The Mix of Methods: towards a framework for anticipating validity threats in evaluations of agricultural value chain support

Giel Ton
Agricultural Economics Research Institute - LEI Wageningen UR

9th European Evaluation Society International Conference 2010
Prague, 6-8 October 2010

LEI
WAGENINGEN UR

Our room of maneuver in impact evaluations

- Each researcher in LEI has its own expertise and methodological wish-list: multiple ways to do evaluations:
  - Models/scenarios
  - Econometrics
  - Case studies
  - Stakeholder processes
- My subgroup mixes economists, anthropologists and engineers
  - We adopt Theory-Based Evaluations
  - We want to improve our research designs in a peer-to-peer process of design quality checks

LEI
WAGENINGEN UR

Theory-based evaluation

- We know that these value chain development processes are complex, and, therefore:
  - We need 'program theories' to be evaluated and tested
  - We need to explore the conditions that make them work (towards 'good principles'/'good practices')
  - We want to maximize evaluation outputs that facilitate learning for 'better' intervention theories

Challenges

1. Focus on key aspects!
2. Methods that can face scrutiny!
3. Outputs that facilitate cross-site learning!

LEI
WAGENINGEN UR

1. Focusing the impact evaluation

- We propose a process to focus in a process in which stakeholders define/refine their intervention logic
- We want to reflect with them to identify the critical assumptions in that logic
  - One of the obvious assumptions, is IMPACT: outcomes can be attributed to the intervention
  - Less obvious, but interesting for social research, are the assumed CAUSALITIES: the assumption that one process causes/triggers other processes

LEI
WAGENINGEN UR

- **Impact Logics / Result Chains:** How do we think value chain performance will be enhanced by our support intervention?
  - On what basis do we think that the proposed interventions are likely to be successful in improving the performance of the value chain?
  - What mechanisms are assumed to work that translate our activities into (intermediate) outcomes? (the arrows!)
  - Can we collect evidence to make the most important causal links plausible to a skeptical outsider?

2. Mixed method measurement tools

- The lack of credible evidence on outcomes and impact of value chain development support:
  - low priority on measuring impacts by practitioners
  - lack of appropriate, lean and credible instruments to do so.

Therefore:

- Step 1: Choice/Negotiation of a core methodology that fits with the main evaluative questions (and 'real-world constraints')
- Step 2: Add to this core method with some additional methods:
  - That responds to the most challenging validity threats of the expected evaluative conclusions
  - That anticipates eventual implementation issues related to the core method

LEI
WAGENINGEN UR

■ We propose to check the core research method design on the most obvious threats to validity, exploring the issue from four different angles:

    a) statistical conclusion validity

      – when using statistics, do it properly

    b) internal validity

      – resolve the issue of causality/attribution

    c) construct validity

      • are the concepts used properly defined and operationalized

    d) external validity

      • under what conditions/settings does the conclusion/recommendation apply

Source: Shadish, W. R., T. D. Cook, et al. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Houghton Mifflin Co. Boston, MA.

LEI
WAGENINGEN UR

3. Facilitate cross-site peer-to-peer learning

- Need to reflect on common elements in value chain development pilots

- Focus on mechanisms in context:
  - What works for whom under what conditions?

- Realist case studies: Context-Mechanism-Outcome Configurations

LEI
WAGENINGEN UR

Application in the design of three
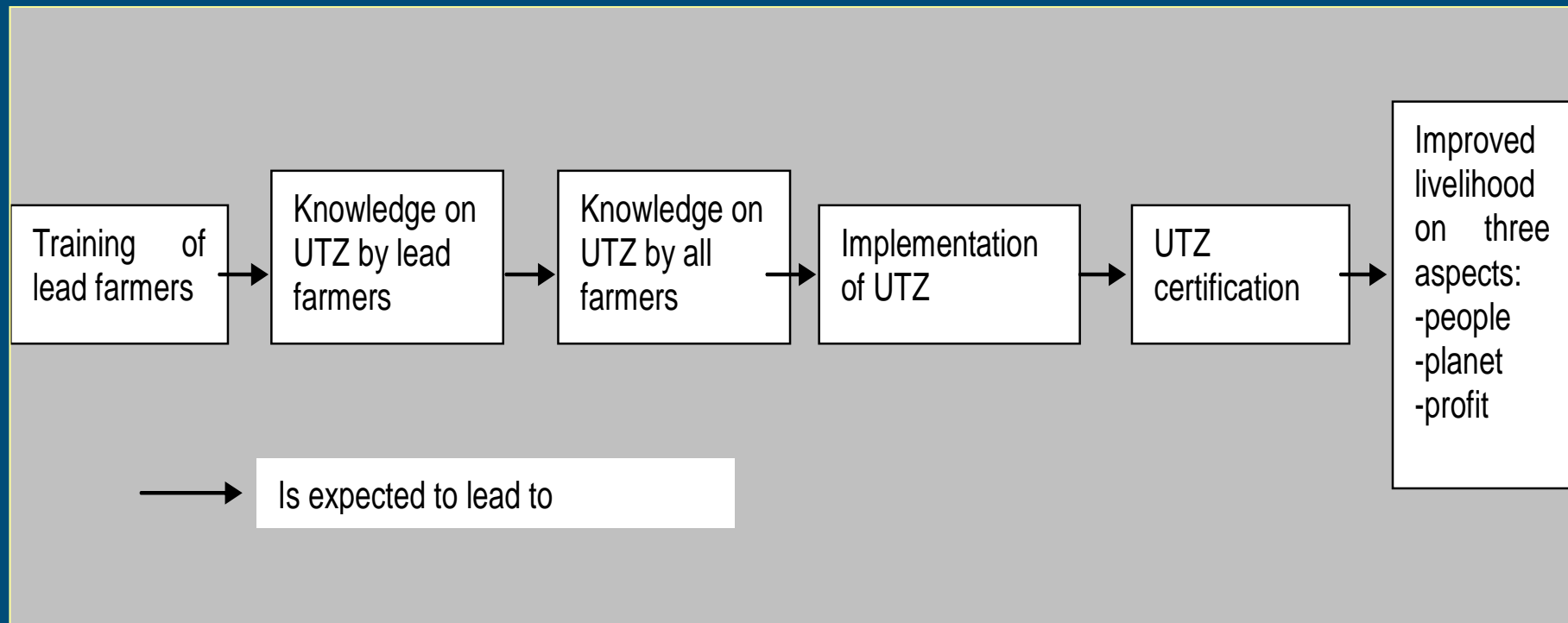# of our evaluation assignments

Income impacts of micro-irrigation technology
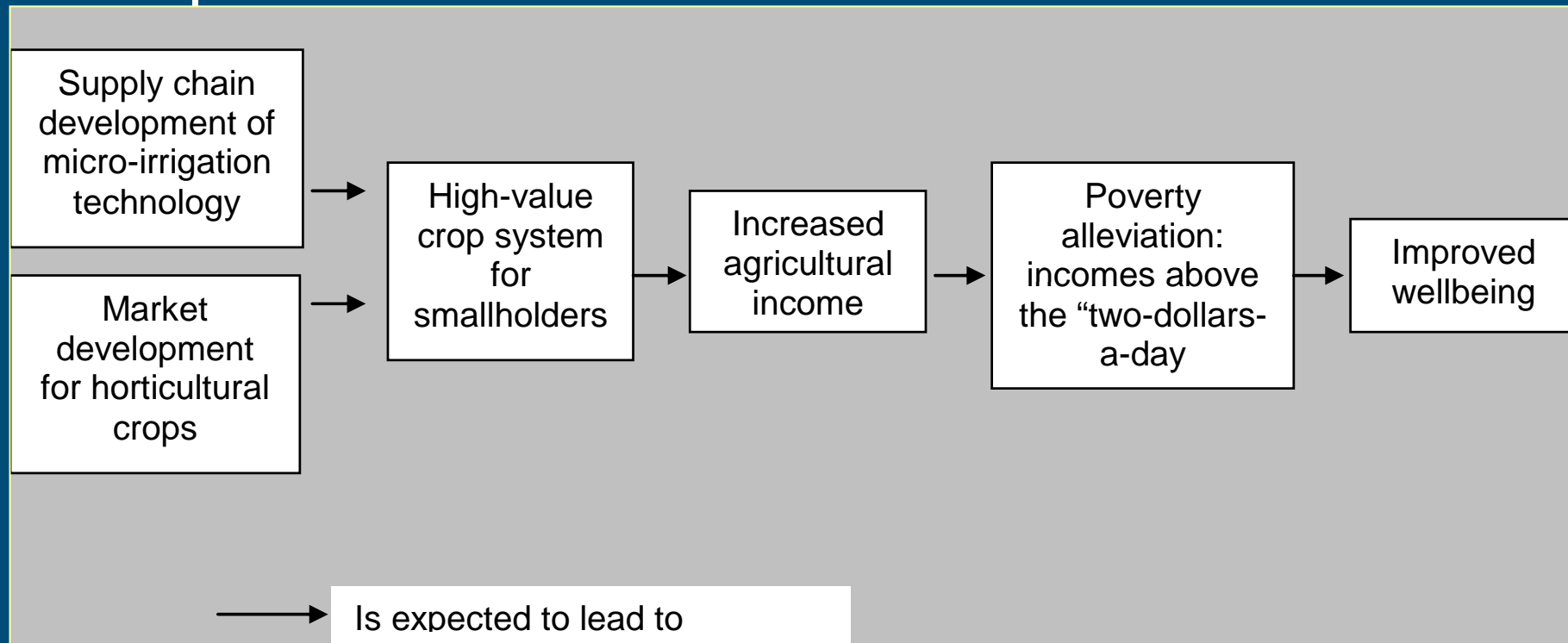
- **Critical assumption**: attribution of household income changes to micro-irrigation technology use
- **Core method**: 'pipeline design' with retrospective baseline
  - Comparing income streams between yearly customer cohorts
  - Asking there about before and after adoption
- **Added mixed method**:
  - On key assumption in program theory
    - Livelihood impact <u>case studies</u>
    - <u>Sector-studies</u> on dynamics in markets and institutional environment
  - On methodological assumptions
    - <u>Recall bias test</u> (repeating measurements in the same households with different recall period)
    - <u>Selection bias</u>: compare between inter-cohort characteristics and their differences with a random sample of the population

Impact of training on tea sustainability

Training of lead farmers → Knowledge on UTZ by lead farmers → Knowledge on UTZ by all farmers → Implementation of UTZ → UTZ certification → Improved livelihood on three aspects:
-people
-planet
-profit

→ Is expected to lead to

LEI
WAGENINGEN UR

- **Critical assumption**: trickle down of training contents outside the core group
- **Core method**: before-after scan on good agricultural practices
- **Added mixed methods**:
  - On key assumption in program theory
    - Realist case studies on differences between tea factories (e.g. meeting intensity, additional stimuli)
    - Check on differences in access conditions of households for some 'necessary' equipment
  - On methodological assumptions
    - Additional checks on selection bias during baseline and discuss what to do with that knowledge

Impact of investment subsidies on associative business
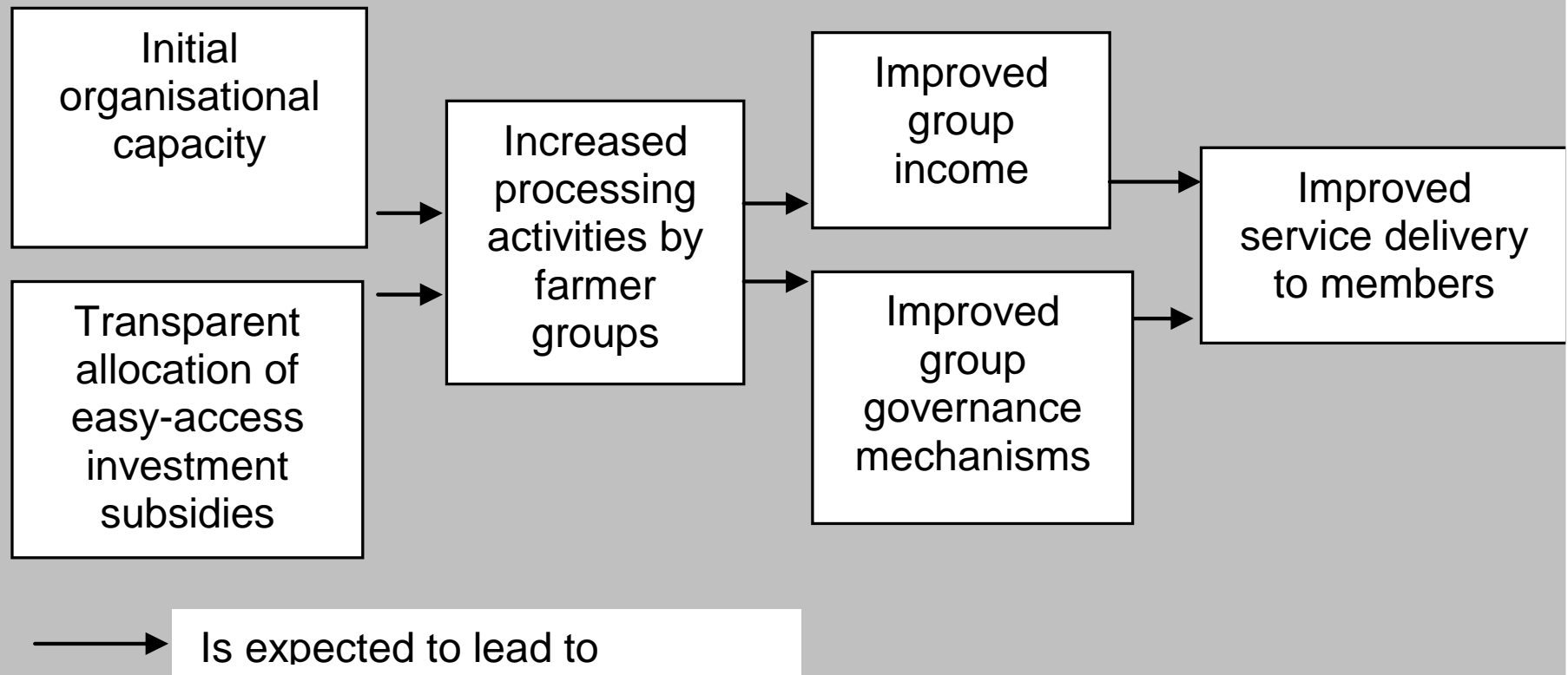
- **Critical assumption**: attribution of changes in organisational capabilities to collective processing
- **Core method**: base-line and three years after organisational scan on a random sample of organisations, analyzed with case-based statistical methods
- **Additional mixed methods**:
  - On key assumption in program theory
    - Descriptions of learning experiences on resolving tensions in collective action (realist case studies)
    - Comparison of preparation process with rejected plans:
  - On methodological assumptions
    - Repeated measurements of scan-instrument in the same organizations, with different respondents (robustness of measurement instrument)
    - Peer-to-peer workshops to discuss the generalisation domain of supposedly 'stronger' governance mechanisms

Conclusions

- **One-method research might be good for publication in top journals, but rarely for generating convincing evidence to stakeholders**
- **Need for proper mixed-methods design:**
  - Theory-based evaluation to find the right evaluative questions
  - Scan on validity threats to the expected 'type of conclusion', in order to find a proper mix of methods
  - Realist case-studies to place mechanisms in context
  - Case-based comparative methods to make sense of diversity in development pathways

Thanks!

Giel Ton, giel.ton@wur.nl