Artimino, June 27° 2016 - Connected Health Summer School Smarter Environments for Smarter Living

Monitoring and Evaluating the Multidimensional Impact of Complex Social and Health Care Services: Field Experiences and Methodological Issues

Federico Ciani

University of Florence and ARCO Lab

federico.ciani@unifi.it

Some **definitions**…

- **Monitoring:** tracks key indicators of progress over the course of a program
- **Operational Evaluation**: examines whether there are gaps between planned and realized outcomes of a program;
- **Impact Evaluation**: Studies whether the changes in well-being are indeed due to the program intervention and not to other factors

Some definitions…

- **Evaluation ex-ante**: predict program impact using data before the intervention (simulation, structural model)
- **Evaluation ex-post**: examine outcome after program have been implemented through the difference in participant outcomes before and after the programme implementation (or across participants and non participants $\Rightarrow$ collecting data on actual outcome for participant or not participants)
- Qualitative vs(?) Quantitative evaluation

Some definitions…

- **Programme**: a well defined intervention targeted to a well defined population, with the purpose of inducing a change in a well defined state
- **Target-population**: a well defined set of units upon which the intervention will possibly operate at a particular time. Units can be persons, households, firms, schools, villages, countries. All units of the target population could in principle take part in the intervention
- **Intervention**: an intervention (i.e. treatment), the effect of which on the outcome variable the analyst wishes to assess relative to no intervention
- **Outcome variable**: an observable characteristic (i.e., some particular measurement) of the population unit, on which the intervention may apply/not apply and may have an effect.

What can’t be evaluated

- “macro-policies”: monetary, fiscal, industrial, environmental, etc. (exception: change of a policy regime; e.g., a monetary shock);
- major infrastructure and public works, basically unique and irreversible
- the current provision of services, chiefly by public administrations (exception: a reform/discontinuity in the provision of services)

Counterfactual: the key concept

- IE aims to determine what would have happened to the beneficiaries if the program had not existed
- A beneficiary's outcome in the absence of the intervention would be its counterfactual;
- But..... the counterfactual cannot be observed in the real world

Counterfactual: the key concept

- The potential outcome framework and notation (counterfactual analysis, sometimes referred to also as the Rubin's Causal Model) will be exploited throughout: for all individuals, a set of outcomes is logically defined across treatment states.
- Let (Y_1,Y_0) be the two outcomes corresponding to a specific population member being treated or not treated, respectively. The two outcomes are logically defined, but only one of them is observed depending on the treatment actually experienced.

Counterfactual: the key concept

- If a specific member of the population is exposed to the intervention Y_1 is observable, while Y_0 is irreversibly unobservable on that specific member
- The counterfactual outcome for a member of the population who did participate in the programme is Y_0 , that is what we would have observed for the same member had not participated.

Counterfactual: the key concept

$$Y = Y_1 D + Y_0 (1 - D) = \begin{cases} Y_1 & \text{if } D = 1 \\ Y_0 & \text{if } D = 0 \end{cases}$$

	Factual outcome	Counterfactual outcome
Participants ($D = 1$)	Y_1	Y_0
Non-participants ($D = 0$)	Y_0	Y_1

From counterfactual to causality

- For each member of the reference population the causal effect of an intervention corresponds to the difference

$$\beta = Y_1 - Y_0$$

- It is well defined for all members of the reference population, that is, irrespectively from participation.
- In words, $\beta = Y_1 - Y_0$ is the outcome change of a specific individual that is due to switching from state D=0 (no treatment) to state D=1 (treatment).

The problem of the problems

- For each unit of your population you can observe only one outcome (Y_0 or Y_1)
- Basically IE is a problem of missing data
- All IE methodologies are efforts to find a good substitutes of the outcome you can’t observe

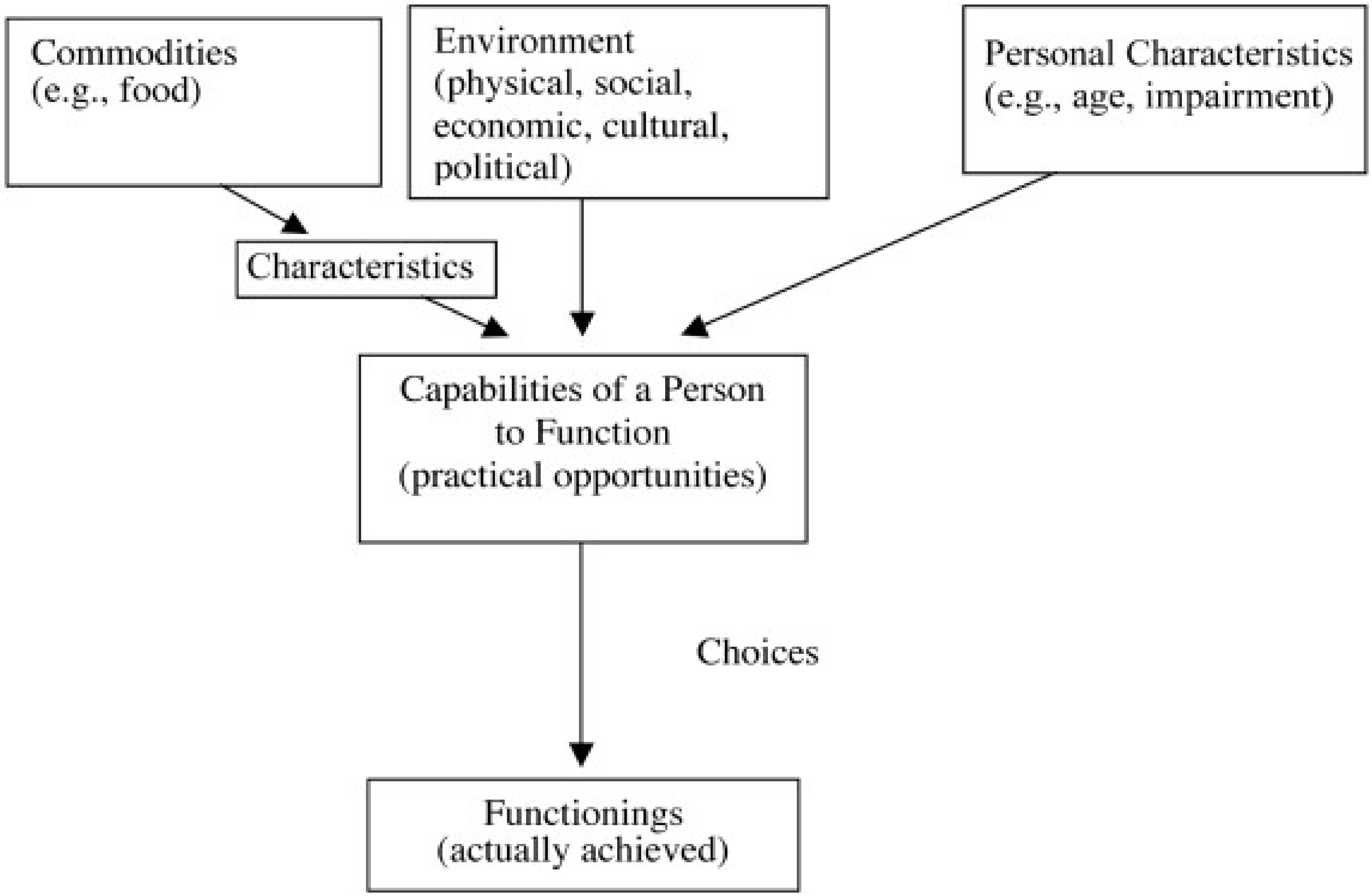
Possible solutions

- Two main roads:
  - Modify the targeting strategy of the program itself to wipe out differences that would have existed between the treated and non-treated groups before comparing outcomes across the two groups (experimental methods: randomization)
  - Create a comparator group through a statistical design (non experimental methods: PSM; DD; RDD; IV)

A theorethical framework for social and health care services: Capability Approach

- Normative approach to wellbeing and social justice elaborated by the Nobel Laureate Amartya Sen

- Multidimensional wellbeing
- Person centered
- Value to freedom
- Good conceptual approach to deal with disability and ageing
- Coherent with UNCRPD approach


Commodities
(e.g., food)
Environment
(physical, social,
economic, cultural,
political)
Personal Characteristics
(e.g., age, impairment)
Characteristics
Capabilities of a Person
to Function
(practical opportunities)
Choices
Functionings
(actually achieved)

Social and Health Care services

- Conciliate key features of the human development and capability approach and the need of assessing the impact of policies
- Needed to have a full operationalisation of human development and capability approach
- Evidence based policy making for human development policies
- Needed to overcome the limits of “mainstream” impact evaluation

- A first set of challenges deals with the kind of variables you need and the ways you can structure data collection

Key issues and possible solutions

- Focus on the ends of the intervention process and not on the means
- Choice of appropriate outcome variables
- No rules but deep analysis of the context

Key issues and possible solutions

- Human development is multidimensional and deprivation is multidimensional
- Going beyond mere health outcomes, income, wealth
- Take into consideration a reasonably large set of variables starting from a theoretical idea of development and wellbeing

Key issues and possible solutions

- <u>Human development focuses of distribution, equity and inclusion/exclusion dynamics</u>
- Take into consideration side effects
- Take into consideration vulnerable groups (e.g. the impact on the poorest, children, persons with disabilities) during sampling and tools design
- Cross cutting issues (gender, ethnic minorities)

Key issues and possible solutions

- Human development and capability approach is a person centered approach
- Include personal preferences, beliefs and behaviors in the evaluation
- Collect data about perceptions (e.g. subjective wellbeing)
- Involvement of stakeholders to identify key dimension for the evaluation

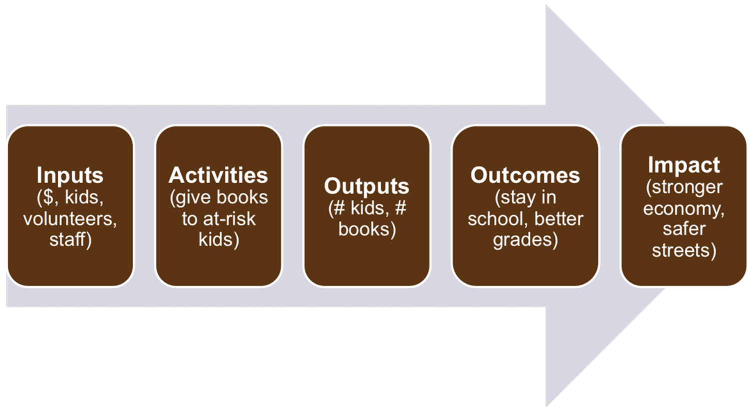
Key issues and possible solutions

- Human development and capability approach gives value to the way a result is achieved
- Value to democratic and participatory processes
- Centrality of agency
- Collect information about critical dimensions such as participation, empowerment (in particular for marginalised groups), collective action
- Need to understand processes and not only results

- A second set of challenges clashes against some structural characteristics of mainstream impact evaluation

The issue of complexity

- Human development based policies are usually multi-level and multi-stakeholders
- Human development and capability approach based policies embodies personalized and tailored treatments
- Often impossible to describe this complexity within a single linear theory of change



Mixed methods

- Two main objectives of IE
  - Measuring the impact→quantitative methods
  - Understanding the process→qualitative methods
- Need to combine qual. and quant. methods
- Definition of a mixed-method based study

“a study qualifies as adopting a mixed methods approach if qualitative data collection and analysis are explicitly included in the study design” (White, 2008)

- Very broad definiton→many different ways of combining approaches
  - ➢ Integration of methodologies
  - ➢ confirming/reinforcing, refuting, enriching, explaining the findings
  - ➢ Merging findings

Mixed methods

Type of Mixing	Type of Design	Why Mixing Occurs	Where Mixing Occurs in Research Process
Connecting	Sequential	One phase builds on the other	Between data analysis (Phase 1) and data collection (Phase 2)
Merging	Concurrent	Bring results together	After analysis of both quan and qual – typically in discussion
Embedding	Sequential or Concurrent	Either building or bringing results together	Either between phases or in discussion after analysis

Which sample size?

The first step in designing a randomized experiment is to choose a sample size and allocation that maximize precision given existing constraints.

For this purpose, it is useful to measure precision in terms of **minimum detectable effects**. Intuitively, a minimum detectable effect is the smallest true treatment effect that a research design can detect with confidence.

Formally, it is the smallest true treatment effect that has a specified level of statistical power for a particular level of statistical significance, given a specific statistical test.

95% confidence interval:

$$(\bar{Y}_{treated} - \bar{Y}_{control}) \pm 1.96 \times \sqrt{\frac{\sigma^2}{np(1-p)}}$$

If the expected sign for the effect is positive, the lower bound should be above zero:

$$(\bar{Y}_{treated} - \bar{Y}_{control}) - 1.96 \times \sqrt{\frac{\sigma^2}{np(1-p)}} \geq 0$$

A common convention for defining minimum detectable effects is to set statistical significance at 0.05 and statistical power at 80 percent. Statistical significance and statistical power translate into a multiplier of the standard error.

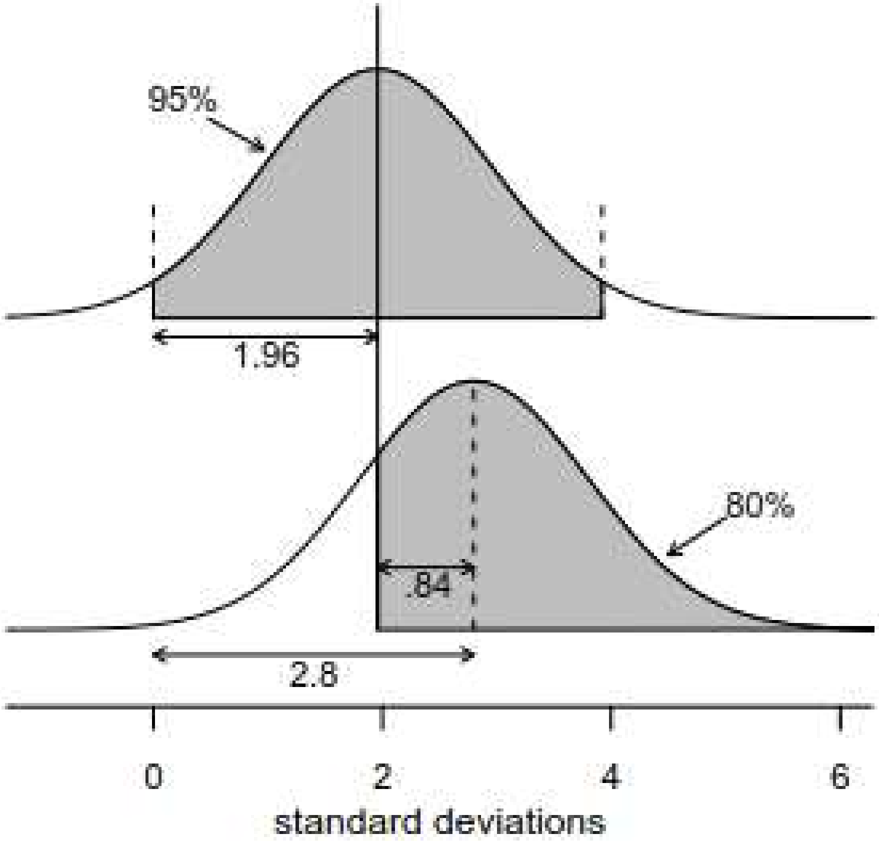
When the number of degrees of freedom exceeds about 20, the multiplier equals roughly 2.5 for a one-tail test and 2.8 for a two-tail test.

When the outcome measure is a one/zero binary variable the variance estimate is p(1-p)/n where p is the probability of a value equal to one.

The usual conservative practice in this case is to choose p=.5, which yields the maximum possible variance = 0,25.

$$MDE_{(P_T - P_C)} = 2{,}5\sqrt{\frac{0{,}25}{0{,}25 * N}}$$

$$MDE_{(P_T - P_C)} = 2{,}5\sqrt{\frac{1}{N}}$$



to obtain 80% power for a 95% confidence interval, the true effect size must be at least 2.8 standard errors from zero (assuming a normal distribution for estimation error).

The top curve shows that the estimate must be at least 1.96 standard errors from zero for the 95% interval to be entirely positive.

The bottom curve shows the distribution of the parameter estimates that might occur, if the true effect size is 2.8. Under this assumption, there is an 80% probability that the estimate will exceed 1.96.

The two curves together show that the lower curve must be centered all the way at 2.8 to get an 80% probability that the 95% interval will be entirely positive.

Reference Values

P(1-P) N	0,5/0,5	0,3/0,7	0,1/0,9
100	25,0%	27,3%	41,7%
400	12,5%	13,6%	20,8%
1000	7,9%	8,6%	13,2%

These are the reference values for a binary variable (the simplest case).
In the case of continuous variables, impacts are measured as a standardized mean difference or “effect size,”

How to improve precision

Enlarge sample size

Increase the effect size

Use other covariates

Thus the value of the MDE crucially depends on the standard error of the regression coefficient.

$$y = \alpha + \beta D + u$$

To lower the standard error and gain efficiency one could think of conditioning for additional regressors·

$$y = \alpha + \beta D + \gamma X + u$$

Stratification: to block or stratify experimental sample members by some combination of their baseline characteristics, and then randomize within each block or stratum.

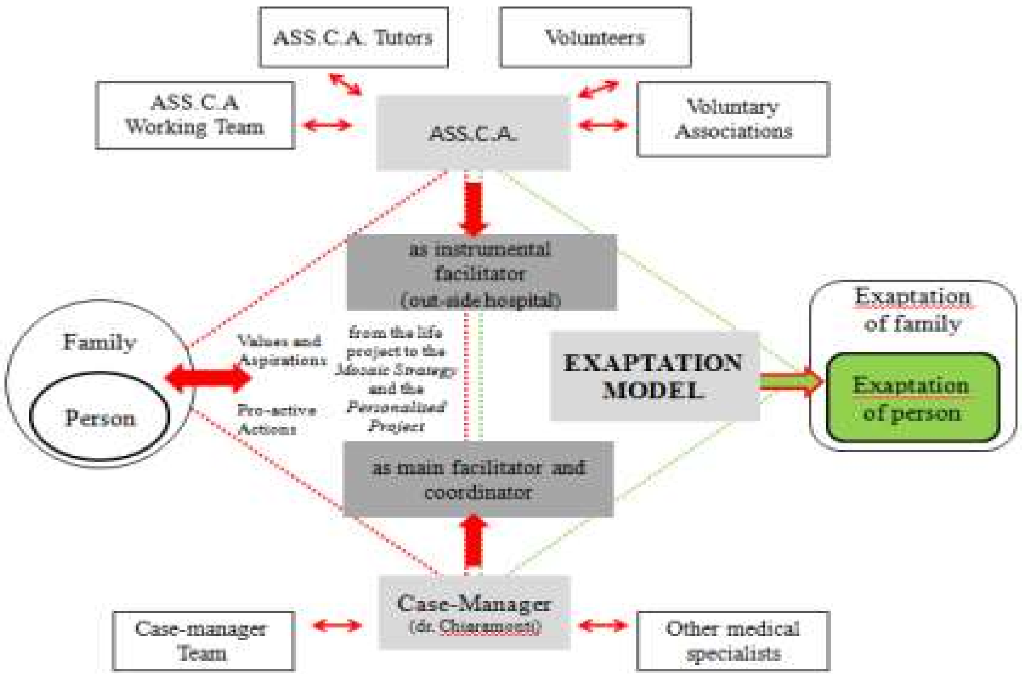
In two-stage cluster sampling, a sample is performed within each sampled cluster.

Rule of the thumb: take a lot of clusters with few observation each

Exploit intra class correlation

Case 1 – Brain Injuries

Evaluate the effectivness and the impact of a social and health care integrated intervention for persons withbrain injuries



Challenges…

To find a control group→ structure of the questionnaire to link it with other datasets

The treatment is not homogeneous→ quali quantitative and preliminary analysis

Sample size–> which elaborations?

The Research: Tools (iv)

- Structured Focus Group Discussion and party numbers

Dimension Opportunity Functioning	BENCHMARK: Opportunity for a person, age 25-50, living in the Province of Florence (1 – 10) (1=no opportunity; 10= the highest opportunity)	Opportunity for a person, age 25-50, living in the Province of Florence with a medium physical and cognitive disability, supported by familiars, who is treated with the *Exaptation Model.* (1 – 10) (1=no opportunity; 10= the highest opportunity)	Contribution of the *Exaptation Model* in determining each level of opportunity. Does it contribute? Positively or negatively? How much? (1-10) (1=very low contribution; 10=very high contribution)	Opportunity for a person with ABI outcomes, age 25-50, living in the Province of Florence, with a medium physical and cognitive disability, supported by familiars, who is not treated with the *Exaptation Model* (1 – 10) (1=no opportunity; 10= the highest opportunity)	Opportunity for a person with ABI outcomes, age 25-50, living in the Province of Florence, with a medium physical and cognitive disability, who is treated with the *Exaptation Model* but not supported by familiars (1 – 10) (1=no opportunity; 10= the highest opportunity)
(a) Physical Health	7/8	7	+ 9	5	3
(b) Mental Health	6	7	+ 10	4/5	2
(c) Work	5	4(6)	+ 7	3	2
(f) Interpersonal relations	8	6	+ 7	3	2

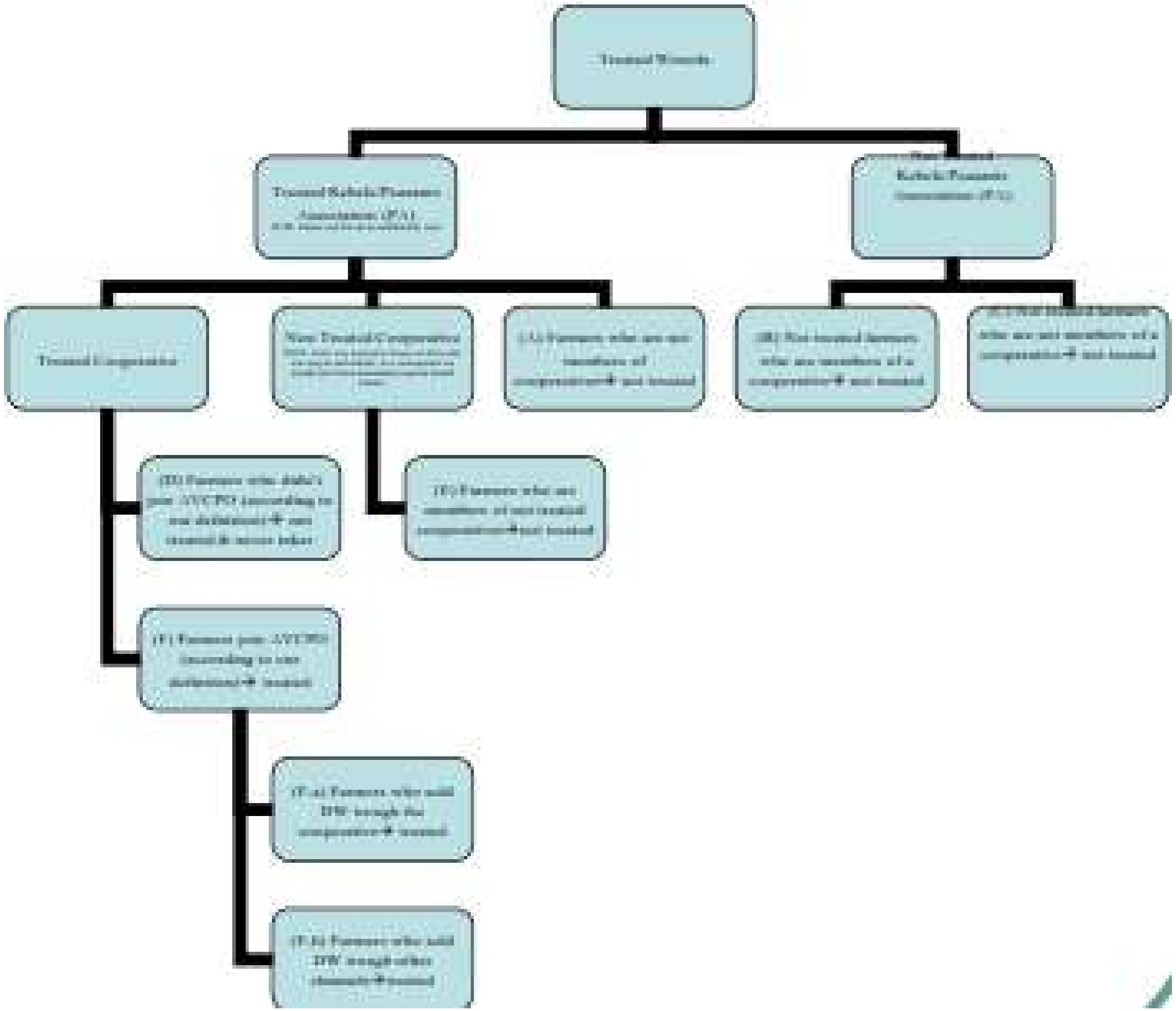
AVCPO

- Main Objective: to directly link organised smallholders to food industry (DW→pasta) Funds: MAE→IAO
- Challenge: smallholders needs to provide the right quantity and quality with the right timing
    - Seed value chain
    - Cooperatives
    - Quality assessment
- How to do it?→ economic incentive

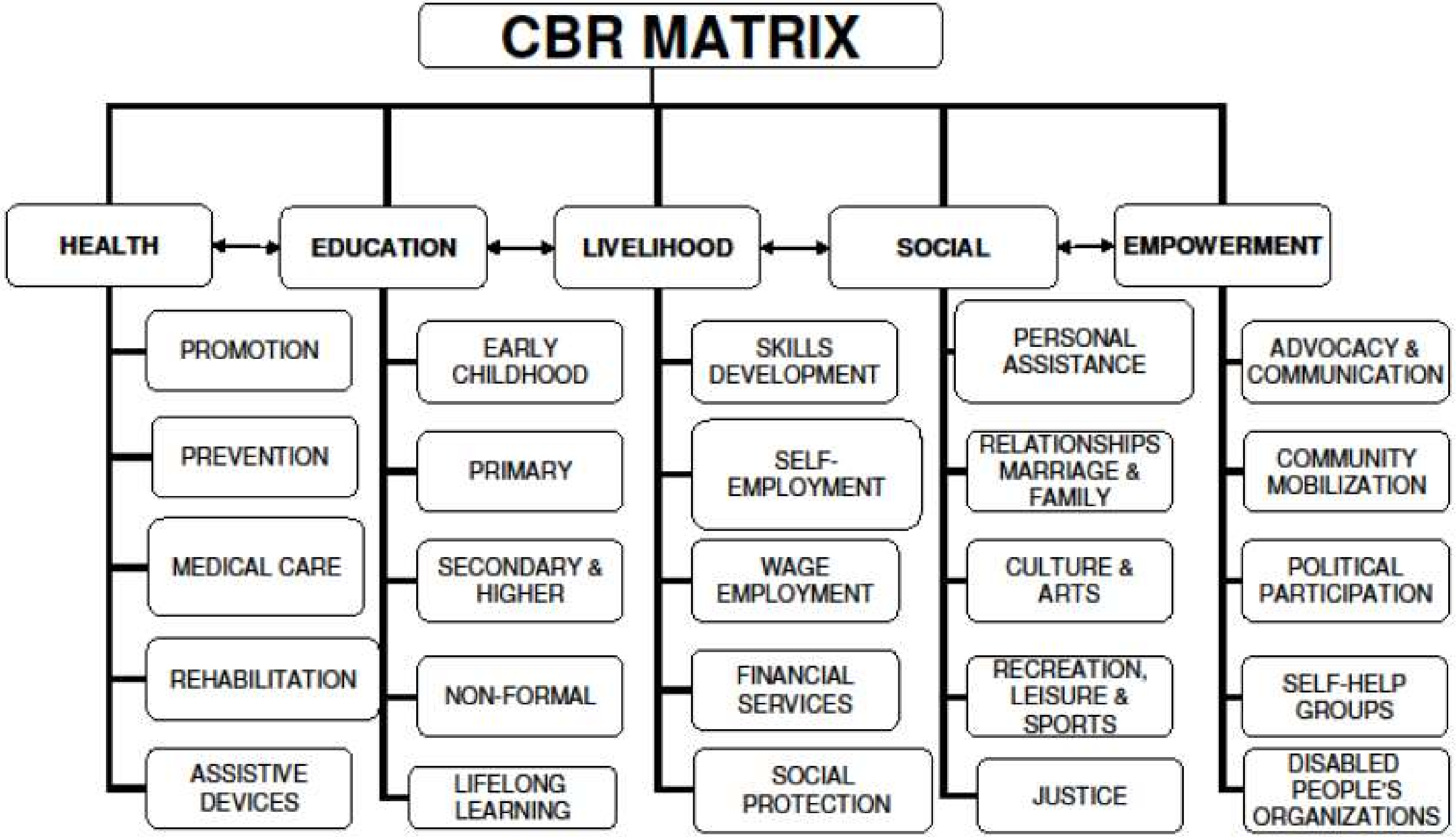
1000 ETB – 650 ETB = 350 ETB

Sample size? At first calculated with standard power analysis tecniques (clustered sampling) but…

Once on the field…



CBR Impact Evaluation in India



A lot of information available ex ante

The CBR project managed by SRMAB (Sri Raman Maharishi Academy for Blind) called Malavalli Project was initiated in 1997 in 25 villages, now about 1300 villages spread over 5 taluks (sub-districts) with around 11,000 persons with disabilities belonging to all the different groups of disabilities.

The CBR project managed by MOB (Maria Olivia Bonaldo) called Mandya Project was initiated in 1998 in 4 villages, now 1200 villages spread over 4 sub-districts and reaches about 9,000 persons with disabilities.

Two-stage Cluster Sampling

using available information to stratify the villages and improve efficiency

Sampling scheme

Total: 9x2x10=180

25 non-empty strata

Sample size
CBR areas
= treated villages

strata2	Freq.	Percent
chen-big-0	6	2.53
chen-small-0	11	4.64
krpt-big-0	5	2.11
krpt-small-0	23	9.70
madd-big-0	9	3.80
madd-small-0	13	5.49
mala-big-0	4	1.69
mala-big-1	13	5.49
mala-small-0	8	3.38
mala-small-1	24	10.13
mndy-big-0	6	2.53
mndy-big-1	16	6.75
mndy-small-0	13	5.49
mndy-small-1	11	4.64
naga-big-0	3	1.27
naga-small-0	31	13.08
pand-big-0	6	2.53
pand-big-1	1	0.42
pand-small-0	9	3.80
pand-small-1	1	0.42
ramn-big-0	3	1.27
ramn-small-0	11	4.64
srir-big-0	4	1.69
srir-big-1	2	0.84
srir-small-0	4	1.69
Total	237	100.00

Suggested readings

IMPACT EVALUATION

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.

POWER ANALYSIS

Cleves, M. (2008). *An introduction to survival analysis using Stata*. Stata Press. (Chap 16)

Hayes, R. J., & Bennett, S. (1999). Simple sample size calculation for cluster-randomized trials. *International journal of epidemiology*, *28*(2), 319-326.

CBR

Mauro, V., Biggeri, M., Deepak, S., & Trani, J. F. (2014). The effectiveness of community-based rehabilitation programmes: an impact evaluation of a quasi-randomised trial. *Journal of epidemiology and community health*, *68*(11), 1102-1108.

To deepen the topic…