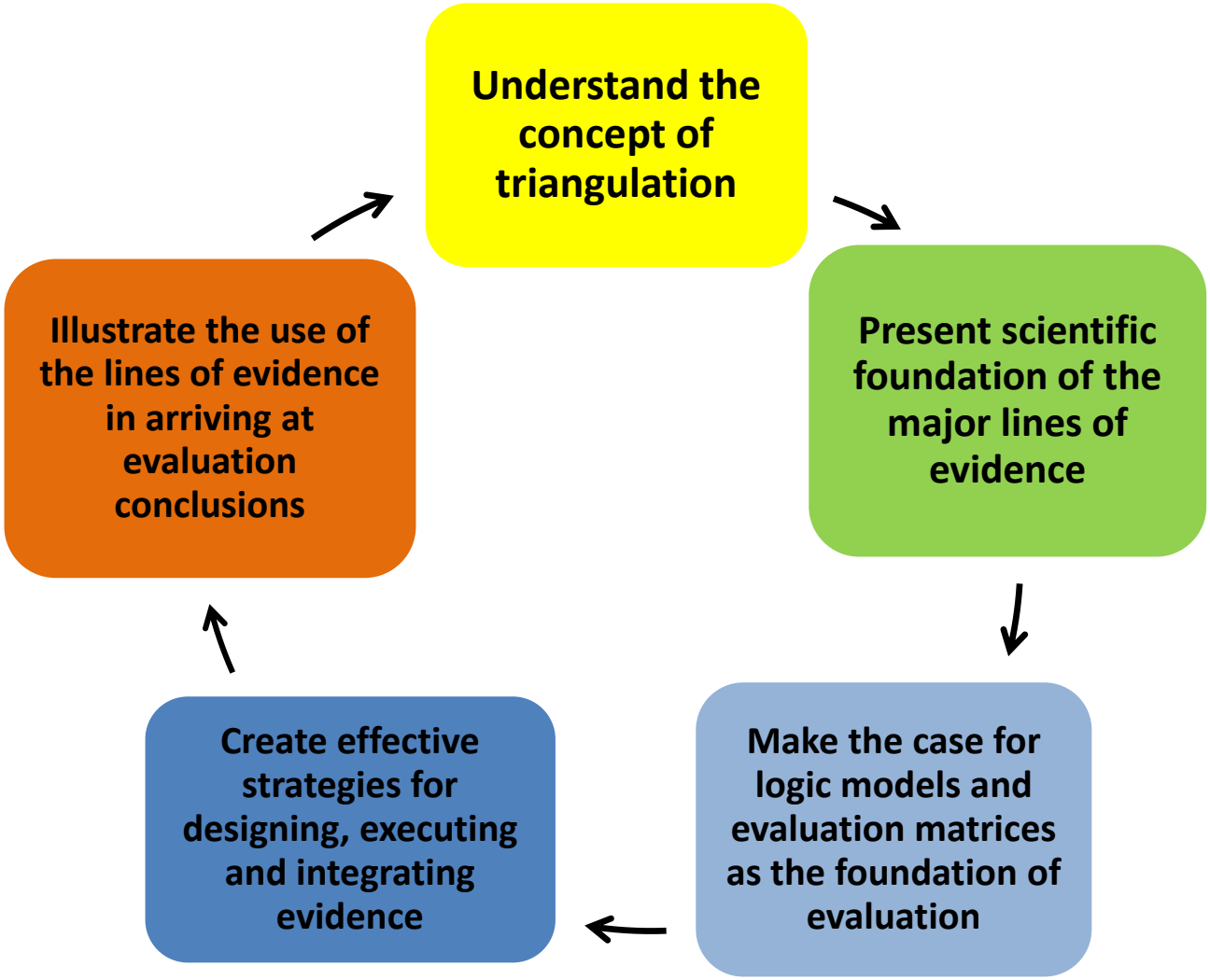
# Mixed Methods: Combining Quantitative And Qualitative Data

Greg Mason
PRA Inc. & University of Manitoba

Ottawa, November 15, 2013

# Learning Objectives



# Outline of the day

1. **Foundations**
   a. Principles of mixed-methods evaluation research
   b. Overview of quantitative and qualitative research
   c. Why quantitative and qualitative data cannot be combined… only integrated
   *d. Case example 1 – First summative evaluation of the NCB*
2. **Assess the information content of common lines of evidence**
   a. Myths and mistakes underlying the use of triangulation and multiple methods
   b. Evaluation methods comprise analytical framework, data collection, data analysis, and interpretation
   *c. Case example 2 – Evaluation of the federal response to the BSE crisis*
3. **Analyse the information requirements of the evaluation**
   a. Value for money and the TB issues/questions are the starting point for any evaluation and the choice of methods
   b. The evaluation matrix is the integrative platform for the lines of evidence
   *c. Case example 3 – The Farm Improvement and Marketing Cooperative Loan Act*
4. **Align the evidence to the evaluation matrix**
   a. Match the method to the question/indicator
   b. Reporting mixed methods research
   *c. Case example 4 – Horizontal Summative Evaluation of the Government of Canada's Investment in the 2010 Olympic and Paralympic Winter Games*

# Part 1 - Foundations

# Social research draws on two broad methods

**Positivism** and **Interpretivism** (post-modernism) are the predominant social science paradigms in North America.

- **Positivism**
    - **Induction** — generalizations (hypothesis formation) from observation
    - **Deduction** — testing hypotheses by data organization, experimentation and prediction
- **Interpretisvism (post-modernism)** (proposes and tests causal relationships using social scientific processes based on everyday existence

# Principles of mixed methods evaluation research (1)*

- Evaluation research answers three key questions ***what, why, how***
  - **"Why"** is about understanding
  - **"What"** deals with the impact ("who" and "when" are subsets)
  - **"How"** describes the intervention
- **"What" and "how"** also generate hypotheses about existence and process (causal relations)
- **Philosophy of research**
  - **Ontology** generates alternative theories of existence
  - **Epistemology** presents alternative ways of knowing and explaining.

* Adapted from Blaikie, 2010

# Principles of evaluation research (2)

- Evaluation research starts as a qualitative exercise - the specification of evaluation issues/questions or the "matrix".
- Quantitative and qualitative data are rooted in specific ontological and epistemological frameworks
- No necessary connection among the philosophical foundations of each method
- Each type of evidence rests on a unique and specific ontological and epistemological foundation
- Different methods of data collection (lines of evidence) can be combined algebraically ***only*** when they share the same ontological and epistemological foundation.
- However, different methods of data collection can be used sequentially building toward increased understanding and insight

# Data collection methods (Lines of evidence)

**Quantitative**

- Questionnaires (large sample, self-administered)
- Structured interview ( large sample, fixed response with interviewer)
- Observations (Structured)
- Content analysis of documents
- Administrative files (program activity, client activity, output/outcome counts, financial data)

**Qualitative**

- Participant observation (observer takes unstructured notes)
- Observation (semi-structured and unstructured)
- Oral history
- Content analysis of documents

**In general, qualitative research is more labour intensive (costly) than quantitative research.**

Key informant interviews are not a line of evidence

A key informant is a line of insight

# Quantitative Research (1)

## Uniform Units of Analysis

- **Unit of analysis** aligned to the 1) program target focus
  - Individuals
  - Families/households
  - Firms
  - Organizations
  - …
- **Unit of analysis** aligned to 2) the program delivery focus
  - Managers
  - Organizations
  - …

**Key idea: Quantitative methods rely on "counting" similar units**

"use of standardised measures so that the varying perspectives and experiences of people can be fit into a limited number of predetermined response categories to which numbers are assigned" (Patton, 2001, p.14).

# Quantitative research (1)

- Emphasize facts and causal relations between facts
- Large sample survey and administrative data sets dominate
- Inferences from a sample to population mandate probability sampling
- Information can be classified and grouped into standardized categories using statistical analysis with sufficient cases
- Key quality attributes

**Reliability**

a. The stability of a measurement over time
b. Control of intervening factors and concepts of "stability" are important ideas

**Validity**

a. Often defined as "construct validity". The construct is the initial concept, notion, question, or hypothesis that determines which data is to be gathered and how it is to be gathered.
b. A key challenge is that researchers may alter the construct in the face of disconfirming data.

# Validity and reliability

- The commonly stated goal is to reduce bias and increase reliability
  - **Bias** is the difference between what is measured/observed and what is true
  - **Reliability** is generally defined as consistency in measurement


http://explorable.com/


87% OF THE 56% WHO COMPLETED MORE THAN 23% OF THE SURVEY THOUGHT IT WAS A WASTE OF TIME

# Qualitative Research

- Data that cannot be counted and processed statistically
- Common manifestations in evaluation are
  - Interviews
  - Focus groups
  - Case studies
- Two core challenges
  - is selecting subjects (as opposed to sampling) for their information value.
  - Managing the tension between researcher as actor and researcher as observer

"the researcher is the instrument" (Patton, 2001, p. 14).

- Key quality attribute are subject to some controversy

**Reliability**

a. Concept of trustworthiness is core for some researchers
b. Others maintain that reliability is a construct that pertains only the quantitative studies.
c. Finally others argue that validity dominates the measure of quality

**Validity**

a. Not an absolute, but based on the theoretical framework and data collection/analysis process.
b. Many researchers stress rigour in the process as the guarantor of validity

# Data Reduction

## Quantitative

- Coding (pre-coding – post coding)
- Scales/indexes (Likert, magnitude)
- Factor/cluster analysis to refine constructs

## Qualitative

- Coding (classification)
- Thematic development (detect story lines)
- Typology/metaphor development (analogies)

Quantitative and qualitative data **both** engage in manipulation/processing

# Data Analysis

**Quantitative methods test hypotheses**

- Univariate/bivariate
- Multivariate (regression and other linear models)
- Instrumental Variables
- Structural equation
- Quasi-experiments

**Qualitative**

- Description – no inference on cause
  - Summary (Synopsis)
  - Thick (Extended story)
- Grounded theory
  - Coded concepts supports
  - Analytic induction to form a conclusions.

# Domains for lines of evidence

**Quantitative research focuses on**

- Measuring concepts (income inequality, cost-effectiveness, etc.)
- Establishing likely causality
- Generalizing from a sample to population
- Replicating and aggregating using standardized methods based on….
- ….discrete and uniform units of analysis

**Qualitative research focuses on**

- Explicating concepts and theories
- Supports insight and hypothesizing
- The actors' points of view
- Thick description
- Social processes

## Qualitative versus quantitative

Eyewitness account of your birth

Your birthdate

# Triangulation - Origins

- Social scientists in the sixties became concerned that single methods (interviews/questionnaires) were inherently biased.
- Corroborative evidence was advocated to increase validity

> "When a hypothesis can survive the confrontation of a series of complementary methods of testing, it contains a degree of validity unattainable by one tested within the more constricted framework of a single method" (Webb *et al* 1966: 174).

> "No single method is always superior. Each has its own special strengths and weaknesses. It is time for sociologists to recognise this fact and to move on to a position that permits them to approach their problems with all relevant and appropriate methods, to the strategy of methodological triangulation." (Denzin, 1970b: 471).

# Rationale for triangulation

- Many use the analogy from land surveying to justify triangulation evaluation

"a single landmark can only provide the information that they are situated somewhere along a line in a particular direction from the landmark. With two landmarks, however, their exact position can be pin-pointed by taking bearings on both landmarks; they are at the point where the two lines cross.

In social research, if one relies on a single piece of data there is the danger that undetected error in the data-production process may render the analysis incorrect… diverse kinds of data *(that)* lead to the same conclusion, one can be a little more confident in that conclusion… *(because)* different kinds of data have different types of error built into them" (Hammersley and Atkinson,1983: 198).

# *Case example 1 – First summative evaluation of the NCB*

**Or how qualitative data supported a reinterpretation of puzzling quantitative results**

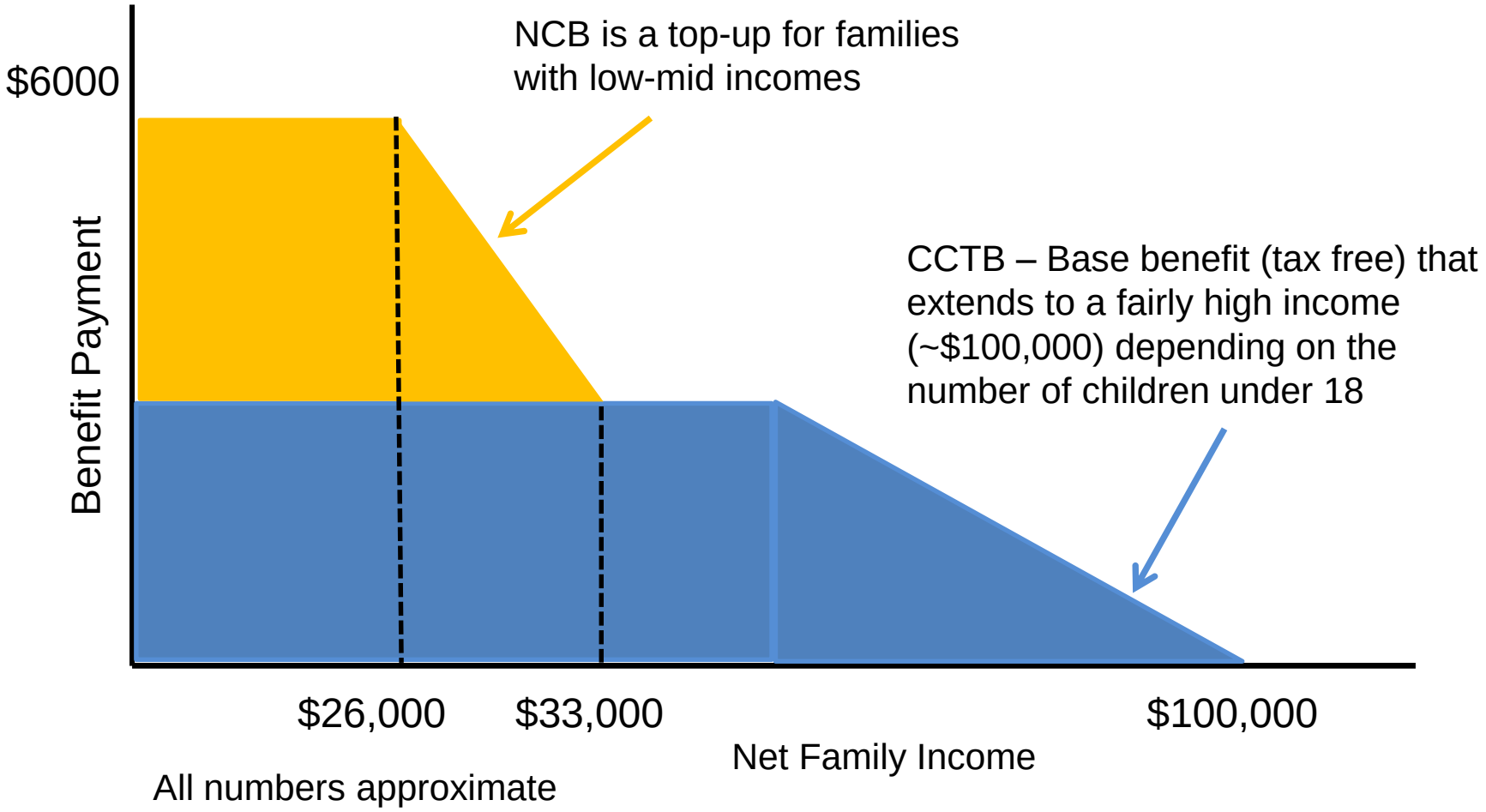
**National Child Benefit (NCB)**

The NCB Initiative is a joint initiative of federal and provincial/territorial governments intended to help prevent and reduce the depth of child poverty, as well as promote attachment to the workforce by ensuring that families will always be better off as a result of working.

It does this through a cash benefit paid to low income families with children, a social assistance offset, and various supplementary programs (childcare, additional cash benefits, employment support, health care, etc.) provided by provinces and territories.

# National Child Benefit

(two children < 18)



# Main methods – Information Role

- **Literature review**
  - Focus only on key theoretical foundations of the program to create the theory of change
  - Review similar programs elsewhere (EITC in US) to identify potential outcomes and causal inference
  - Identify work response to cash benefit programs to frame expected results
  - Differential response of low income families that did/did not receive social assistance (cash and in-kind)
- **Interviews** with FPT managers
  - Orient client survey
- **Client survey** (n=5000+) using sample from CRA tax records
  - Discover role of NCB in household budgets
- **Statistical analysis** of work response comparing low income families with children (program) and without children (control)
  - Use client survey data joined with tax data
  - Test whether receipt of the NCB reduced the incidence/depth of poverty; test whether NCB recipient parents changed work attachment
- **Focus groups**
  - Just because we thought it was a good idea… these turned out to be instrumental.

# *Key Interaction of quantitative and qualitative methods*

**Quantitative Analysis**

- On average the NCB did not significantly reduce poverty.

**Qualitative Analysis**

- Some parents expressed a strong preference for parenting their children as opposed to working for low wages and using child care.

**Quantitative re-Analysis**

- Isolating recipients with an another income earner in the household, and/or child with specials needs, and stated preference for parenting confirmed the reduction in work effort that negated the goal of poverty reduction.

**Qualitative analysis supported important re-interpretations of the work response and led to a deeper understanding of why the NCB seemed to have a limited impact on poverty.**

# Part 2 - Assessing the information content/potential of common evaluation lines of evidence

# Review of the common lines of evidence used in federal evaluations

| |
|---|
| Documents |
| Administrative data |
| Literature/expert review |
| Sample surveys |
| Key informant interviews |
| Focus groups |
| Case studies |

## Typical documents – information potential

| Document type | Information content – potential role in the evaluation *[1 = qualitative 2 = quantitative]* |
|---|---|
| Foundation documents (TB Subs, MCs, policy background) | • Program rationale and relevance *[1]*<br>• Program origins *[1]*<br>• Authority (financial, governance) *[1]*<br>• Outcomes [1,2]<br>• Targets *[1]* |
| Performance reports (G&Cs) | • Outputs and outcomes *[1]*<br>• "Thick" descriptions (implementation, outputs, outcomes) *[1]* |
| Audits and evaluations | • Program history *[1]*<br>• Benchmark for costs, implementation outputs, outcomes *[2]* |
| Applications (G&Cs) | • Applicant/client attributes *[1,2]* |

## Typical administrative files – information potential

| Document type | Information content – potential role in the evaluation *[1 = qualitative; 2 = quantitative]* |
|---|---|
| Management files (meeting minutes, HR records, etc.) | • Number *[2]* and type of employee *[1, 2]*<br>• Minutes of meetings to<br>– describe implementation *[1 and 2]*<br>– participation of partners: number *[2]* and type *[1]*<br>– implementation timing *[2]* and processes *[1]* |
| Financial records | • Payments (individual and aggregate) *[2]*<br>• Distribution and fairness *[2]*<br>• Payment timing and delay *[2]* |
| Client services | • Services delivered *[2]*<br>• Participation in program *[2]*<br>• Sample frame to support survey *[2]* |

| Typical literature reviews – information potential | |
|---|---|
| Review type | Information content – potential role in the evaluation [1 = qualitative 2 = quantitative] |
| Scan | • Top line summary [1]<br>• Program context [1]<br>• Implementation context [1] |
| Integrated review | • Program "arc" (history, context, and evolution) [1]<br>• Theory of change [1] |
| Integrated literature and expert review | • Program "arc" (history and evolution) [1]<br>• Theory of change [1]<br>• Theory of context and program evolution [1] |
| | |

| Typical large probability sample surveys reviews – information potential | |
|---|---|
| Survey type | Information content – potential role in the evaluation [1 = qualitative 2 = quantitative] |
| Interviewer mediated | • Respondent self-report<br>– Fixed response – number/category [1 and 2]<br>– Verbatim [1]<br>• Interviewer probes [1 and 2]<br>• Interviewer-respondent interaction creates a complex qualitative data field [1]<br>• Potential to increase reliability and validity [1 leads to insight on quantitative results**, if data are presented to interviewees**<br>• And decrease reliability and validity **[interviewer knowledge and skill paramount]** |
| Self-completed | • Respondent self-report<br>– Fixed response – number/category [1 and 2]<br>– Verbatim [1 → 2 on coding] |

# *Yes Prime Minister*
# Sir Humphrey teaches questionnaire design

Classic British TV comedy *Yes Prime Minister* has important lessons for those who want to interpret questionnaire data.

This clip shows two civil servants discussing a policy suggestion. Bernard Woolley, who we see first, thinks the public are in favour of the policy – the minister has had an opinion poll done. Senior civil servant, Sir Humphrey Appleby sets him straight.

Fans of cognitive biases, note that Sir Humphrey uses at least three in his illustration of a biased questionnaire: framing, priming, and acquiescence bias.

This example is exaggerated, but the moral still holds: questionnaires can be designed to encourage the answers you want. People's opinions are not objective facts like their height and weight. They change, depending on the context and on how they are asked.

Yes-Minister-SUrvey.wmv

# Typical key informant – information potential

| Interview Subject | Information content – potential role in the evaluation[1 = qualitative 2 = quantitative] |
|---|---|
| Expert | • Theory of change[1→2 surveys]<br>• Program antecedents [1]<br>• History of and projected need for intervention [1]<br>• Unique role for government vs other delivery options |
| Senior Manager | • **Program** origins and implementation [1]<br>• Strategic management (program) issues (e.g., FPT relationships) [1]<br>• Resource allocation (macro) [1]<br>• Expected/actual results (macro) [1]<br>• Alternatives (strategic/global) |
| Line Manager | • **Project(*s*)**origins and implementation [1]<br>• Local management (project(*s*) issues (e.g., community/organizational relationships) [1]<br>• Resource allocation at regional level (micro) [1]<br>• Expected/actual results at regional (micro) [1]<br>• Alternatives (program delivery) |
| Project Proponents (G&Cs) | • **Project** origins and implementation [1]<br>• Local management (project) issues (e.g., community/organizational relationships) [1]<br>• Resource allocation at project level (micro) [1]<br>• Expected/actual results and project (micro) [1]<br>• Alternatives (project implementation/delivery) |

## Typical focus groups – information potential

| Group type | Information content – potential role in the evaluation [1 = qualitative 2 = quantitative] |
|---|---|
| Client | • Program implementation [1]<br>• Program impact [1]<br>• Field experiment [2]* |
| Management | • Program implementation [1]<br>• Program impact [1] |
| * Certain quantitative methods are ideally implemented in a small group setting. Conjoint analysis applied to program/policy design is an example that should be more widely used | |

Focus groups are often seen as supplementary evidence designed to gather context about program implementation and impact, as well as ideas for program revision

The interaction among the participants means that the information whole is greater than the sum of the information parts.

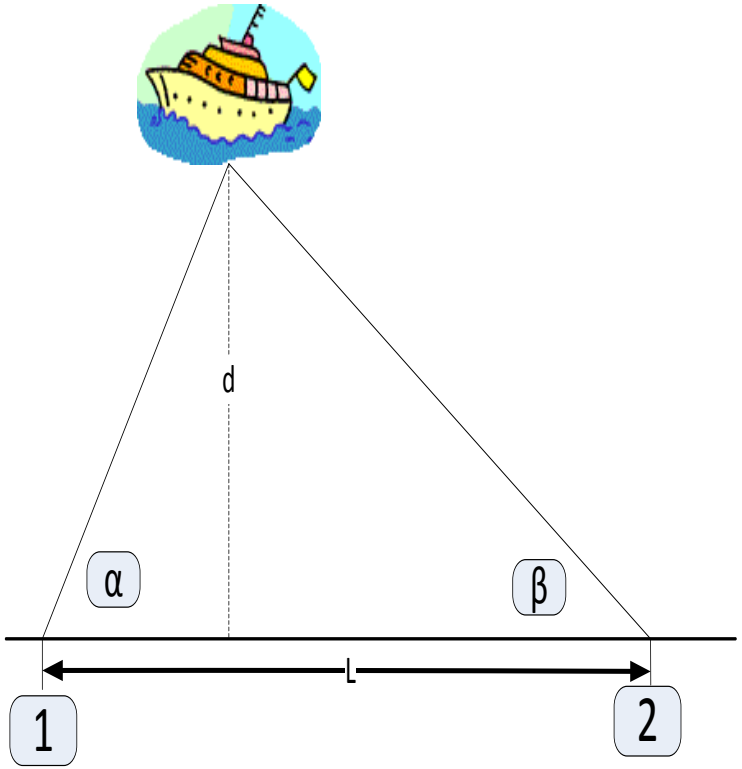
| Typical case studies – information potential | |
|---|---|
| Case study type | Information content – potential role in the evaluation [1 = qualitative 2 = quantitative] |
| Maximum variation | • Identify key patterns and variation (needs relatively large number.( >10) |
| Typical case | • Identifies the norm |
| Extreme (successes) | • Best practices (feel good) |
| Extreme (failures) | • Corrective |
| Politically critical | • Gain wanted positive or suppress unwanted negative attention |
| Convenience | • Low cost – low information |

# Sample surveys represent a blend of qualitative and quantitative methods

1. Design phase
   - Literature
   - Standard scales
   - Prior surveys
   - Expert interviews
   - Focus groups
2. Pretest
   - Expert assessment (detect hesitancy)
   - Active probes to secure meaning
   - Follow-up debrief
3. Data collection
   - Verbatim entry
4. Analysis
   - Coding and categorization
5. Reaction to results by experts and KIs


*"It's my fervent hope, Fernbaugh, that these are meaningless statistics."*

# Triangulation – one more time




Two observers can "triangulate" the location of the boat (distance from the short) by measuring the angles α and β and using the distance L and the law of sines.

Observer 1 and 2 measure the angles and the length L

**A man with one watch always knows the time.**

**A man with two watches is never sure.**

The key is that both observers are using the same theoretical framework (plane trigonometry

# Problems with triangulation

- Does not necessarily increase validity – competing perspectives fail to converge or collectively converge on a wrong idea
- May offer differing perspectives, but in social science this may not lead to less bias
- Mixing of quantitative and qualitative methods that draw from different theoretical frameworks usually results in the quantitative data dominating
- The analogy with surveying presents serious theoretical problems for mixing quantitative and qualitative methods

Using the problem to determine the distance of the boat from the shore, imagine observer 1 gave the angle as 23$^{o}$ while observer 2 gave the angle as "somewhat acute. " Further the helper hired to measure L decided to count the number of paces and not a standardized unit.

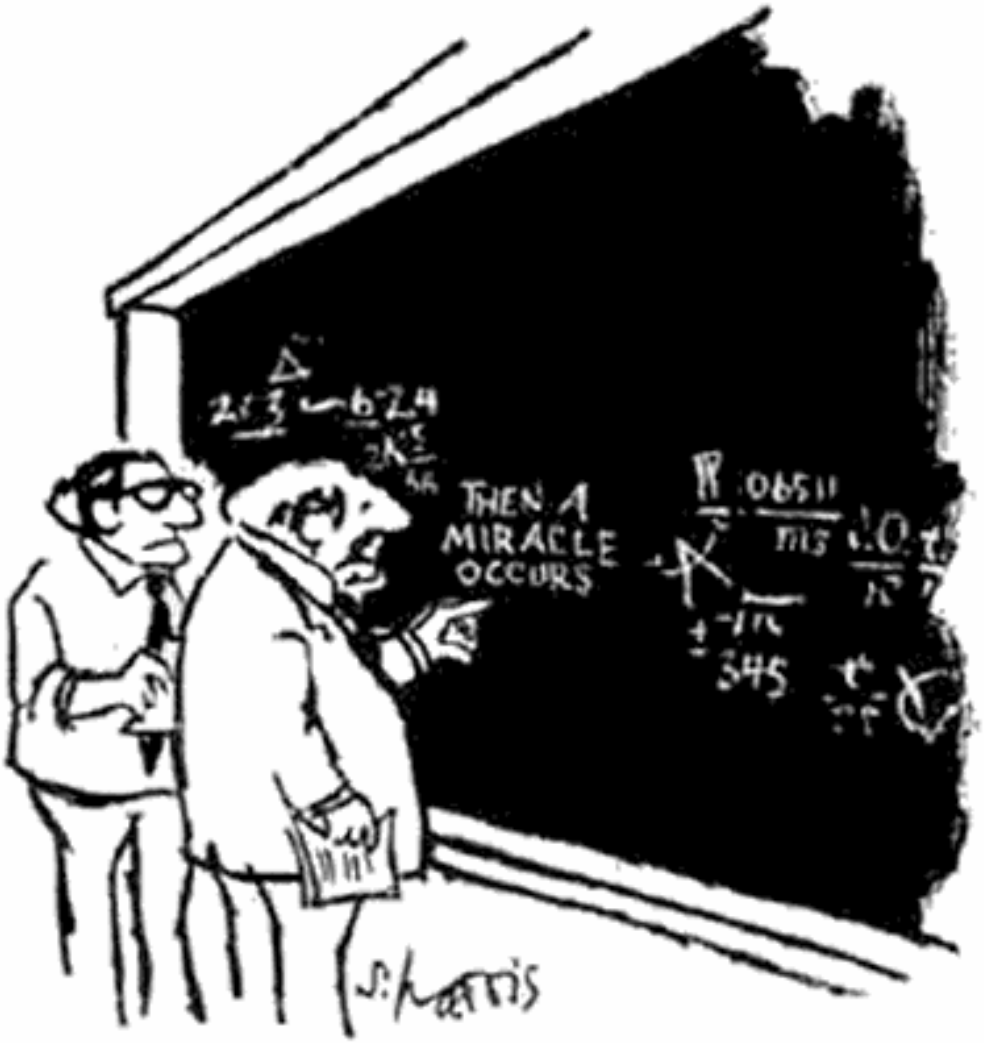
Triangulation has come to mean different things, depending on the ontological and epistemological framework of the researcher

# Multiple Triangulation

- **Data sources** – multiple sources (diverse key informants)
- **Investigators** – multiple interviewers
- **Multiple theoretical perspectives** – more than one theory of change
- **Multiple methodologies** – quantitative diversity and qualitative diversity

***Within method*** (strategies within a method) vs. ***across method*** triangulation (dissimilar methods to assess the same unit)

"the flaws of one method are often the strengths of another, and by combining methods, observers can achieve the best of each, while overcoming their unique deficiencies" (Denzin, 1970a: 308)."




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

# *Case example 2 – Evaluation of the federal response to the BSE crisis*

**The metaphor is the message**

- One Canadian cow tested positive for BSE on May 20, 2003
- Borders closed to more than 40 countries, including the US
- Exports of live cattle ceased, slaughter rates decreased, producers delayed marketing cattle
- Producer concerns about affording to feed animals and having to dispose of animals themselves
- In response, governments introduced programs to sustain and reposition the industry

Programs comprised financial compensation to offset losses by beef producers

- Special BSE programs (Federal)
  - BSE Recovery Program
    - Slaughter Element
    - Inventory and Pricing Element
  - Cull Animal Program
  - Fed Cattle Set-Aside
  - Feeder Calf Set-Aside
- Existing farm safety net

**The goal of the programs were to avoid a mass cull and to stabilize finances of primary producer and secondary processors.**
**Avoiding the UK experience was important**

# Main Methods
## *(Information Content; 1 = Qualitative 2 = Quantitative)*

- **Literature review *(2)***
  - *BSE Experience and programming elsewhere*
  - *The potential for BSE to affect human health created strong justification for action*
  - *The mass cull in the UK could not be repeated in Canada*
- **Document and file review *(2)***
  - *Origins of the program - Response of certain provincial governments (Alberta) impelled a federal response*
- **Market and economic analysis (simulation on Farm Model) *(1)***
  - *Simulated the financial and economic impact of the program on the livestock sector*
  - *Analysis of industry market data*
- **Analysis of program financial data *(1)***
  - *Financial payments to sector and individual producers/processors (fairness of support)*
- **Producer survey (n=1200) *(2)***
  - *Summary statistics on impacts across sector*
  - *Satisfaction with the program*
- **Interviews (n = 15 mangers FPT) (1)**
  - Understanding of program operation s
- **Case studies (n=6) *(1 and 2)***
  - Insight into impact of crisis at the farm level
  - Varied type and size (cow-calf, backgrounder, dairy, genetics)

# *Interaction of quantitative and qualitative methods*

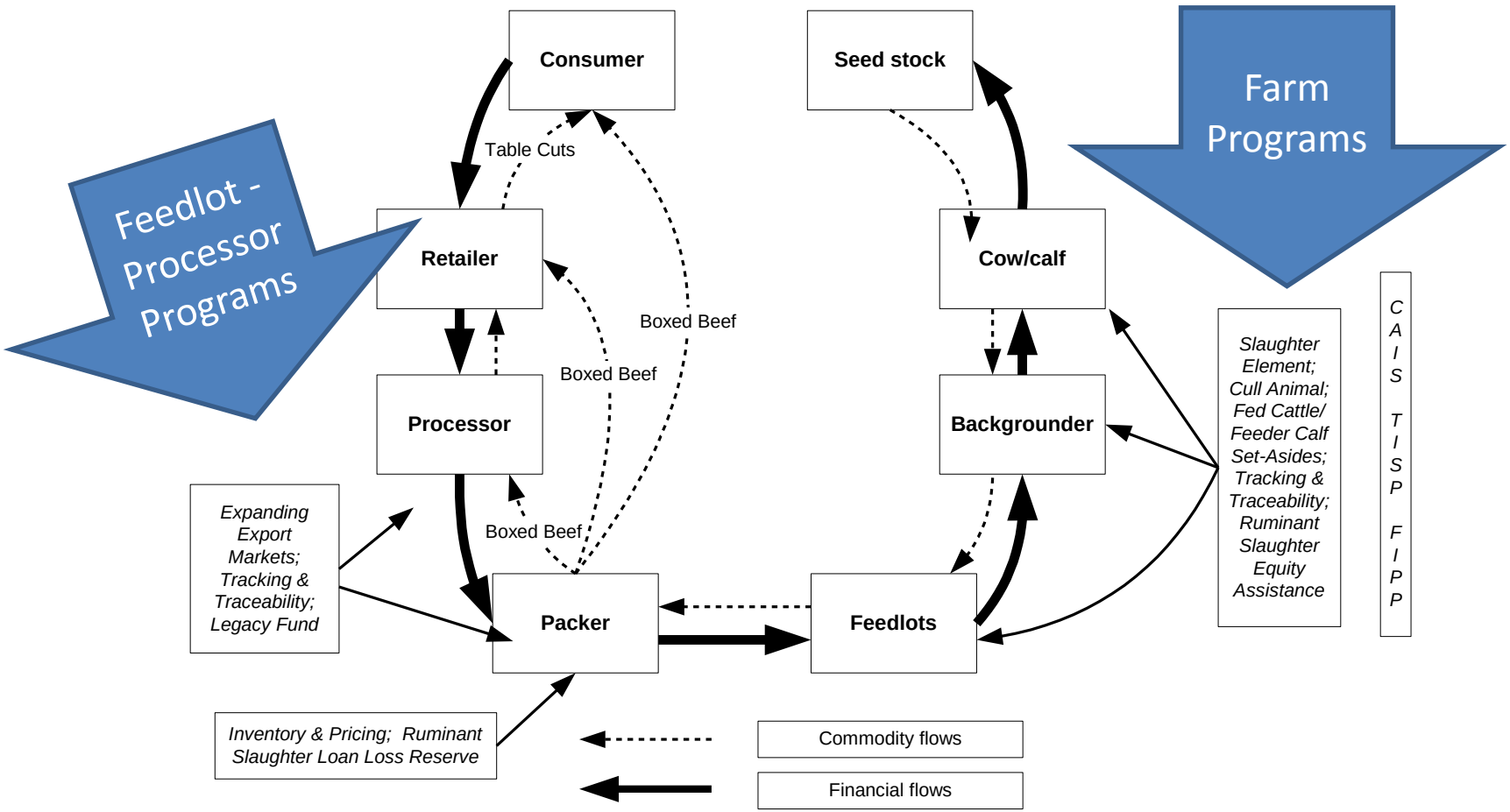
- The range in type and size of producers in the industry and vertical integration makes program decisions inherently difficult.
- No way exists for AAFC and the provinces to estimate who benefited and by how much across all programs.
- Future crises will benefit from improved understanding of the market structure of major food chains.
- The emotional and social impacts from the closure endure, are significant, and cannot be deduced from the balance sheet of individual producers.

The quantitative data (financial) illustrated the overall magnitude of the crisis, but a strong theory of change (the beef "cycle", the concept of vertical integration, and an understanding of the political dynamics (US)) revealed the impossibility of an orderly response to the crisis

Case studies illustrated key issues:

- How the financial crisis prompted extreme stress at the farmer level
- How this farm-level stress created political will to support a rapid response to the crisis
- The resiliency of the sector as some producers innovated
- The process of accelerated exit for older farmers

# Ultimately we needed a metaphor



# Part 3 – Analyse the information requirements of the evaluation

Develop informative matrices that guide the data collection and analysis

# Aside on the Policy on Evaluation

## Goal

1. Defines the obligation for departmental evaluation plans to demonstrate progress toward achieving coverage of direct program spending over five years

2. Plans that do not demonstrate evaluation coverage of all direct program spending need to use a “risk-based” approach to planning coverage

The evaluation plan needs to either show 100% coverage or identify the programs that will be assessed (and not assessed) within the 5-year cycle using a **risk-based criteria**.

# Focussing Evaluations

**Because of resource constraints (time and money), and the increased demands imposed by effective triangulation, evaluators and managers will need to *focus* evaluations.**

**This focus will be driven by a risk assessment based on**

- **materiality and**
- **strategic importance**

# Risk assessment: materiality

**Bob Woodward:** The story is dry. All we've got are pieces. We can't seem to figure out what the puzzle is supposed to look like. John Mitchell resigns as the head of CREEP, and says that he wants to spend more time with his family. I mean, it sounds like bullshit, we don't exactly believe that...

**Deep Throat:** No, heh, but it's touching. Forget the myths the media's created about the White House. The truth is, these are not very bright guys, and things got out of hand.

**Bob Woodward:** Hunt's come in from the cold. Supposedly he's got a lawyer with $25,000 in a brown paper bag.

**Deep Throat:** **Follow the money.**

*All the President's Men*

# Risk assessment: strategic importance

**Programs that are integral to the Department Agency priorities**

**Programs that, if they fail:**

- Compromise the delivery of other programs
- Compromise the priorities of the department
- Cause social, economic, and political cost disproportionate to their magnitude

**Programs that are constitutionally and legislatively required may be assigned a lower priority since the discretion on spending is limited**

- Note that the Policy identifies these as requiring only an "administrative review"

# The PAA level determines emphasis


**Focus on relevance of activities and coherence of sub and sub-sub activities. Less emphasis on performance except as established by evaluations at lower levels**

**Key methods**

- **Literature reviews**
- **Expert interviews**
- **Senior management interviews**
- **Document reviews**
- **Performance (aggregation from evaluations of lower order PAA)**

**Emphasis on relevance and alignment to immediate level of PAA.**

**Key methods**

- **Limited literature review (operational, delivery, and outcomes of cognate programs)**
- **Management interviews**
- **Project level reporting (aggregation to sub and sub-sub activity level)**
- **Quantitative measures (surveys, administrative files, etc.)**

# Core issues have transformed for federal evaluations

**Reduced from**

- Rationale/relevance
- Design/delivery
- Success/impacts
- Cost effectiveness/alternatives

**To**

- Relevance
- Performance

- The two new issues offer less guidance to development of evaluation matrices
- Issues within the two themes of relevance and performance covered by 5 Core Issues
- Many evaluation matrices either
  - reproduce the core issues with minor adjustments or
  - create massively detailed questions with multiple data source

# Relevance

| Core Issues | |
|---|---|
| **Relevance** | |
| Issue #1: Continued Need for program | *Assessment of the extent to which the program continues to address a demonstrable need and is responsive to the needs of Canadians* |
| Issue #2: Alignment with Government Priorities | *Assessment of the linkages between program objectives and (i) federal government priorities and (ii) departmental strategic outcomes* |
| Issue #3: Alignment with Federal Roles and Responsibilities | *Assessment of the role and responsibilities for the federal government in delivering the program* |

- #1 Implied in the term *demonstrable need* is
  - •whether “private sector” opportunities exist (or have been displaced)
  - •whether other orders of government offer sufficient service.
- #2 Program documentation supports the response – however horizontal initiatives are a complication
- #3 A key issue in roles and responsibilities is federal jurisdiction and constitutional alignment.

# Performance

| Performance (effectiveness, efficiency and economy) | |
|---|---|
| Issue #4: Achievement of Expected Outcomes | *Assessment of progress toward expected outcomes* (incl. immediate, intermediate and ultimate outcomes) *with reference to performance targets and program reach, program design, including the linkage and contribution of outputs to outcomes* |
| Issue #5: Demonstration of Efficiency and Economy | *Assessment of resource utilization in relation to the production of outputs and progress toward expected outcomes* |

- #4 – We see a blurring of the old formative/summative evaluation. Also apparent is the need to show a causal/attribution link (contribution) and a validation the theory of change.
- #5 – This is the cost effectiveness issue linked specifically to **economy** (are we acquiring resources/inputs at the lowest cost?) and **efficiency** (are the outputs being produced at the lowest unit cost?). The **cost- effectiveness** question (cost per unit outcome) is implied in the term “progress toward expected outcomes.”

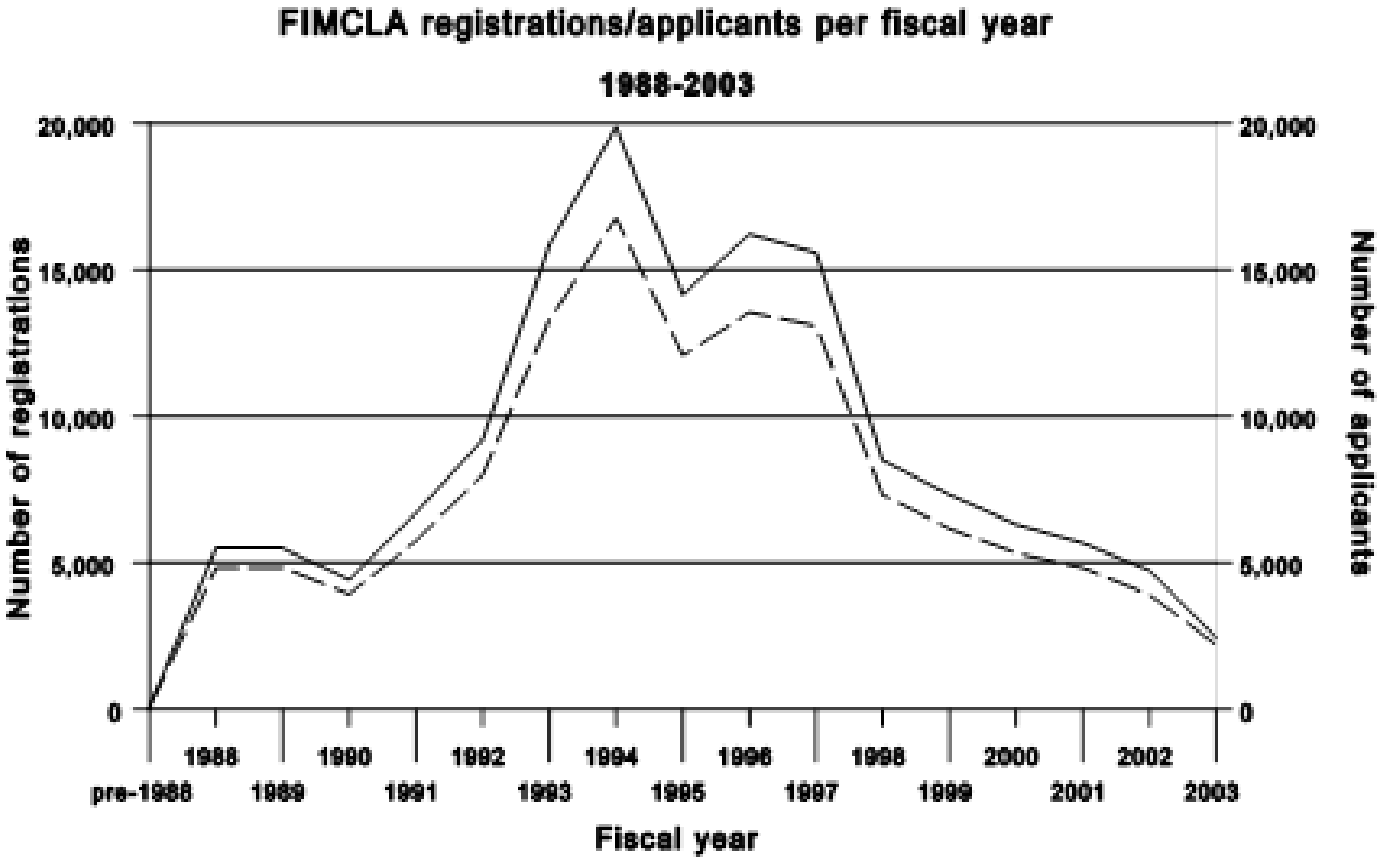
# Translating TB issues into an effective evaluation matrix

- These five questions represent an abstract structure
  - evaluators must translate these into concrete issues and questions pertinent to management needs.
- In general, the evaluation matrix must develop specific questions that support TB questions, **but…**
  - The TB questions would rarely be posed on any evaluation in this format.
  - Rephrase the TB questions to support reliable, valid, and concrete indicators that reflect the program context, goals, and implementation.
  - Management will always introduce specific issues for review

# Case example 3 - Farm Improvement and Marketing Cooperative Loan Act

**When one "fact" dominates, do we need any other information?**

- FIMCLA-guarantees bank loans to:
  - Farmers who are actively engaged in farming for the purpose of earning a profit in Canada
  - Farmer defined as "an individual, partnership, corporation, or cooperative association that is engaged in farming in Canada."
- Loans are repayable with interest fixed at 1% above prime
- The rationale framed during high interest era




This chart captures the essence of the program --- it had become a solution in search of a problem

All the other lines of evidence (survey of farmers, interviews, etc.) served to explain this decline in need

## *Interaction of quantitative and qualitative methods*

- FIMCLA as originally intended is no longer relevant to farmers
- Expansion of FIMCLA into new loan uses (e.g., first time farm purchase) requires a needs assessment and risk analysis
- Lenders are finding the loan default and claims processing activities cumbersome and costly.
- The continued need for the program depends on several factors such as potential for interest rate increases, the need to finance recovery from crises (BSE, etc.).

**The evidence that program use had effectively fallen to zero(because Canada had transitioned to a low interest environment) dominated the research process once it was revealed.**

**The single line of evidence of program decline cannot be "triangulated" with any qualitative evidence to modify the conclusion that the program was no longer relevant.**

# Part 4 – Align the evidence to the evaluation matrix

# Weave data integration into evaluation matrix design

- **Methodology** explanation adds important detail for each data source and indicator
- The matrix specifies the data collection methodology and integration process

KISS – Keep it simple and sophisticated

Evaluation matrices tend to be repetitive with redundant questions, often fail to guide the collection and *interpretation* of data

# From logic model to matrix

- Management shapes the purpose of evaluation by selecting the questions of interest
- The evaluation matrix guides the integration of data collection and analysis
- The matrix determines the success of the evaluation
- It shapes the direction and depth of analysis

## IQIDM

# Principles of evaluation matrix design (1)

- **Issues** need to align with the mission/goal of the program
  - High-level language is ok for the TB issues, but concrete and "grounded" plain language specifications are preferred
  - This translation will support measurement
  - Ensure that the service lines (program pillars) emerge at the issues level
- Do not replicate the TB structure … it is too general
- Parsimony (a few focused issues) is preferred

**Frameworks that reproduce the TB issues/questions reflect lazy work that sabotages mixed methods and effective triangulation (passing the buck)**

# Principles of evaluation matrix design (2)

- **Questions** are operational and specific to the program and service lines
  - Use the results chain and logic model to identify key delivery points/times/processes for *outputs*
  - Focus on immediate *outcomes*
- Questions align with indicators and data collection
- Rank questions within an importance level (H,M,L) to direct the allocation of evaluation resources.

# Principles of evaluation matrix design (3)

- **Indicators** describe the information needed to answer the question
- Detailed descriptions support reliable and valid data collection
- Align indicators with source based on expected information content and quality

*Example: What immediate outcomes (first five years) were expected at the program's inception?*

- Client opinion *(poor)*
- Line management opinion *(slightly better)*
- Senior federal and provincial manager opinion *(even better)*

# Principles of evaluation matrix design (4)

- **Data sources must align to each indicator**
- Detailed descriptions of sources must be specified to support efficient evaluations

*Example: Senior federal and provincial managers' opinion on alignment to departmental goals*

a. Poor - Unspecified key informant interviews with a single guide
b. Better – Align interviews into cognate groups [Interview with federal agreement managers (n=3); Interview with ADM(s) (n=2); Interview with Provincial/territorial Agreement managers (n=13)
c. Even better - Align types of key informants to specific questions and create specialized interview guides for each class of key informant

Core challenges for federal evaluators are

- poor inherited logic models,
- general evaluation matrices and
- truncated time/budget which limit interview segregation and feedback

# *Case example 4 – Horizontal Summative Evaluation of the Government of Canada's Investment in the 2010 Olympic and Paralympic Winter Games*

## The power of the evaluation matrix to unify lines of evidence

**Immediate outcomes**

- Leverage 2010 Games to advance existing federal priorities
- Positive exposure and heightened recognition of the Government of Canada as a key partner in the 2010 Games
- Successful delivery of the mandated essential federal services.

**Intermediate outcomes**

- Pan-Canadian engagement in sport, economic, social, and cultural activities related to the 2010 Winter Games
- Enhance Canada's domestic and international profile
- Canadians and international participants experience safe and high-quality Games

**Final outcomes**

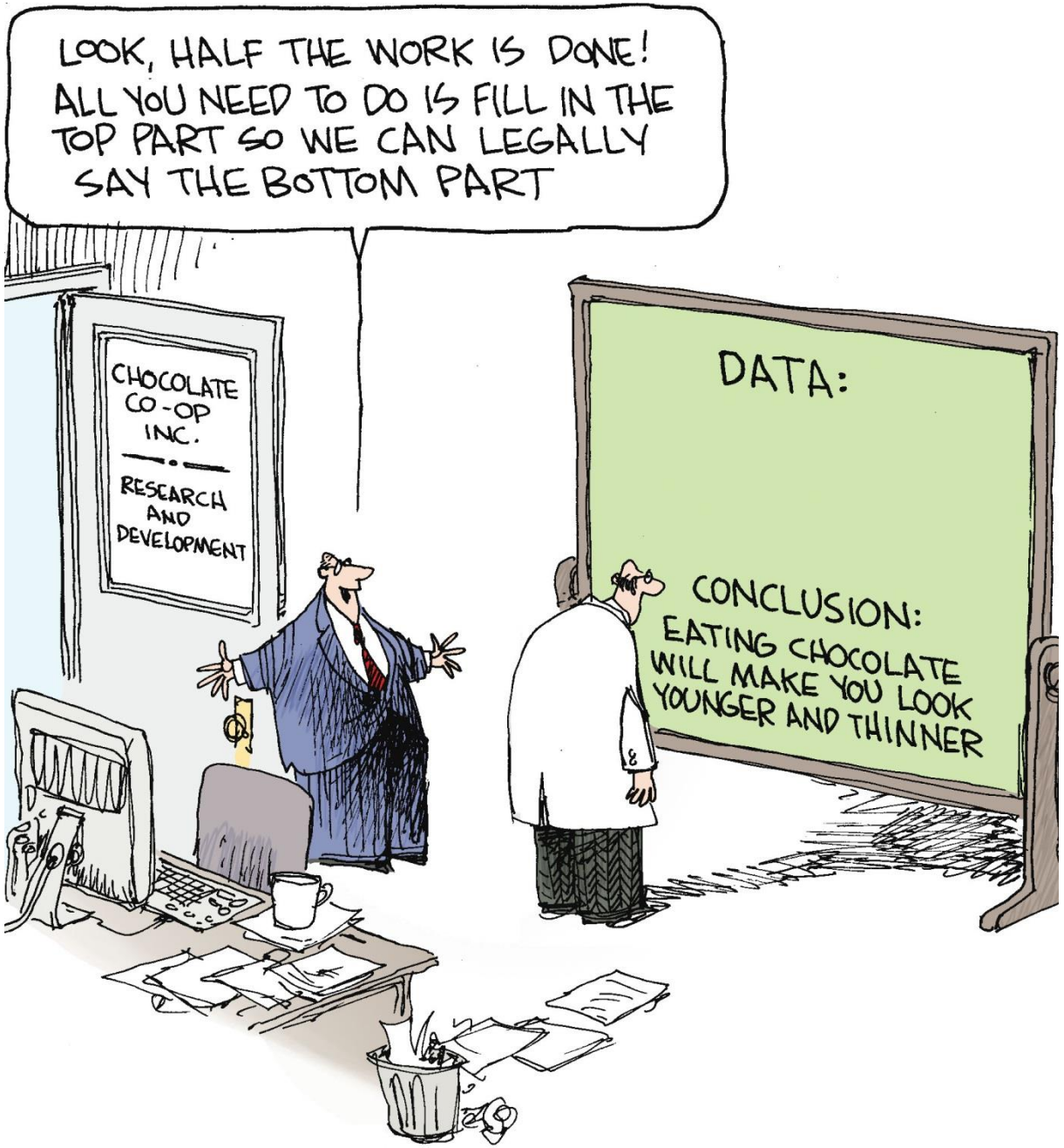
- Sport, economic, social, and cultural legacies are established for the benefit of Canadians
- Canadian excellence and values are promoted domestically and internationally
- Canada is recognized as a capable and inclusive host.

•(

# 2010 Winter Games - Main methods

- Main evaluation focussed on the role of the Federal Secretariat and the contributions of:
  - 15 federal departments provide services deemed essential for conducting successful Games
  - Auxiliary services
    - PCH received support for opening ceremonies
    - INAC received funding to ensure legacies, benefits and participation of fist nations
- Security (separate evaluation conducted by RCMP)
- Eight other studies comprised the evaluation
- Polling used to track “national pride”
- Media analysis complemented quantitative polling
- Interviews (directed to specific classes of respondents).

# Concluding observations

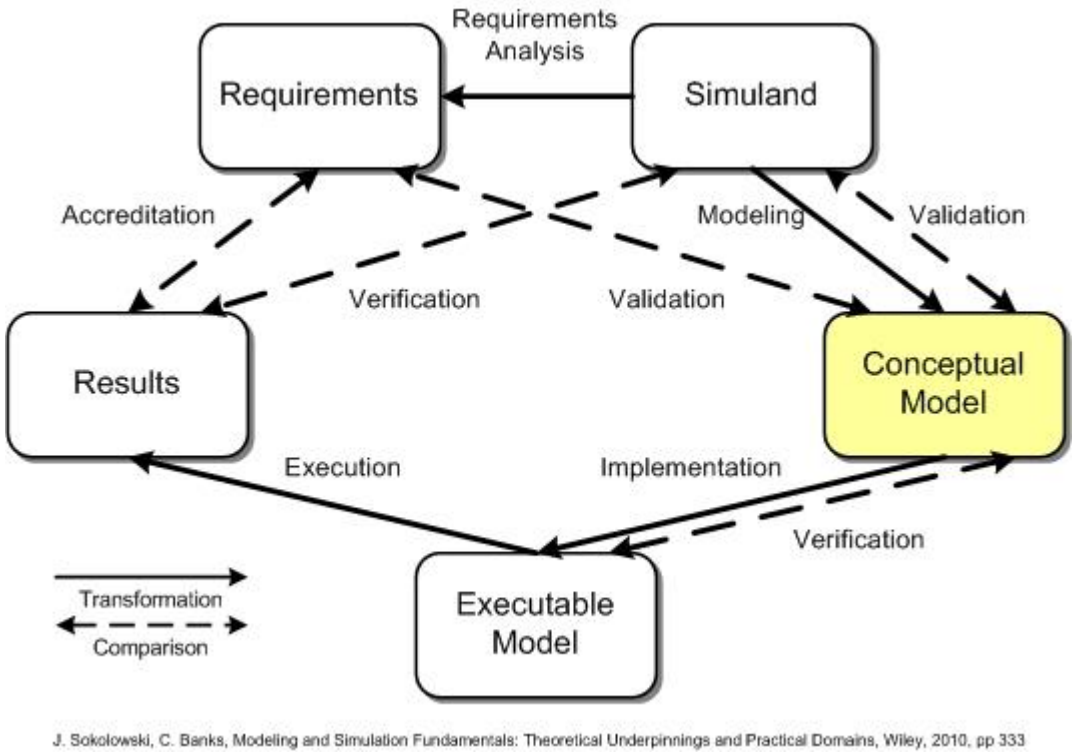



© Wiley Ink, inc./Distributed by Universal Uclick via Cartoonstock

# What combining qualitative and quantitative data is ***not***

$$r = \frac{1}{n-1} \sum \left( \frac{x - \overline{x}}{s_x} \right) \left( \frac{y - \overline{y}}{s_y} \right)$$

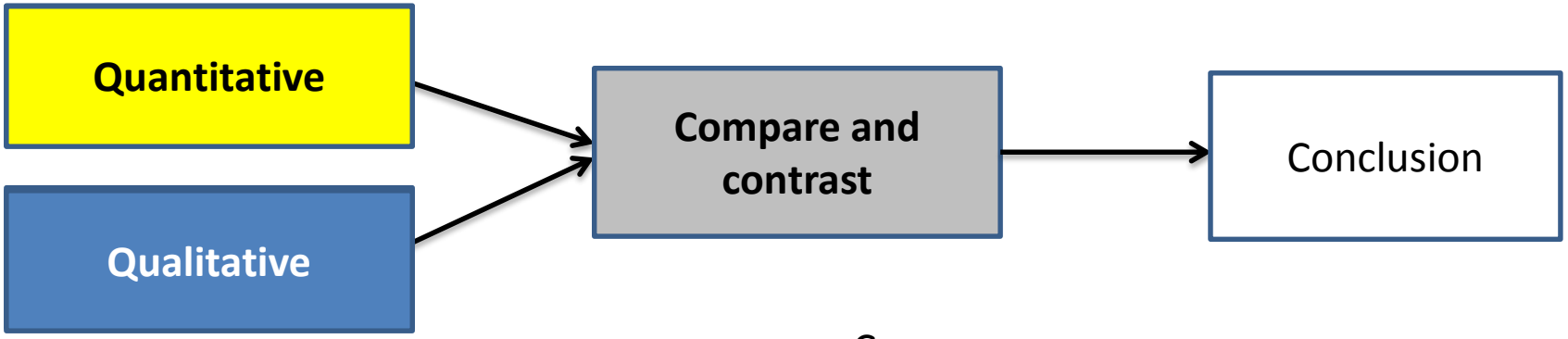
Or….




J. Sokolowski, C. Banks, Modeling and Simulation Fundamentals: Theoretical Underpinnings and Practical Domains, Wiley, 2010, pp 333

# Interaction of quantitative and qualitative methods
# Four models*

## 1. Triangulation




Pros

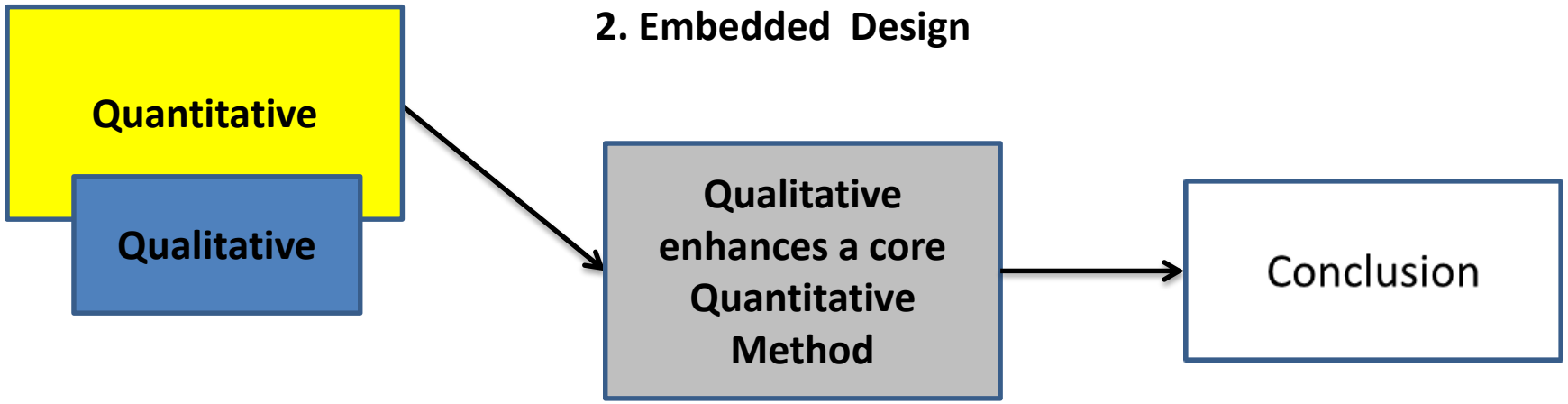
- The most popular concept
- Seen to align Quant and Qual methods as complementary
- Qual data are often transformed to Quan (using coding)
- Intuitive approach – appears to balance all types of data
- Less costly

Cons

- Reflects poor methodology
- The time and resource demand are high to ensure all methods are developed
- Procedure to combine different types of data must be explicit
- It can be very challenging to design Quant and Qual data to address the same question

# Interaction of quantitative and qualitative methods
# Four models*

## 2. Embedded Design




Pro

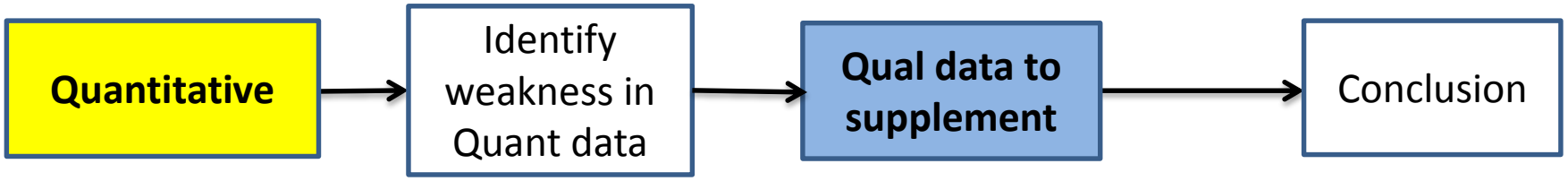
- Strong Qual data support measurement of performance and attribution (development of instruments and interpretation)
- Supports sequential analysis
- Emphasis on Quant data tends to be familiar to most

Con

- Weak method when Quant data are poor (opinions of impact are not credible)
- The role of Qual data as "support" to Quant methods needs explanation
- Poorly executed Qual data will bias Quant methods.

# Interaction of quantitative and qualitative methods
# Four models*

### 3. Explanatory Design




Pros

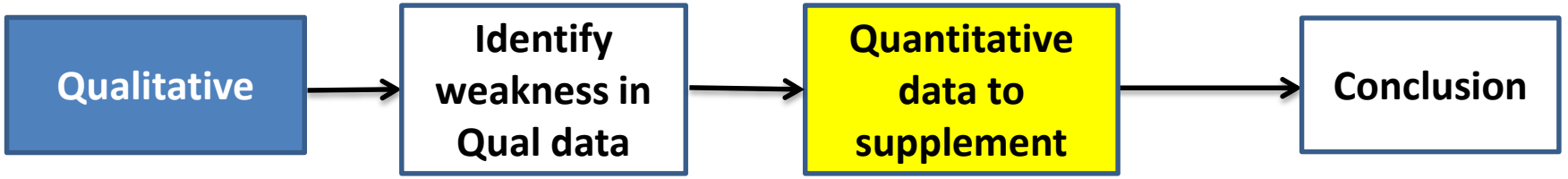
- Quant precedes Qual data collection
- Tends to emphasize Quant results
- Qual used to explain and add insight to Quant results
- Quant results can be used to design Qual research (e.g., selecting focus group participants and case studies from client survey)
- Study design is more manageable

Cons

- Sequential phasing can lengthen the research
- Quant data collection will dominate Qual data collection
- Qual data may reveal weaknesses in Quant data that compromises the overall design, requiring repeated data collection.

# Interaction of quantitative and qualitative methods
# Four models*

## 4. Exploratory Design




Pro

- Qual data used to explore a concept
- Qual data and analysis tend to dominate
- Quant data used to generalize the Qual results
- Easy to implement and tends to make Qual results more acceptable

Con

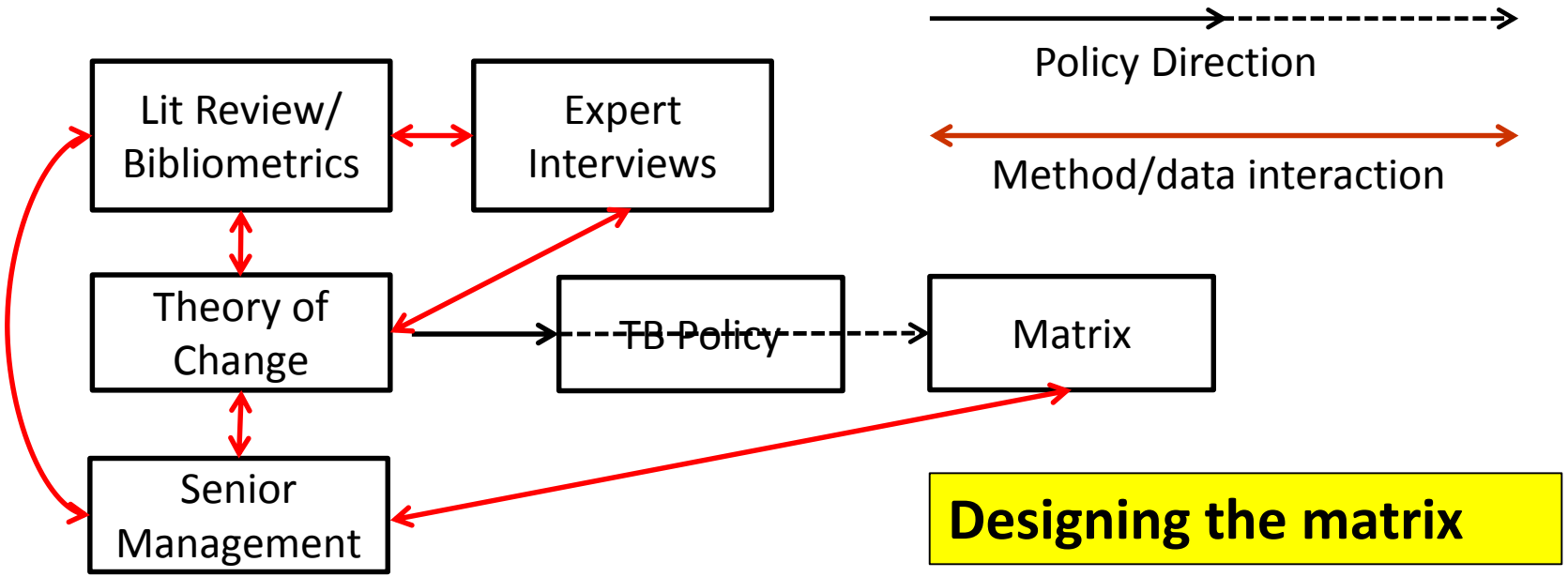
- Tends to take a long time
- Quant data collection may need to be revised in the light of Qual results

# Integrating evidence in evaluation

**Integration/synthesis are superior terms to the term "triangulation**

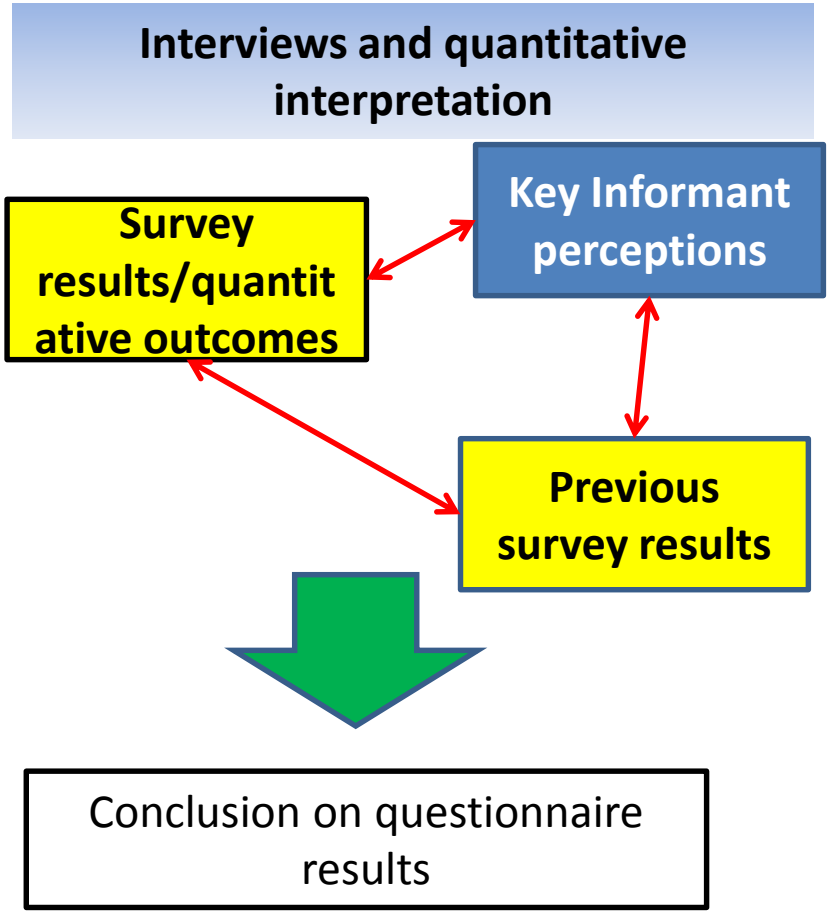
**Integration of Quant/Qual occurs every step of the analysis and not at the end of the study**

**Active hypothesizing, evidence review and revision of findings occurs throughout the evaluation**



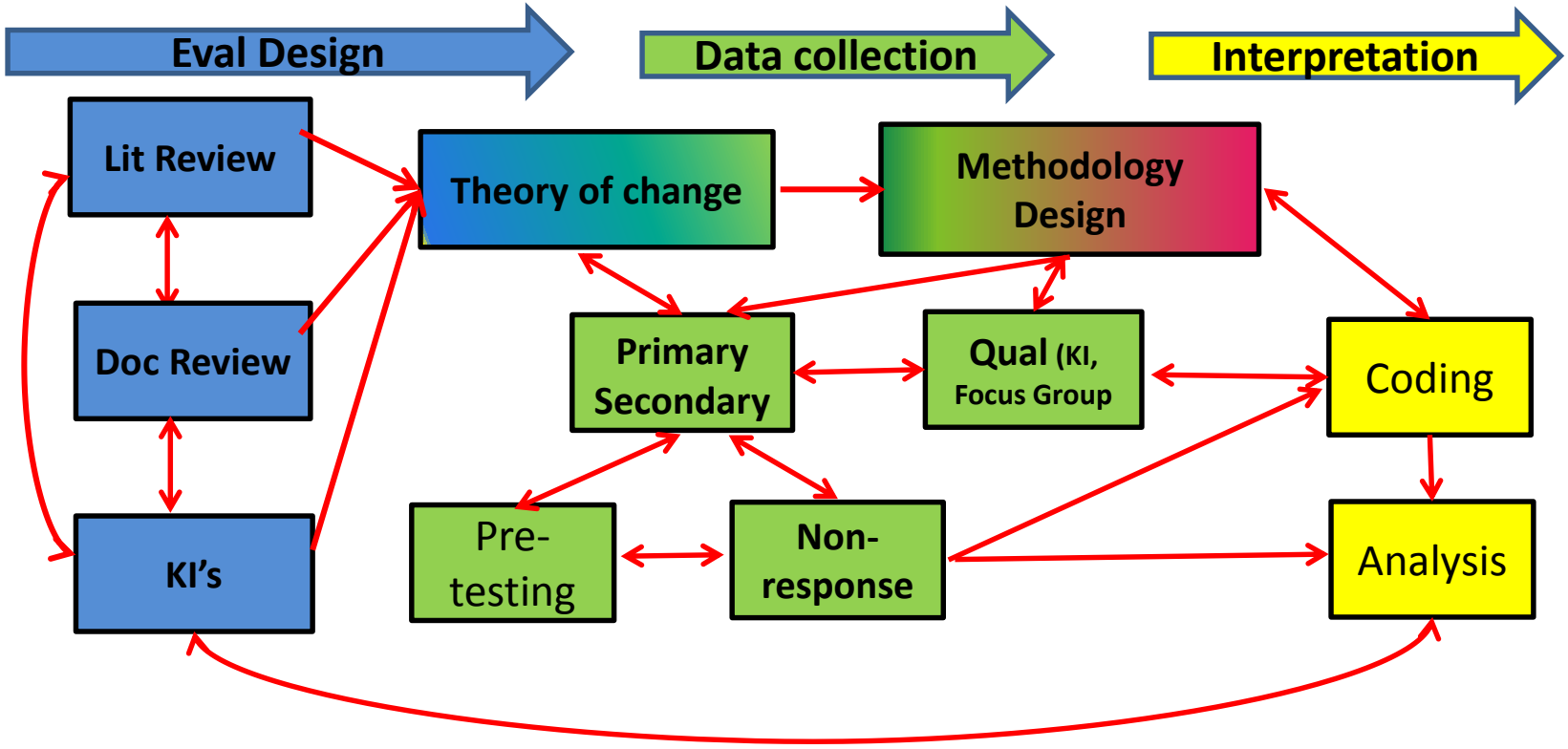
# Integrating Qualitative/Quantitative Evidence
# Key informant interviews

- **Segment the key informants**
  - Subject matter experts
  - Senior managers
  - Project proponents (G&Cs)
  - Line managers
  - Client
- **Align key informants, cases, focus groups, etc.** to specific and limited questions
- Create **short, focused guides** aligned to specific questions
- Ask **appropriate questions**
  - Limit questions solely to areas of legitimate experience
  - Do not ask clients about program relevance…do ask about the project and direct service experience
  - Do not ask senior managers about quantitative outcomes **except** where you are presenting data for their reaction and interpretation.



# Integrating Qualitative/Quantitative Evidence
# Client participant/surveys

Integrating quantitative and qualitative data collection can proceed without theory and perspective in design *and* theory and perspective in interpretation



# Data integration vs. Triangulation

- Data integration starts with the evaluation design
- Data integration
  - Occurs during each data collection step
  - Bridges transitions among data collection steps
  - Supports “looping back” for re-analysis
- Data integration never starts after data ceases.

- Triangulate only within a data methodology applied to similar data:
  - Alternative statistical models using the *same data*
  - Contrast the views of *similar key informants* (within national managers, within local project leaders, within line social workers….)
  - Across multiple *homogeneous* focus groups to understand multidimensionality of experience and perceptions within that type of participant.

# Guidelines

- The evaluation matrix is the foundation of mixed methods in evaluation.
- *Information integration* occurs at every stage of the evaluation design, data collection and interpretation.
- Weakness in one line of evidence (e.g., quantitative measures of performance) cannot be compensated by intensifying other (qualitative evidence).
- Adding uninformed interviewees will never increase reliability and validity –
  - the best outcome is higher cost
  - the likely outcome is higher cost and weaker results.
- Do not apply the rules of quantitative data collection/analysis to qualitative data collection/analysis.
- Align evidence to the question.
- Be clear about the specific (and varied) information offered by line of evidence for each evaluation question.
- Do not add a line of evidence without a clear and detailed understanding what incremental benefit will be offered.
- KI's, case studies, focus groups must be selected for their information value and not in a false attempt to increase reliability.

# Bibliography