Should the Ministry hire Econometricians to Learn whether what it has Done Works?

Balázs Váradi

Kyiv School of Economics
Kyiv, November 9th, 2018

INTRODUCTION

- Language
- Ukraine + evidence-informed policy making under liberal democracy
- Practical considerations on policy evaluation for civil servants in Ukraine
  - Based on my background in Academia/civil service/independent evaluation (at the Budapest Institute for Policy Analysis)
- Evidence-informed policy making

TABLE OF CONTENTS

BENEFITS OF EVALUATION

- Costly luxury…
  - E.g. in the case of a one-shot policy intervention
  - Status symbol
- a tactical device…
  - E.g. Hungarian Cohesion Policy Evaluations in preparation of the 2013-19 budget cycle
- …or *sine qua non* of evidence-based policy making?
  - Could help formulat future policies
  - Way to
    - Learn from past mistakes (also for other departments, agencies, countries
    - Improve policy instruments

THE USE OF EVALUATION

- General learning
- Affecting the policy area
  - Changes or even termination of policies
- Without evaluation, policies can survive their original goals indefinitely
- E.g. free milk to pupils in Hungary

POLITICAL COSTS AND CONDITIONS OF EVAL.

- Time horizons
  - E.g. the fate of our big anti-poverty scheme for Hungarian ex-PM Bajnai in 2011-2012
- Evaluation could be critical
  - Secret, (partially) leaked, or public
  - E.g. our evaluation of the Way-out micro-credit program
- …or *sine qua non* of evidence-based policy making?
  - Could help formulat future policies
  - Way to
    - Learn from past mistakes (also for other departments, agencies, countries
    - Improve policy instruments

TYPES OF EVALUATION

- „Political evaluation"
  - Democracies: at the ballot box: „hypothesis of economic voting"
  - Infrequent, bundled, noisy, distorted and subjective
- „Judicial evaluation"
  - Mostly in liberal democracies
  - Courts can have the power to judge, even to cancel policy programs
  - Often concerns procedures, not outcomes
- Administrative evaluation

►*What we discuss here*

WHO SHOULD DO IT?

- Civil Service
  - Make sure administrative evaluation takes place
  - Should belong to a different unit to be impartial
  - Often specialized units
  - Policy *learning* depends on the properties of the civil service
  - E.g. Previous studies on the same topic under P.M. Gyurcsány
  - External evaluation → need for civil service unit to procure
    - Terms of Reference is key
    - No answers to questions not asked!
    - No good answers to unanswerable questions!
    - E.g. Service Integration study for the EC DG Employment

WHO SHOULD DO IT? (CONT.)

- Think tanks, independent evaluators (incl. Academia)
  - Specialized expertise could be needed
  - Might be good at the scientific techniques of evaluation
  - Might have ideological preconceptions
  - Might want to keep the government's business
  - Might be partial to not the best suited methods
- The role of the public and the media
  - Role in *political* evaluation
  - Will take up administrative evaluation (often in a sensational or distorted form)
  - If no evaluation, judgement could still form
  - E.g. Kézdi-Surányi's and our evaluation of school integration of the Roma in Hungary for the Ministry of Education in 2008/10

TYPES OF ADMINISTRATIVE EVALUATION

- Three kinds of evaluation, depending on its place in the policy cycle:
- *Ex ante*: before it starts
  - More properly a part of policy formulation
  - Concentrates on:
    - Consistency
    - Coherence
    - The links between goals and instruments (relevance)
- E.g. Evaluating Hungarian spending plans for the next EU budget cycle

TYPES OF ADMINISTRATIVE EVALUATION (CONT.)

- *Interim/midway*: while it is implemented
  - More properly a part of implementation
  - Concentrates on minor changes in implementation
  - E.g. our micro-credit evaluation for Kiút Programme
- *Ex post*: after implementation
  - Tries to judge the issues like the results and the efficiency of the program
  - Piloting big interventions if possible

  ►*What we concentrate on mostly*

EX POST ADMINISTRATIVE EVALUATION

- Narrow or broad
  - Ouput, outcome, result
- Different questions requiring different methods:
  - Do the stated policy goals solve the social problem they are meant to solve?
  - Are the procedures well designed? Do they work as planned?
- Often compares inputs and outputs
  - E.g. How much money invested in training produces how many retrained workers?

TYPES OF EVALUATIONS

Types of evaluations (from Rossi, Lipsey, Freeman: Evaluation, a Systematic Approach)	**Questions addressed**
Needs Assessment	What is the problem?
Program Theory Assessment	How, in theory, does the program fix the problem?
Process Evaluation	Does the program work as planned?
Impact Evaluation	Were its goals achieved? What is the magnitude of the effect?
Cost Effectiveness	Given magnitude and cost, how does it compare to alternatives?

EX POST ADMINISTRATIVE EVALUATION (CONT.)

- Different questions requiring different methods (cont.):
  - Do they produce the planned outcomes?
  - How efficiently?
  - Do the outcomes deliver the social results of the program?
  - How effectively?
  - E.g. Healthcare Instrument Cataster Failure, Hungary, 2012

APPROACHES AND METHODS

- Many methods are used during ex post evaluation
- There is no one ideal best method, because of the diversity of:
  - Policy programs
    - E.g. school reform, military reform or building a dam
  - Timeframes
    - E.g. change in government
  - Data availability
    - E.g. lack of cohesion spending evaluation on many ALMP training efforts in Hungary
  - Scales: evaluation costs should be in proportion to the costs of the policy measure
    - E.g. rationalizing waste collection in a village
  - Questions asked (as discussed above)

APPROACHES AND METHODS (CONT.)

A dichotomy of methods:

- Quantitative:
  - Scientific methods relying on Mathematics
  - Based on numbers, experiments or quasi-experiments
- Qualitative:
  - Closer to soft social sciences
  - Based on interviews, documents, the evaluators' own experience
- Often a combination of both are needed
  - E.g. Roma integration in Hungarian Schools

CASE STUDY: EVALUATING MICROFINANCE

- Public problem/challenge: helping poor potential entrepreneurs to exit poverty
- Scope: Social and development policy, India, 2006-2008
- Main measures/steps: launching microfinance and evaluating it with a randomized study
- Implementation: calculations
- Outcomes & result: success but no long-run effects
- Lessons: even the most celebrated measure can prove worse than expected on rigorous evaluation

CASE STUDY: INTEGRATED EDUCATION OF ROMA CHILDREN

- Public problem/challenge: Public and school resistence to teaching Roma and non-Roma together
- Scope: Educaton and Social Policy, Hungary, 2005-
- Main measures/steps: A few schools integrated, and evaluating that with a quasi-experimental study
- Implementation: Calculations, observation
- Outcomes & result: Integration is good, even for the non-Roma
- Lessons: Evaluation can decide ideological debates

PROCESS EVALUATION

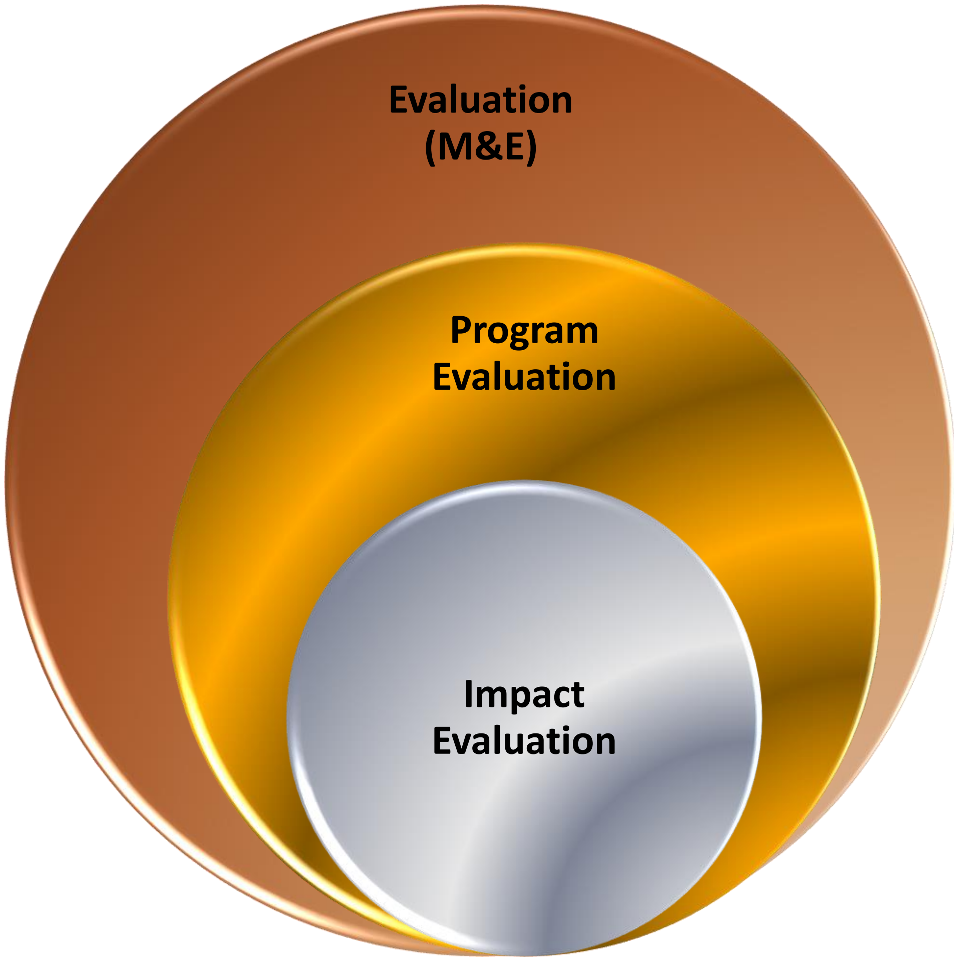
- Are basic tasks being completed?
- Are the services being delivered?
- Is the intervention reaching the target population?
- Is the intervention being completed well or efficiently and to the beneficiaries' satisfaction?

HOW DOES IMPACT EVAL. DIFFER FROM PROCESS EVAL.?

- When we answer a process question, we need to describe what happened.

- When we answer an impact question, we need to compare what happened to what would have happened without the program

IMPACT EVALUATION



IMPACT EVALUATION

In impact evaluation, we are interested in the following questions :

- *Is there an impact?*
- *What is the nature and magnitude of the impact?*
- *Is there heterogeneity?*

DIFFICULTY OF IMPACT EVALUATION

- Want to know what would have happened in the <u>absence</u> of the program (X, the “counterfactual”):

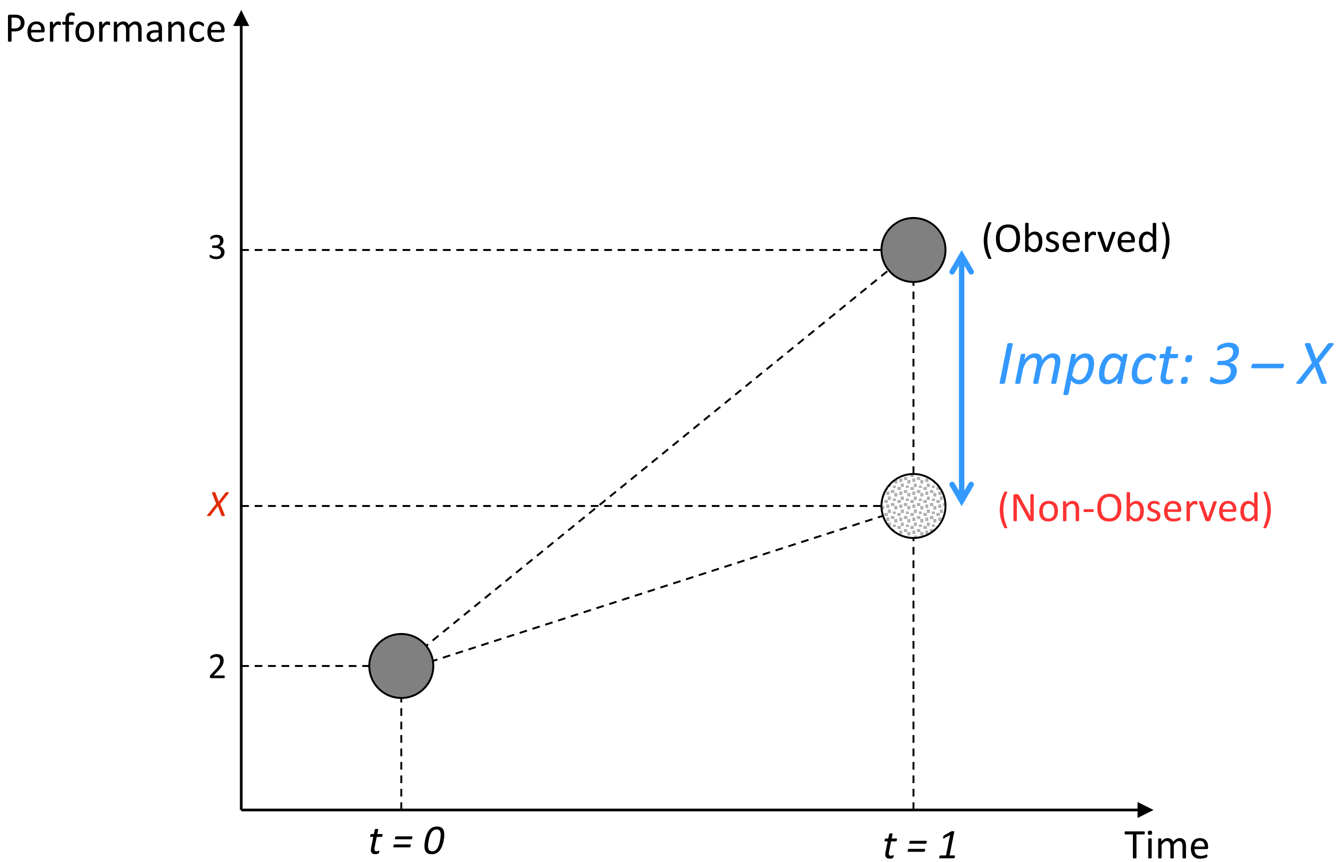
Impact is defined as a comparison between:

1. the outcome some time after the program has been introduced
2. the outcome at that same point in time had the program not been introduced (the “*counterfactual”*)

THE COUNTERFACTUAL BASELINE

- A crucial issue of ontological depth for impact evaluations is the definition of the counterfactual baseline:
  - What do we compare results to?
  - What if the program had not been introduced?
  - The dangers of naïve before-after comparison
  - E.g. employment policy interventions programs concurrent with the economic downturn of 2008

WHAT WOULD HAVE HAPPENED IN ABSENCE OF THE PROGRAM?



DIFFICULTY OF IMPACT EVALUATION

- But how is it possible to know what would have happened?
  - The program exists or it does not
  - Impossible to observe the two “states” at the same time
  - Thus, fundamental problem of evaluation is a problem of missing data
  - Impact evaluation is correct only if the estimation of X is correct
- The “art” of impact evaluation is to reconstruct X correctly

Solution: We need to “mimic” or construct the counterfactual

INDIVIDUAL VS POPULATION OUTCOMES

- Some reforms seek to influence the behaviour of participants to the scheme = individual-level outcomes
- Other reforms aim to modify the size and composition of the inflow into a scheme = population-level outcomes
- Very difficult to measure population-level outcomes and individual-level outcomes simultaneously

INDIVIDUAL-LEVEL COUNTERFACTUALS

- To estimate the impact of an intervention on individual-level outcomes (e.g., labour-market participation, net income, benefit duration), one needs to build individual-level counterfactuals.
- Individual-level counterfactuals amount to comparing beneficiaries of a new intervention with beneficiaries of existing provisions.

COST-EFFECTIVENESS

- Individual-level counterfactuals can also be used to estimate the impact of different aspects of the reform and to elicit the most cost-effective ones => highly valuable from a policy perspective
- Besides, individual-level counterfactuals help estimate the heterogeneity of impact on different sub-populations => highly relevant to some policies, such as those aiming to activate minimum-income recipients

POPULATION-LEVEL COUNTERFACTUALS

- To estimate the impact of the reform on the inflow into the scheme, one needs to build population-level counterfactuals
- Two similar groups are compared: one is given access to the new scheme, whereas the other continues benefiting from the existing scheme
- The difference in application rates, enrolment and characteristics of individuals entering the two schemes provides a measure of the change in inflow size and composition
- Can be difficult to undertake, especially when service provision is fragmented between different organisations or departments

MANY STATISTICAL METHODS…

- Non-experimental
  - Before-after
  - Difference in differences
  - Regression
  - Statistical Matching
- Quasi-experimental
  - Instrumental Variables
  - Regression Discontinuity Design
- Experimental:
  - Randomized Controlled Trials (RCTs)
  - As practiced in medical research
  - The „gold standard"

WHERE CAN WE GET DATA?

	The good . .	. . and the bad
Administrative data • Collected by a government or similar body as part of operations	• May already be collected and thus free • Can be extremely accurate (e.g. electricity bills)	• May not exist or not answer the question you want • May itself change behavior (e.g. taxes) • Can be biased (healthcare reimbursement mechanisms)
Other secondary data • Collected for research or other purposes not admin	• May already be collected and thus free • Can inform the larger context of a project	• May not exist or not answer the question you want • Dubious quality
Primary data • Collected by researchers for study	• Address the exact question of interest • Cover channels and assumptions	• Very costly and time consuming • May be biased

WHY NEED LINKED ADMIN DATA?

What is linked administrative data?

- Data on the same individual kept in various official records, linked at the individual level, but anonymised for research and policy analysis

Why need such data?

- Many advantages over survey data
- Cost of linking and anonymisation is marginal compared to the original cost of collection
- EU requirement on secondary use of public section information (PSI directive, reinforced by GDPR)

Legal barriers to compiling such datasets

LINKED ADMIN. DATA IN HUNGARY

Bureaucratic tradition

- Austro-Hungarian Monarchy (1867-1918), planned economy (1945-1989)

Statistical tradition

- household budget survey: 1949, wage survey: 1986, LFS: 1992; single personal identifier: 1978

1990s

- Constitutional Court abolishes use of unique personal identifier in 1991
- Ombudsman and new law on personal data protection in 1992
- media scandals on abuse of personal data by public administration/officials
- failed attempts to link admin data to identify free-riders of welfare system

2002-2004

- Socialist-Liberal gov elected in 2002: EU accession, NPM agenda
- research unit set up in Finance Ministry by UK graduates in 2003
- FinMin requests and receives admin data from Tax Authority (TA)
- files data requests with plan to link admin data of TA and Treasury and Health and Pension Insurance using Residence Register
- stakeholder workshop on admin data access
- failed attempt to amend personal data protection law

2005-2007

- prepare/negotiate new law on use of admin data for research and policy making
- new law on anonymization passed by Parliament in July 2007

LEGAL BARRIERS

- 1992 Law on personal data protection and public information defines extremes:

  (1) public information can be *accessed by all* and anonimised admin data is public information

  (2) data are considered personal as long as relation to the person it refers to (data subject) is *possible to restore* – with no qualifications

  cf UK rules for research data: enough to ensure that probability of re-identification is “remote”

- personal data can only be processed if approved by the person it relates to or if its use fulfils a legal obligation (ie. use for official purpose, explicitely stated by law)
- data owners have no legal basis for processing PD for the purpose of anonimisation as supporting research or statistical analysis is not legal obligation

[Note: legislation on trade secrets exists but does not prohibit anonomisation of admin data on firms]

INSTITUTIONAL/POLITICAL BARRIERS

- lack of clear protocols and earlier scandals increase risk of bad press / legal suits if processing personal data in any way „off the beaten track"
- anonimisation may (be perceived to) require a lot of staff/computer time and not acknowledged/appreciated by managers
- external users may discover shortcomings of data quality
- data owner loses monopoly for publishing (or selling) the data
- external users may discover unlawful/corrupt practices
  (e.g. award of ESF grants: very reluctant to share data on rejected applications)

- poor understanding of why researchers need individual level data

- limited knowlege of / trust in anonimisation techniques at the executive level
- low demand for evidence based policy making
- lack of trust within and between institutions *„No hidden agenda?"*
- few precedents of grassroot initiatives in public administration

THE IMPERFECT SOLUTION: THE LAW OF 2007

To please data owners:

- Right to request data anonimisation is restricted and varies by complexity of request
- But anonymised data can be accessed by all

To please Privacy and Legal concerns:

- Complicated linking procedure based on irreversible identifiers (hash codes: unique but cannot be traced back to original)
- One agency responsible for linking data
  - keep track of accummulation of data by owners
  - searchable, public database of anonymised datasets for secondary use
- Explicit and overly strict rules of anonymisation (§7): 50% sample, max small-region
- Owner must delete both the code and the data soon after sending it to intermediary.

DATASET: EXAMPLE

A large administrative panel linking 5 sources

Institute of Economics: Admin2

- Employment status, monthly accrual days, earnings, occupation + firm's balance sheet + registration as U, benefits, program participation + social transfers + proxies of state of health + participation in and graduation from education + place of residence in 2003
- Organised as a monthly panel, 2003-2011. Over 4 million people.
- Identifiers: SSN, Firm ID, zip code
- Ongoing research covers:

Employment in MNEs and health outcomes - Unemployment, benefits and health outcomes - Plant closures and their effects on displaced workers - Incentive effects of sick pay - Impact of acquisitions on the unobserved quality of new hires – Network effects on productivity and wages

RESEARCH – AN EXAMPLE

Cseres-Gergely Zs, Á Scharle and Á Földessy (2015): Evaluating the impact of a well-targeted wage subsidy using administrative data, *Budapest Working Papers on the Labour Market*, No 3

- **Subject**: impact of wage subsidy for long term unemployed jobseekers aged over 50, with at least secondary education introduced in 2007
- **Sample** drawn from ”Admin1” covering 50% of the population over 7 years. Estimation sample: 1053 men, 871 women
- **Data** from the Pension Directorate, Health Insurance Fund, and Public Employment Service (incl level of education) merged using SSN
- **Identification**: Discontinuity with a cutoff at age 50.
- **Result:** significant positive effect for men, esp with vocational secondary education
  -> improve targeting, need additional incentives for women
- **Advantage of admin data:** Ex post, can observe past + 18 months after exit, much cheaper than a survey

THANK YOU FOR YOUR ATTENTION!

balazs.varadi@budapestinstitute.eu