or RTD Diffusion Activities, Outputs and Interactions:

- Public funds expended for these RTD or Diffusion programs ; Leveraged investments by private sector
- Translational or cross-functional teams; Presence of intermediary organizations
- Technical milestones achieved, prototypes built/scaled up, additions technical knowledge and infrastructure
- Dissemination, exchange of knowledge; consultation; citation
- Additions to diffusion/adoption infrastructure (capabilities, delivery, etc.)

Application of Research, Progress toward Outcomes:

- New technology development advances (movement through stages, functionality)
- Product commercialized; policy /practice implemented; attitude or behavior changed
- New "technology" commercialization/diffusion advances (supply chain develops, adoption of new process technology)

For each of the above:

- Utilization/influence, sustainability of influence on decisions, behavior, physical or financial factors

Sector, Social and Economic Outcomes/Impacts:

- Modeled monetized benefits
- Health status
- Security, safety measure
- Sustainability measure
- Income levels
- Jobs
- Benefit to cost ratio
- Quality of life
- Environmental quality
- Production levels
- Cost savings
- Competitiveness

Related Programs and Major Influencers:

- Date of formal handoffs to or take up from partners, others
- Chronological account of who else did what, when

Logical Framework Example: NSF Human and Social Dynamics Program

Research Program		For/ With	Near Term Results
Resources HSD projects funded	Activities & Outputs Research published. Interdisci-plinarity	Target Audience Non-HSD fields, collabor-ators	Science Outcomes Knowledge diffused, Influences other fields

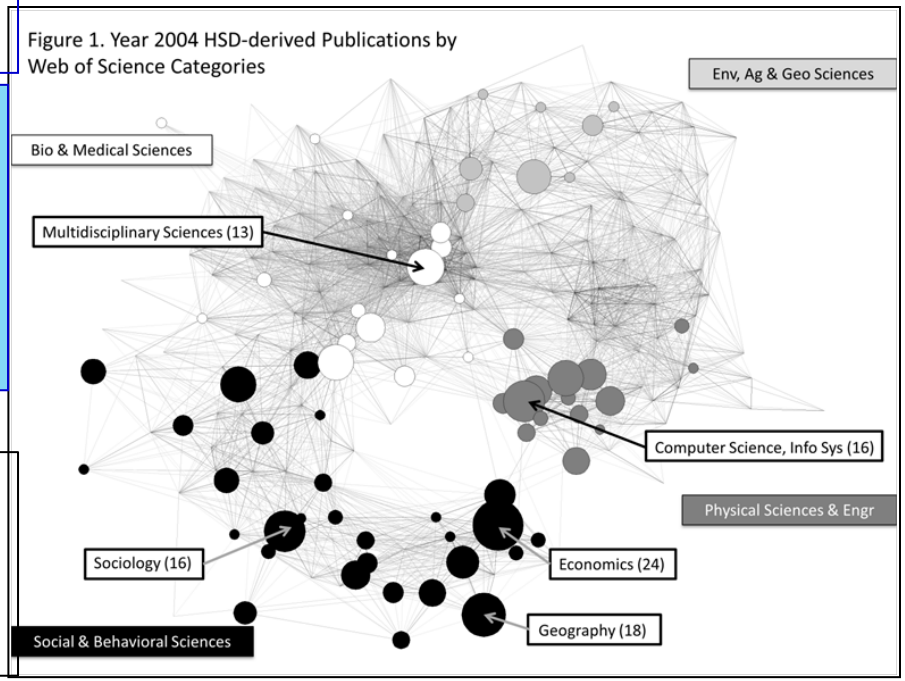
Indicators

Expendi-tures Topics funded	Publications -number -co-authorship	Integration scores Network characteris-tics	Publication maps Citations - #, distance -velocity

Design, Methods

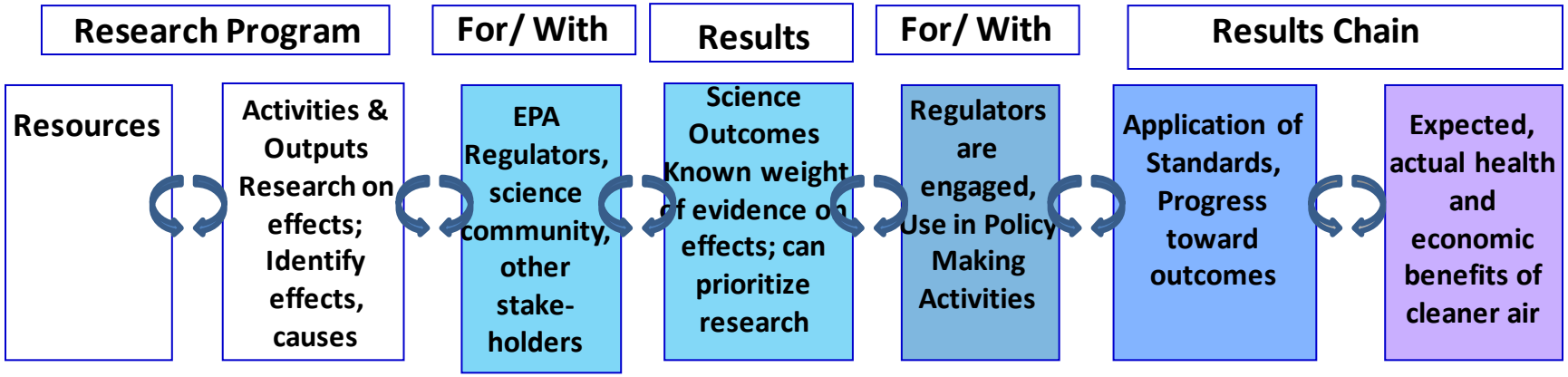
Control group
Bibliometrics, Network analysis
Visualization of diffusion patterns

Contextual Influences: Social/Cultural, Technical, Business/Economic, and Political/Legal)



Source: Garner J, Porter AL, Borrego M, Tran E, Teutonico R. (2013). *Research Evaluation*,22(2.

Logical Framework Example:
Research and Science Judgments That Inform Health Standards



Indicators

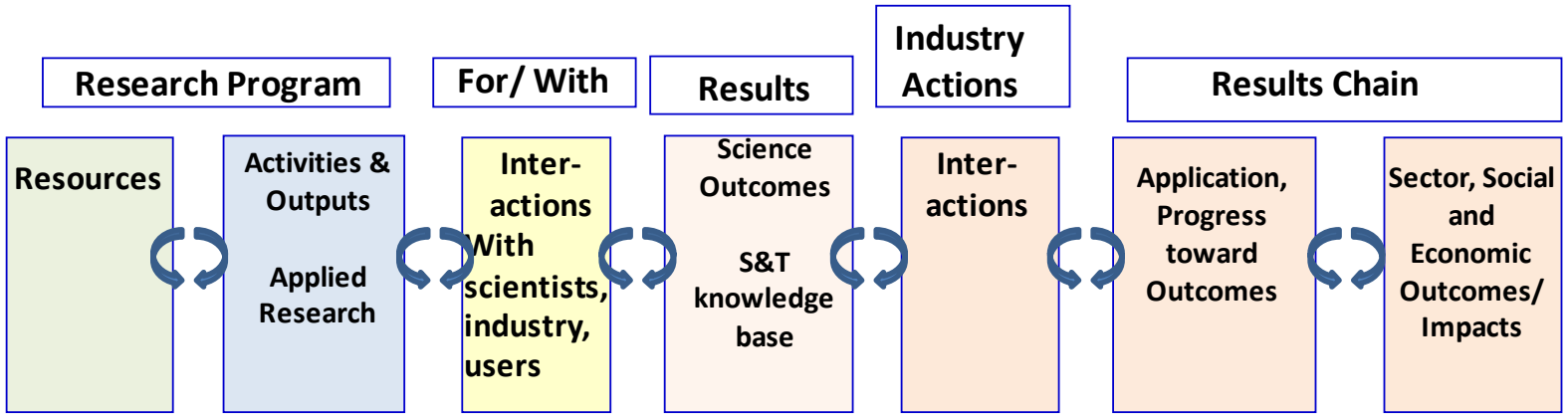
Funds Laboratories Scientists Expert Advice	Publications; Reduced scientific uncertainty; New methods, tools, models.	Influence on research agenda; Findings in HERO, used in ISA process	Citations; Experts view evidence provided as strong	Regulatory assessments; Health-based standards; Judicial decisions	Emissions reduced; Air quality improved; Human exposure reduced.	Reduced risk to human health; Benefits from reduced morbidity & mortality

Design, Methods

Mine knowledge database (Health & Environmental Research On-line (HERO))
Exposure and Risk assessments
Independent expert review of science

U.S. EPA. (2009). *Integrated Science Assessment for Particulate Matter (Final Report).*

Logical Framework Example: U.S. DOE Wind R&D Linkages with Commercial Wind Generation



Indicators

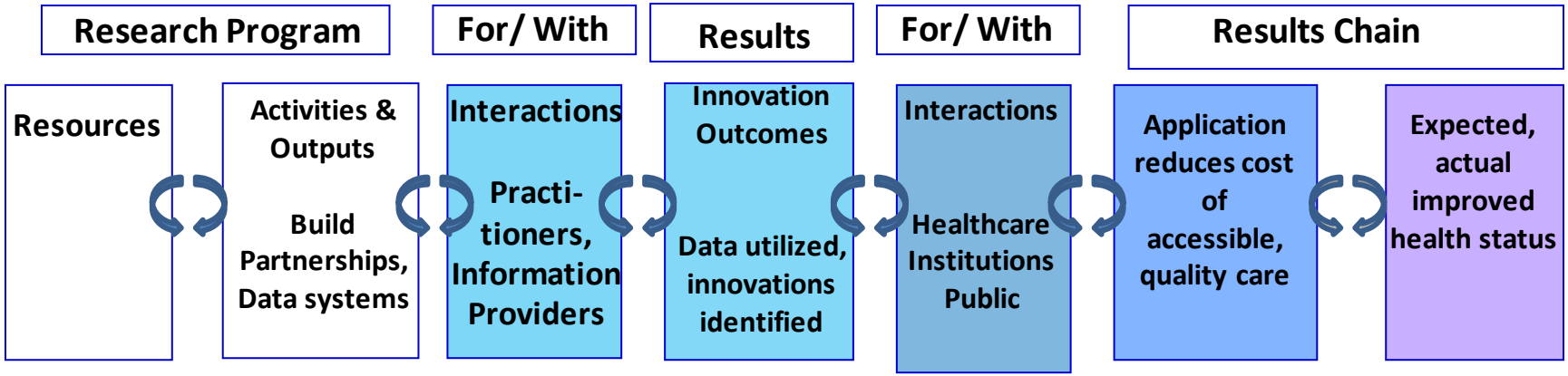
$ spent by year Topics funded	Publications, reports, patents. Prototypes of turbines, blades, control systems	Partnerships with universities, research labs, technology & engineering firms, utilities, user groups	Co-authorship and citation of publications. Patent tracing & citation of DOE research. Testimonials.	Commercialization of DOE prototype turbines. System Integration. Other new innovations.	Cost per MW reduced. Improved system reliability, durability. Market growth. Spillovers.	Change in wind power capacity (MW). Fossil fuel, pollution avoided.

Design, Methods

Historical Tracing study that included: Pre-post overview, publication and patent analysis; interviews with researchers, program managers, firms, and technology and market experts; network analysis.

Ruegg and Thomas, Linkages from DOE's Wind Energy Program, 2009

Logical Framework Example: Innovation in Healthcare Delivery to Reduce Costs



Indicators

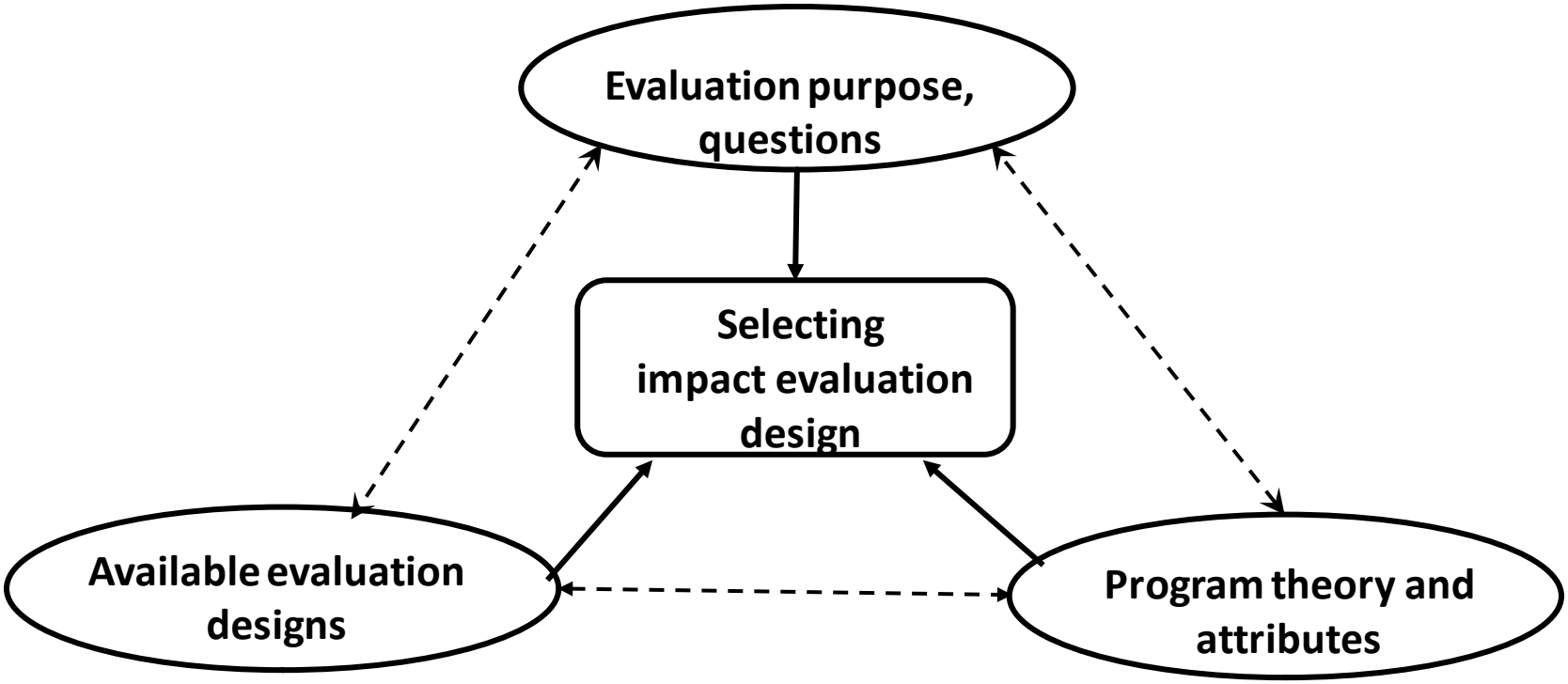
Funds Expertise Existing relationships	Partnerships Goal alignment Standard-setting Data base of cost, activities, outcomes	Number of practices, size of population involved Dissemination mechanisms	New innovations Randomized trials to validate innovations Public accesses, believes data	Practitioners, public become aware of results, change attitudes	Public demands, practitioners deliver more effective care	Benefits from reduced morbidity & mortality Reduced medical costs

Context: Related programs, Interest groups, Financial incentives, Determinants of health, Social, Economic, Political/Legal factors

Recommendations: Use of Appropriate Methods

- Clarify purpose and questions before deciding on a method
- Choosing a study design for outcome evaluation, attribution
- Consider Contribution Analysis
- Using mixed evaluation methods
- Valuing economic and other societal outcomes
- Evaluation synthesis and aggregation

Purpose-, Question- and Theory-Driven Design



Source: Adapted from Figure 6 in Impact evaluation of natural resource management research programs (Mayne and Stern, 2013)

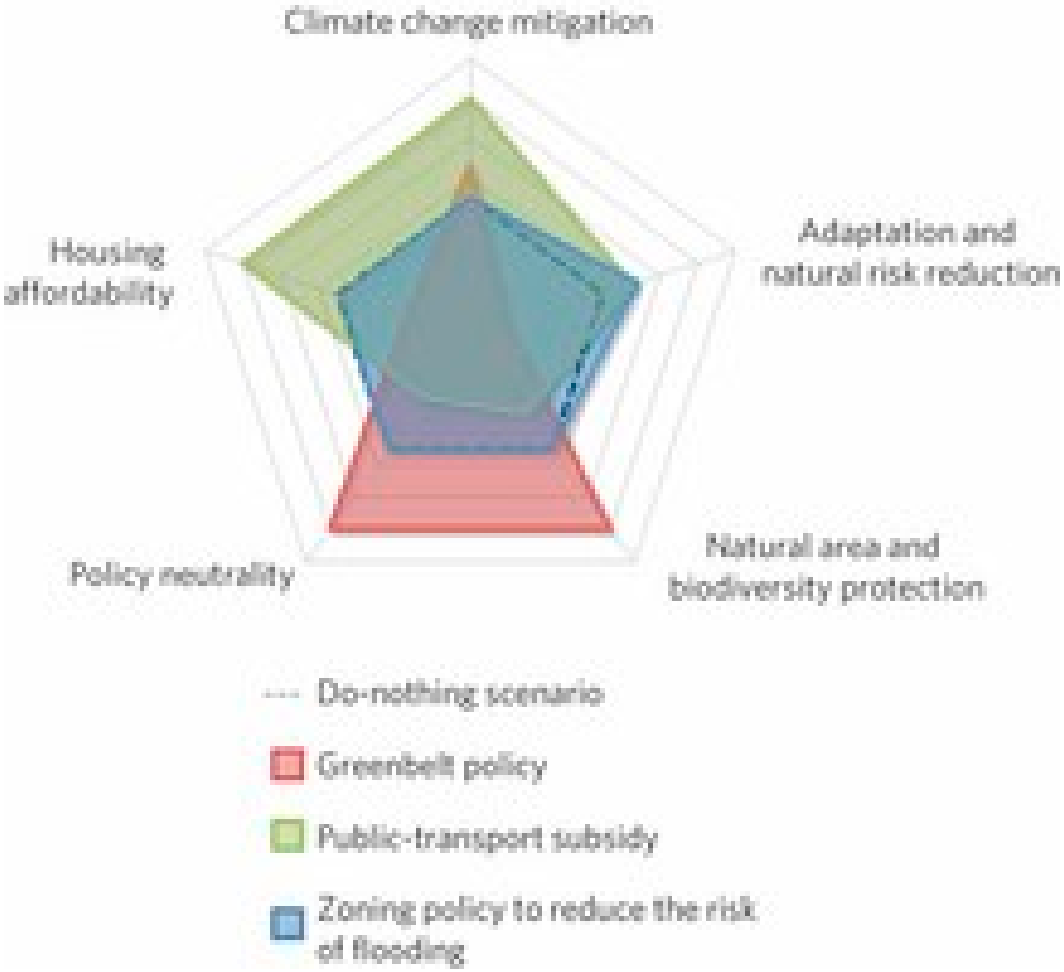
Attribution Using Frameworks and Context

- Three conditions required to establish cause and effect:
  - a logical explanation for why the investment can be expected to have led to the observed outcome.
  - a plausible time sequence of the investment occurred and the observed change relative to an appropriate baseline follows.
  - compelling evidence that the investment/actions are the partial or full cause of the change when competing explanations are taken into account.
- Reliable control groups in experimental or quasi-experimental study design is seldom possible for RTD. A sampling of participants and non-participants may not be truly random, groups not comparable.

Contribution Analysis – An Alternative

- A use of program theory
- Examines the role the program played (plays) in the larger system.
- Shares the credit
- Has the advantage of also informing next steps
- Being used more in Europe and Canada
- Specifically, Contribution Analysis examines *context, mechanisms,* and *outcomes* to see what worked under what circumstances (John Mayne, 2012)

Impacts Are More Than Economic



http://www.nature.com/nclimate/journal/v2/n5/full/nclimate1434.html?WT.ec_id=NCLIMATE-201205

Evaluation Synthesis

- Takes existing studies, and based on the quality of the study and strength of evidence, uses findings as a database of what is known at that time.
- Helps answer policy questions that no single study could answer because a single study cannot be large enough in scope.
- After conflicts in findings can be resolved, looking across studies points to
  - features of an intervention that matter most, that are not visible in a single study.
  - which may be background variables, or research design, or stability across groups.
- Can show where there are gaps in knowledge that call for further targeted evaluation studies or new policy experiments.

Source: U.S. Government Accountability Office (GAO) 1992, *The Evaluation Synthesis*, GA/PEMD-10.1.2, Washington, DC.

For Example, Standardized Case Studies

- Standardized case studies share a common framework and characterize key aspects of a program and its context, so study data can be aggregated and hypotheses tested with combined data (French National for Institute for Agronomic Research (INRA))
- Tools standard across the studies
  - Chronology: time frame, main events, turning points
  - Impact Pathway: productive intermediaries/interactions, contextual factors
  - Impact Vector: Radar chart of impact dimensions
- Identified
  - Production of actionable knowledge, Lag before impact
  - Program roles on two dimensions: Upstream or downstream and Exploring new options or insuring existing.

Joly, Pierre-Benoit, Laurence Colinet, Ariane Gaunand, Stéphane Lemarie, Phillipe Laredo, Mireille Matt, (2013). A return of experience from the ASIRPA (Socio-economic Analysis of Impacts of Public Agronomic Research) project. *www.fteval.at/upload/**Joly**_session_1.pdf and http://www6.inra.fr/asirpa_eng/ASIRPA-project.*

AEA RTD Group Recommendations

Recommendation #1: Build into each new program and major policy initiative an appropriate evaluation framework to guide the program or initiative throughout its life.

- Evaluation should be undertaken because evaluation is a valuable management tool at all stages of the program life cycle;
- Evaluations should be planned using a logical framework that reflects the nature of RTD in a meaningful way; and
- Decision makers' questions may call for both retrospective and prospective evaluation, and for evaluation of outputs and early outcomes that are linked to longer term outcomes.

Recommendation #2: More needs to be done to develop appropriate methods for designing programs and policies, improving programs, and assessing program effectiveness.

- More can be done to use or insist on the use of the robust set of methods that exists for evaluating RTD outcomes;
- Evaluation methods for demonstrating program outcomes should be chosen based upon the specific questions being answered and the context;
- Mixed methods are usually best, especially when outcomes of interest go beyond knowledge advance to include social or economic outcomes, where neither expert judgment nor bibliometrics are sufficient; and
- There are options for assessing attribution, although it is recognized that experimental design is seldom an option and contribution to a causal package is more useful.

Recommendation #3: The RTD community should move toward the utilization of agreed upon evaluation frameworks tailored to the RTD program type and context in order to learn from synthesis of findings across evaluations.

- There needs to be continued movement toward a common language and common evaluation frameworks by type of RTD program and context, with common questions, outcomes, indicators, and characterization of context; and
- Methods need to be further developed and used in relation to evaluation synthesis and the research designs and data collection and analysis that support it.

Summary, Next Steps

- The objective of the AEA RTD interest group is to provide a document with which to engage RTD evaluators, program managers, and policy makers in a dialogue about a current RTD evaluation language and practice.
- The end goal is consensus on a common RTD evaluation language and practice that is then broadly implemented.
- The paper is a DRAFT Final. Barring major concerns, it will stand as a RTD TIG paper, Version 1.
- We will post the paper on the TIG website, put it under a Creative Commons license (share with attribution), and welcome suggestions and additions for Version 2.

Acknowledgement
Volunteers from the RTD TIG

Team Leaders

Gretchen Jordan	360 Innovation LLC
Dale Pahl	US EPA
Liza Chan	Alberta Innovates - Health Solutions, Canada
Kathryn Graham	Alberta Innovates - Health Solutions, Canada
Deanne Langlois-Klassen	Alberta Innovates - Health Solutions, Canada
Liudmila Mikhailova	CRDF Global
Juan Rogers	Georgia Tech
Rosalie Ruegg	TIA Consulting Inc.
Josh Schnell	Thomson Reuters
Robin Wagner	US National Institutes of Health
Madeleine Wallace	Windrose Vision LLC
Brian Zuckerman	IDA Science and Technology Policy Institute

Acknowledgement
Reviewers providing written comments

- Erik Arnold, Technopolis Group
- Frederic Bertrand, Independent Consultant
- Mark Boroush, U.S. National Science Foundation
- Irwin Feller, Pennsylvania State University
- Laura Hillier, Canada Foundation for Innovation
- Jean King, University of Minnesota
- Jordi Molas-Gallart, Spanish Council for Scientific Research
- Lee Kruszewski, Alberta Research and Innovation Authority
- Al Link, University of North Carolina at Greensboro
- Steve Montague, Performance Management Network
- Cheryl Oros, Independent Consultant and Liaison, AEA EPTF
- Richard Riopelle, Ontario Neurotrauma
- Anita Schill, National Institute for Occupational Safety and Health
- Bill Valdez, U.S. Department of Energy

Comments on Version 1 are welcome through November 1, 2014.

Send these to

Gretchen Jordan
gretchen.jordan@comcast.net