Comparing treatments: Adjusting for missing respondents and different group characteristics

Hans Olav Melberg
(hom@sirus.no)
SIRUS
(National Institute for Alcohol and Drug Research)

Preliminary results,

Not for citation

Research funded by SIRUS and The Research Council of Norway

The problem

- Cost benefit analysis: Treatment A > B
- B complain:
  - selection bias I: There are fewer missing people/dead in our group!
  - selection bias II: Our group is more difficult to treat!
  - you use the wrong success criterion!

Example: Cost-Benefit Project at SIRUS:

- Treatment for drug abuse, 407 individuals in treatment interviewed using the Addiction Severity Index (ASI) each year for three years
- Compare:
  - Hierarchical Therapeutic Communities (HTC)
  - Psychiatric Youth Teams (PYT)
  - Large units/collectives (Large)
  - Small units/collectives (Small)
  - Also: Children and (later) methadone (Both ignored here)

Causes of the problem

- Key cause: Non-experimental data?
- Theory: Rubin Model (Holland, 1986)
  - Effects of a treatment (a) (measured by the value of Y) on a unit (u) is given by:
    - $Y_a(u) - Y_n(u)$
  - where $Y_n(u)$ represents the control i.e. non-treatment (n)
  - The fundamental problem of causal inference, Holland argues, is that we cannot observe these two values on the same unit
  - Two possible "solutions"
    - 1. Scientific solution: Use assumptions (e.g. temporal stability and causal transience)
    - 2. Experimental solution

      $T_a = E[Y_a(u) - Y_n(u)] \rightarrow T_a = E[Y_a(u)] - E[Y_n(u)]$ Stil not observable

      Assume independence, then we have an observed estimate:

      $E[Y_a] = E[Y_a|S=a]$
      - <u>Randomisation makes assumption of independence plausible</u>
      - Problem for non-experimental data: Since the units in the different treatments are not selected at random we cannot assume that the differences between groups cancel out when we take the average (not independent). Result: Bias! Estimated difference not the result of treatment, but of group selection/group composition.

Solution: Randomised experiment?

- If the cause of the problem is non-random selection of units, then the solution must be randomised experiments … or?
- Problems with this solution
  - 1. Ethical
  - 2. Practical
- Both well known; less well known: Also problems with experiments *even if practical and ethical.*
- What are the problems? See Heckman (1995)

Problems with experiments

- Heckman, JEP, 1995

1. Experiments provide little evidence on many questions of interest (e.g. median)
2. The intrinsic variability in evidence from randomised experiments

   - experiments not produce joint distribution, only marginal

3. Randomisation bias
4. Institutional limitations on social experiments

   - e.g. if voluntary then often reduced external validity

5. Substitution bias

Non-experimental solutions

- If randomised experiments are unethical, impractical or viewed as likely to be plagued by problems, then: Do as well as we can with observational data. How?
- Make assumptions about:
  - Whom to compare?
  - How to adjust (confounding variables, missing respondents)?
  - Different success criteria
- Then test these assumptions empirically (or “make plausible” using theory)
- What follows: Preliminary results. Not advanced (e.g. two stage estimation), Not complete overview or final analysis.

Whom to compare?

Average percentage change in ASI index for drug in return for 100 000 NOK

Whom?	HTC	Large	Small	PYT	Children	Total
1. No selection	11,72 %	6,70 %	8,77 %	7,38 %	3,71 %	7,82 %
2. Exclude missing	10,28 %	6,90 %	7,52 %	6,89 %	3,18 %	7,49 %
3. Exclude "controlled environment"	8,76 %	6,40 %	8,29 %	-7,09 %	3,72 %	6,86 %
4. Exclude "drop-outs"	10,82 %	7,48 %	13,52 %	172,22 %	7,62 %	10,68 %

Note: Blue = ”best”, red = ”worst”

Conclusion depends heavily on assumption about selection. For instance: PYT is both best and worst!

Number of clients in different groups

	Total	HTC	Large	Small	PYT	Children
Answers, Recruitment interview (t=0)	407	94	118	64	100	31
	100,0 %	23,1 %	29,0 %	15,7 %	24,6 %	7,6 %
Answers, 1. Follow up	364	83	103	59	88	31
	89,4 %	88,3 %	87,3 %	92,2 %	88,0 %	100,0 %
Answers, 2. follow up	342	79	98	53	85	27
	84,0 %	84,0 %	83,1 %	82,8 %	85,0 %	87,1 %
Not controlled environment, 1. Follow up	189	43	35	23	77	11
	46,4 %	45,7 %	29,7 %	35,9 %	77,0 %	35,5 %
Not controlled environment, 2. Follow up	248	56	67	35	71	19
	60,9 %	59,6 %	56,8 %	54,7 %	71,0 %	61,3 %
Completed treatment, 1. Follow up	36	9	11	9	7	0
	8,8 %	9,6 %	9,3 %	14,1 %	7,0 %	0,0 %
Completed treatment, 2. Follow up	31	4	18	6	0	3
	7,6 %	4,3 %	15,3 %	9,4 %	0,0 %	9,7 %
Completed treatment, Cumulative	67	13	29	15	7	3
	16,5 %	13,8 %	24,6 %	23,4 %	7,0 %	9,7 %
Still in treatment, 1. Follow up	141	31	50	14	23	23
	34,6 %	33,0 %	42,4 %	21,9 %	23,0 %	74,2 %
Still in treatment, 2. Follow up	52	14	19	4	5	10
	12,8 %	14,9 %	16,1 %	6,3 %	5,0 %	32,3 %
Dead before 1. Follow up	9	3	3	1	2	0
	2,2 %	3,2 %	2,5 %	1,6 %	2,0 %	0,0 %
Dead before 2. Follow up	7	0	4	1	1	1
	1,7 %	0,0 %	3,4 %	1,6 %	1,0 %	3,2 %
Dead, Cumulative	16	3	7	2	3	1
	3,9 %	3,2 %	5,9 %	3,1 %	3,0 %	3,2 %

Adjusting

- Using OLS regression to estimate response from missing
- Many problems …
  - Whom to include (again)?
  - Different estimates for different types of non-response (dead, refuse, not find)?
  - Ordinary regression problems (technical)

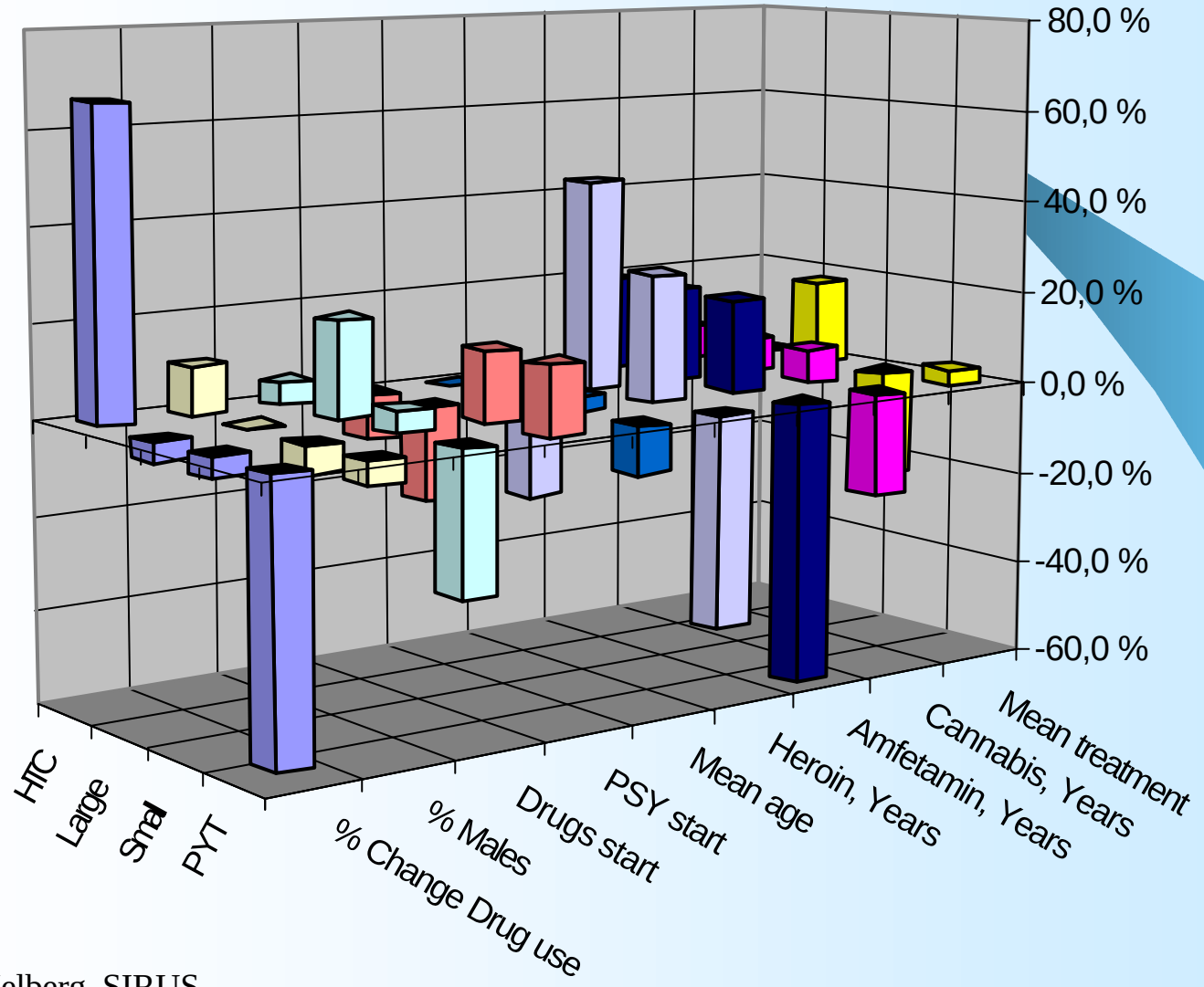
Adjust for ”confounding variables”

- Example: Treatment B seems to do worse than A, but clients in B may be ”worse” (e.g. ”heavy addicts” or ”severe psychological problems in addition to drug use problem”)
- First: Examine possible confounders

Potential confounders I: Numbers

	Total	HTC	Large	Small	PYT	Children
% Change in drug use	**38 %**	**59 %**	**34 %**	**34 %**	**15 %**	**68 %**
% Males	66,8	75,5	68,6	64,1	65	45,2
Drug problem severity (0-1)	0,31	0,33	0,38	0,33	0,22	0,25
Alcohol problem severity (0-1	0,08	0,08	0,08	0,12	0,08	0,04
Psychiatric problems (0-1)	0,28	0,26	0,23	0,33	0,33	0,31
Mean age	29	30	32	31	27	16
Years of heroin use	5	4	8	7	3	0
Amfetamin, Years	5	6	6	6	2	1
Cannabis, Years	10	11	11	11	8	2
Mean treatment, days	401	369	434	283	380	690

Potential confounders II: Visual



Adjusting

- Methods
  - Regression analysis; examine importance of other variables; eliminate this influcence from the result, then redo cost-benefit anlaysis
  - Tried this; quickly run into problems (circular, difficult to interpret, what to include)
  - Easier alternative: Divide respondents people into-sub-groups of ”similar” people and then compare treatment result
  - Which groups?

Results for some sub-groups

Average percentage change in ASI index for drug in return for 100 000 NOK

Whom?	HTC	Large	Small	PYT	Children	Total
Only herion users	13,7	8,1	12,3	3,2	2,9	9,3
n	45	70	36	28	5	184
Psychiatric problems (top 25%)	9,7	8,8	7,7	93,1	5,8	11,5
n	17	16	19	27	4	83
Less psychatric problems (bottom 25	14,2	-6,0	8,7	20,5	6,3	3,6
n	21	25	9	17	1	73
Heavy users (top 25%)	14,6	9,5	16,0	83,6	2,7	11,7
n	5	25	10	7	2	49
Less heavy users (bottom 25%)	7,0	-23,5	-10,2	-24,4	4,7	-5,8
n	11	11	9	33	7	71
Females	14,7	9,8	2,8	-7,5	4,2	6,7
n	18	25	21	25	13	102
Males	11,9	3,1	8,6	38,5	3,9	7,4
n	58	65	31	51	9	214

Note: Not explored sub-groups of sub-groups (e.g. Female heavy drug users with psychological problems compared to some other sub group)

Importance of considering sub-groups

- Treatment A may be ”best” for one sub-groups, but B for another. Hence, must know sub-groups before make recommendation of A or B
- Even worse: Simpson’s paradox
  - Treatment A is best for both sub-groups, but when pooling all individuals into one group (ignore sub-groups) treatment B is best! (Counterintuitive)
  - A possibility. Example follows on next slides (constructed i.e. not based on real data)

Simpson's Paradox I

TABLE 1				
	Cured	Not cured	Total	**Success rate**
Outpatient (Day-care)	20	20	40	**50%**
Inpatient (Residential care)	16	24	40	**40%**
Total	36	44	80	

Simpson's Paradox II

	TABLE 2							
	Easy cases				**Hard cases**			
	Cured	**Not Cured**	**Total**	**Success rate**	**Cured**	**Not cured**	**Total**	**Success rate**
Day-care Outpatient	18	12	30	**60%**	2	8	10	**20%**
Residential Inpatient	7	3	10	**70%**	9	21	30	**30%**
Total	25	15	40		11	29	40	

Changing the success criteria

- So far: Success = % change of ASI drug index
- Some alternatives
  - Cured (No longer using drugs)
  - Absolute change (not percentage)
  - No index, use ”average days”
  - Consider more than changes in drug use
    - Health (mental and physical), crime, work situation.

Results when changing success criterion

Success criterion	HTC	Large	Small	PYT	Children	Total
Cured" addicts	0,082	0,041	0,074	0,406	0,039	0,065
n	32	32	23	25	18	130
Absolute change in drug index (not %	0,034	0,025	0,026	0,016	0,009	0,023
n	76	90	52	82	24	324
Mean number of days with heroin last	2,659	2,397	2,482	3,245	0,133	2,020
n	94	118	64	100	31	407
Crime (Asi)	0,043	0,033	0,026	0,111	0,010	0,031
n	77	89	47	84	20	317
Psychiatric (Asi)	0,024	0,012	0,017	0,270	0,006	0,022
n	77	89	47	84	20	317

Conclusion

- Matrix of possible answers depending of assumptions of whom to compare, how to adjust and success criterion
- Provide whole matrix. Admit uncertainty!
- Sometimes clear results even from matrix!

Comments and answers

- Comment: Using OLS to fill in results for ”missing” does not increase statistical power
- Response: Our estimates would ideally have two properties: Ubiased and as litte variance as possible. Increasing the number of ”similar respondents” decreases variance, but it is true that if the OLS regression filling in values for missing is not based on ”new” information there is nothing to be gained (in terms of reduced variance). It is, however, possible to employ some ”new” information to adjust for missing – both to reduce variance, but – more importantly – to reduce bias. Here is an example: Imagine that the ”worst” cases fail to respond when you do the follow up. Then your estimate of the treatment effect will be biased (too optimistic). If you, as we have, record the number of phone calls and contacts you have made to find the person, this represent ”new” information that can be used to fill in for those missing and adjust for the bias. You extrapolate the trend from the clients you found after many attempts to those you did not find. Hence, reducing bias is probably the most important use of this method (OLS to fill in for missing), not to reduce variance (as I initially – but wrongly - thought).

References

- Holland, P.W. (1986) Statistics and Causal Inference. *Journal of the American Statistical Association* 81, pp. 945-970.
- Heckman, James J. and Jeffrey A. Smith (1995) Assessing the Case for Social Experiments. *Journal of Economic Perspectives* 9 (2), pp. 85-110.
- See also ch.9 about “Models with self-selectivity” in Maddala (1983) Limited dependent and qualitative variables in econometrics. (CUP)
- Many more references can be found in some draft papers I have written on the topic: Available at: http://www.geocities.com/hmelberg/papers/phd_in_progress.htm