Erling Berge
POL3507 IMPLEMENTERING OG EVALUERING AV OFFENTLEG POLITIKK

Ex Post Facto, Subobjectives, Qualitative Method

Ref.: L. B. Mohr 1995 Chapter 10-12

Literature

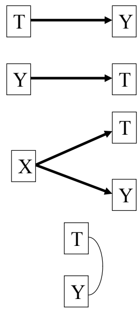
- Allison, Paul D. 2002 "Missing Data", Sage University Paper: QASS 136, London, Sage,
- Breen, Richard 1996 "Regression Models. Censored, Sample Selected, or Truncated Data", Sage University Paper: QASS 111, London, Sage
- Hamilton, Lawrence C. 1992 "Regression with graphics", Belmont, Duxbury, Kap. 1-7
- Hardy, Melissa A. 1992 "Regression with dummy variables" Sage University Paper: QASS 93, London, Sage,
- Mohr, Lawrence B. 1995 "Impact Analysis for Program Evaluation", Sage, London
- Winship, Christopher, and Robert D. Mare 1992 «Models for sample selection bias», Annual Review of Sociology, 18:327-350
- Winship, Chrisopher, and Stephen L. Morgan 1999 "The Estimation of Causal Effects from Observational Data", Annual Review of Sociology Vol 25: 659-707

Ex Post Facto Design

- This is basically about how to perform impact analysis on existing data.
- The basic characteristic of this design is the uncontrolled (decentralised) manner of assigning cases to treatment.
- The problem is usualla called the problem of self selection.

Correlation and causation

- The correlation r(T,Y) can arise in 4 ways
  - T is the cause of Y
  - Y is the cause of T
  - X is the cause of both T and Y. This is called a spurious correlation
  - Chance: The correlation arises by pure chance



Spuriousness and selection

- Two types of study
  - Treatment (often multilevel) and impact are measured for all cases in retrospect, reconstructed
  - Treatment is given to those volunteering for treatment
- Are treatment effects valid?
  - Unmeasured variables may affect both treatment and outcome measure
  - Personality characteristic making people volunteer may also affect outcome measure (Y)

Volunteering

- Many treatments are meant to be applied to volunteers. Then generalising to all subjects is irrelevant.
- Two types of effects from volunteering

  $Y_i = \alpha + \beta_1 X_{1i} + \beta_T T_i + [\beta_2 X_{2i} T_i + \beta_3 X_{3i}] + \varepsilon_i\ (Q_i)$
  - Interactive $[\beta_2 X_{2i} T_i]$ is a problem only for external validity
  - Additive $[\beta_3 X_{3i}]$ is a problem for internal validity and very difficult to counteract in an ex post facto study even extensive controls will not remove doubt

NB error in formula [10.1]: [+ should be + [

Volunteering

Advice

- Working with volunteers only: create a quasi-experimental design
- Use late volunteers as comparison group for early volunteers that just have received treatment
  - But then there may be doubt if these were volunteers at the time of observation, maybe something affetcted them to become volunteers? Then we have a possibility for spuriousness.
- Look for a criterion population as comparison group

Volunteering

- Problem of attrition causing missing data
  - People leaving the treatment group before treatment is completed will cause trouble unless a proper posttest is available
  - Without the posttest we need to impute values to them (see Allison 2002, & next 8 slides,)
    - Conventional ways of handling missing data will usually make the problem worse
    - The best case is missing at random (MAR)
    - Missing at non-random require modelling the process of attrition

Innsetjing av verdi (imputasjon)

- Målet her er å erstatte missing verdiar med rimelege gjettingar på kva verdien kunne vere før ein gjennomfører analysen som om dette var verkelege verdiar, t.d.:
  - Gjennomsnitt av valide verdiar
  - Regresjonsestimat basert på mange variablar og case med gyldige observasjonar
- I enkel imputasjon er parameterestimata er konsistente, men variansestimata er skeive (systematisk for små) og testobservatorar er for store
- Unngå om mogeleg å nytte enkel imputasjon

Oppsummering om konvensjonelle metodar for manglande data

- Vanlege metodar for korreksjon av manglande data gjer problema verre
- Ver nøye med datainnsamlinga slik at det er eit minimum av manglande data
- Prøv å samle inn data som kan hjelpe til med å modellere prosessen som fører til missing
- Der data manglar: **bruk listevis utelating** dersom ikkje **maximum likelihood eller multiple imputasjon** er tilgjengeleg

Nye metodar for ignorerbare manglande data (MAR data): Maximum Likelihood (ML) -1

- I det generelle tilfellet av manglande data finst det to tilnærmingar
  - Expectation maximization (EM) metoden er ein tostegsmetode der ein startar med ein forventa verdi på dei manglande data som vert nytta til å estimere parametrar som igjen vert nytta til å gi betre gjetting på forventa verdi som igjen … (like Iterated Reweighted Least Squares in Hamilton)
  - EM metoden gir skeive estimat av standardfeil
  - Direkte ML estimat er betre (men er tilgjengeleg berre for lineære og log-lineære modellar)

Nye metodar for ignorerbare manglande data (MAR data): Maximum Likelihood (ML) -2

- Konklusjonar om ML
  - Baserer seg på sannsynet for å observere nett dei variabelverdiane vi har funne i utvalet
  - ML gir optimale parameterestimat i store utval når data er MAR
  - Men ML krev ein modell for den felles fordelinga av alle variablane i utvalet som manglar data, og den er vanskeleg å bruke for mange typar modellar

Nye metodar for ignorerbare manglande data (MAR data): Multippel Imputasjon (MI) -1

- MI har dei same optimale eigenskapane som ML, kan brukast på alle slags data og med alle slags modellar, og kan i prinsippet utførast med vanleg analyseverktøy
- Bruken av MI kan vere temmeleg krokete slik at det er lett å gjere feil. Og sjølv om det vert gjort rett vil ein aldri få same resultat to gonger på grunn av bruken av ein tilfeldig komponent i gjettinga (imputasjonen)

Nye metodar for ignorerbare manglande data (MAR data): Multippel Imputasjon (MI) -2

- MI krev ein modell som kan nyttast til å gjette på manglande data. Som regel er det føresetnad om normalfordelte variablar og lineære samband. Men modellar kan lagast særskilt for kvart problem
- MI kan ikkje handtere interaksjon
- MI modellen bør ha med alle variablane i analysemodellen (også avhengig variabel)
- MI fungerer berre for måleskalavariable. Tar ein med nominalskalavariable trengst spesiell programvare

Nye metodar for ignorerbare manglande data (MAR data): Multippel Imputasjon (MI) -3

- Konklusjonar om MI
  - Baserer seg på ein tilfeldig komponent som vert lagt til estimat av dei einskilde manglande opplysningane
  - Har like gode eigenskapar som ML og er enklare å implementere for alle slags modellar.
  - Men den gir ulike resultat for kvar gong den blir brukt

Data som manglar systematisk

- Non-ignorable missing data
- Krev som regel ein modell av korleis fråfallet oppstår
- ML og MI tilnærmingane kan framleis nyttast, men med mye strengare restriksjonar og resultata er svært sensitive for brot på føresetnadene

Oppsummering

- Dersom nok data vert igjen er listevis utelating den enklaste løysinga
- Dersom listevis utelating ikkje fungerer bør ein freiste med multippel imputasjon
- Dersom ein har mistanke om at data ikkje er MAR må ein lage ein modell for prosessen som skaper missing. Denne kan eventuelt nyttast saman med ML eller MI. Gode resultat krev at modellen for missing er korrekt

Contamination and time order

- Contamination: Uncertainty about who actually got the treatment (e.g. persons from the control group getting treatment)
- Time order of T and Y for example in measurements based on recall
- Treatment starting before the actual pretest

Reality: $\boxed{T} \rightarrow \boxed{X} \rightarrow \boxed{Y}$

Model: $\boxed{X} \rightarrow \boxed{T} \rightarrow \boxed{Y}$

High r(T,Y) would in the model be interpreted as spurious after control for X

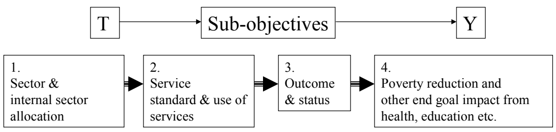
Conclusion

- Avoid ex post facto designs!
- Alas, it is impossible or inappropriate to transfer most public policy into experimental or quasi experimental designs
- So we do as best we can
- My personal advice is to use theory to bolster the data analysis, detailed elaboration of theory is the best aid for interpreting treatment effects

Case of impact analysis

- *Bjørn K. Wold, Mercy Kanyuka, Estrellita Rauan, Malawi Yute, Medson Mkwemba, Stein Opdahl and Randi Johannessen 2005*
- **Tracking Resource and Policy Impact in Malawi.** Incorporating Malawi Poverty Reduction Strategy Paper Indicators, Millennium Development Goals & Poverty Monitoring Across Sectors. Report 27/2005.
- Statistics Norway, Oslo, and National Statistical Office, Zomba

Tracking Resource and Policy Impact on Poverty Reduction in Malawi



Sector expenditure:
- Health sector
- Education sector
- Water and sanitation sector

Poverty indicator:
- GDP per capita

Tracking Resource and Policy Impacton Poverty Reduction in Malawi: water and sanitation sector

Figure 4.7. Input to water and sanitation sector, Malawi, 1990-2000
Expenditure per capita and as %

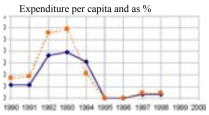


Figure 4.9. Outcome in water and sanitation sector, Malawi, 1990-2000
Diarrhoea Incidence

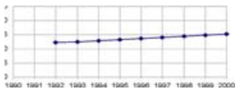


Figure 4.8. Output in water and sanitation sector, Malawi, 1990-2000
(top) Share of population with access to improved (/safe) sanitation
(bottom) Share of population with access to improved (/ safe) water

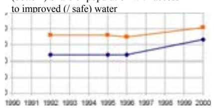


Figure 4.10. GDP per capita, PPP (current international $), Malawi, 1990-2000

Subobjectives, causation and the qualitative method

- Subobjectives permits modelling a causal process so that causal proximity may help validate the treatment inference
- Two types of ”causal” links
  - Physical causality
  - Factual causality (based on belief in the counterfactual)
- The qualitative method do not rely on any inference about the counterfactual: it relies on establishing with high probability a physical cause

Subobjectives

- Subobjective (S) are inserted between T and Y to validate the theory linking them. If the predicted links are found we are more confident in our theory
  - Causal proximity and size of impact are information that may increase confidence
- T → S → Y
- If S casues Y, then we need to finda a T that affects S

Subobjectives

- Model: T → S → Y
- Analyse by recursive structural equations:

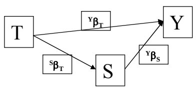
$$Y_i = {}^{Y}\alpha + {}^{Y}\beta_S S_i + {}^{Y}\beta_T T_i + {}^{Y}\varepsilon_i$$

$$S_i = {}^{S}\alpha + {}^{S}\beta_T T_i + {}^{S}\varepsilon_i$$

Model without subobjective:

$$Y_i = {}^{*}\alpha + {}^{*}\beta_T T_i + {}^{*}\varepsilon_i$$

Decomposing correlations



$^{*}\beta_T = {}^{Y}\beta_S\, {}^{S}\beta_T + {}^{Y}\beta_T$

- This is known as path analysis
- For more information see e.g.
  - Ringdal, K. 1987 Kausalanalyse i samfunnsvitenskap, Oslo, Universitetsforlaget

Subobjectives

- If $^{*}\beta_T$ is of appreciable size then
  - The subobjective S does not have much impact on Y and/or
  - There are many more subobjectives with impact

$^{*}\beta_T = {}^{Y}\beta_S\, {}^{S}\beta_T + {}^{Y}\beta_T$

- If either $^{Y}\beta_S$ or $^{S}\beta_T$ is small the subobjectie does not help establishing confidence in the validity of the T-Y relation
- In the S-Y relation spuriousness may be a problem
- If the T-S link is causal, the S-Y link maybe made quasi experimental

Subobjective

- Subobjectives where we conclude with causal links T-S and S-Y may increase our confidence in the causal nature of T-Y relation by
  - The strength of the relation (lage β)
  - Causal proximity
    - In general for good subobjectives the causal distance is greater for T-Y than for either T-S or S-Y
- Case study from Malawi: what is the treatment for poverty and what are the subobjectives?

Qualitative analysis

- Causal inference in quantitative studies relies on the counterfactual hypothesis
- Causal inference in qualitiative analysis will explicitly not involve counterfactual reasoning
  - The modus operandi method of establishing causation
  - Different causes of Y (such as T) are assumed to have their particular signatures
    - A known mechanism linking T and Y
    - One or more additional observations that are known to occur because of T

Causality

- Physical causation (physical processes links T and Y) This is necessary for
- Factual casuation to work
  - X is a factual cause if it is included in a physical causal chain and occupies a ***necessary*** slot here
- Factual causation is related to the counterfactual theory of causation
  - If X then Y and if not-X then not-Y

Reasons as causes

- Assumes that reasons are causes of intentional behaviour
- Assumes that among several reasons there is one operative reason, the ”strongest”
- Neither the operative reason nor the factor(s) that make(s) it the strongest reason is a part of anyone’s thoughts, reasons and their strength are assumed to be entities of an unaware physiological system, the affect-object system: operative reasons are physical causes

Causal reasoning

- Factual causal reasoning
  – Starting from necessary slots in the physical causal chain from T to Y it develops an argument for if not-T then not-Y, or what would have happened in the counterfactualc case
- Physical causal reasoning
  – Looks for the physical mechanism linking T and Y
  – In cases of behaviour it must explicate the operative reason and its link to behaviour

Internal and external validity

- In physical causal reasoning we have to establish that as T occurs so do Y and all other predictable outcomes (the ”signature” of T). We also have to eleminate other competing explanations for Y, the U, V, W, ... factors that in the absence of T can cause Y
- This is a persuasive argument for T causing Y. Internal validity is established
- External validity is another issue, the qualitative method does not address this problem
- But then, all approaches has a residue of doubt
  – Quasi-experiment will be haunted by selection
  – Experiments will be haunted by contamination
  – Statistics are wrong on average at the level of significance

Multiple outcomes

- To assess if an administrative program is effective one needs to:
- *Finding* the outcome dimensions
- *Limiting* the number of outcome dimensions
- *Assessing the impact* of the program on the dimension
- *Common scaling*: Combining estimates of impacts into one performance measure
- *Weighting* different dimensions in the common scale

Alternatives to impact analysis

- Cost Benefit Analysis
  - Little attention to finding and limiting #dimensions
    - Risk of double counting is noted
  - Relies heavily on willingness to pay as measure of utility
  - Assessing impact is central
    - A minimum of data on quantities and prices to be used and
    - Economic theory to extrapolate or impute from data
  - Common scaling and weighting is handled by using monetary value as a measure of utility
  - Different groups (e.g. poor) may be given particular weight
    - Sensitivity analysis of differences in weighting is recommended

Alternatives to impact analysis

- Multiattribute utility technology (MAUT)
  - Finding is prominent. Decisions to be made must involve all stakholders defined as all that will be affected by the decision
  - Rules of limiting: avoid duplication, overlap and relatively unimporant impacts
  - Impact assessment is left to the choice of the analyst. Much is done as subjective judgements
  - Common scaling is done by getting minimum and maximum tolerable measurement scores from stakeholders and using these as anchors for all scales
  - Weighting is also done by getting stakeholders to assign weights sometimes extended by assignig weights to each group of stakeholders also

Impact analysis and multiple outcomes

- Finding: use a modified stakeholder approach
- Limiting: look to the outcome line and the relative value of including the outcome
- Impact assessment: experiment and quasi-experiment
- Common scaling should not be attmpted
- Weighting: done by each participant

Finding dimensions

- 3 categories of outcomes
  - Objectives, constraints, side effects
- 2 approaches to finding
  - Professional: the final –non-partisan - choice is up to the researcher and the conception of the common good
  - Partisan: the final choice is up to the sponsor
- The evaluator should write down all stakeholders, objectives, constraints, and side effects having any plausible effect on decsion making regarding the program

Limiting the number of dimensions

- Organise outcomes into outcome lines and pick an outcome of interest on each line
- Outcomes of interest should be order into
  - Cannot be researched
    - Because of resource constraints or lack of adequate methods
  - Need not be researched
    - Results are obvious
  - Should not be researched
    - They do not have an impact on decisons
  - Of minor interest
    - Outcomes here are unlikely to affect decisions and may be assessed subjectively
  - Of major interest

Common scaling and weighting

- Common scaling is not a priority and should be used only where it can be done with confidence in the result.
- Weighting should be left to the political process making decisions
- Leave the results of the impact analysis in their original units of measurement
- Humans are in general good at making choice among incommensurables

Utilisation

- Comprehensive presentation of outcomes
- Correct choice of outcomes to be submitted to research
- Quality of selection and execution of research design
- Constructiveness in the evaluation findings