

EU-DAP project

**Strategies of statistical analysis and
Differences in results**

Federica Vigna-Taglianti

Multilevel data structure

- Multilevel or hierarchically structured data are the norm in the social, medical and biological sciences
- School education provides a clear case of a system in which **individuals** are subject to the influences of **grouping**
 - Pupils learn in classes
 - Classes are taught within schools
 - Schools are administered within local authorities, communities, countries

Hierarchical levels

- School education: different hierarchical levels
 - Authorities-boards-countries: **level 4**
 - Schools: **level 3**
 - Classes: **level 2**
 - Pupils- students: **level 1**
- Units at one level are recognized as being nested within units at the next higher level (clustered populations)

ICC

- It is expected a positive correlation among observations taken on members of the same group
= ICC = intra class correlation
- It reflects an extra component of variance attributable to the group beyond the variance attributable to the members
- The magnitude of the ICC is used as a measure of the degree to which the scores within the group are not independent of one another (independence of obs is violated)

Critical points

- The ICC increases the variance of any group-level statistic
 - With a limited number of groups in each condition, the degree of freedom available to estimate group-level statistics are limited
- 
- Any test that ignores the extra-variation or the limited df will be inflated (often badly) producing an underestimation of standard errors of regression coefficients

Analytic strategies

- **Appropriate analytic strategies are recommended**
 - **Because of constraints inherent with curriculum delivery the randomisation is performed at the **school-level** while the statistical analysis is generally performed at the student level**
 - **However, the unit of random assignment should also be the unit of analysis**

School as unit of analysis

- School can be used as unit of analysis
- However there are some disadvantages in doing so
 - Type II error rates are much higher
 - Statistical power is reduced
 - Appropriate only when ICC=1
 - Many important processes are well defined only at the individual level (constructs make sense only at the individual level, e.g. self-esteem)

Multilevel modelling

- Provides better estimates in terms of SE
- models and investigates the relative sizes and effects of **level** characteristics (country, school, class) as well as that of **individual** characteristics
- Properly corrects for ICC (intra class correlation)

Multilevel models in school based evaluation

- Different models can be built (at least 3 grouping levels)
- Palmer-Graham-Hansen (Prev Med 1998) re-analysed their data using different multilevel models and found that
 - The model with **school** as grouping level (level 2) produced results slightly weaker than the model with **class** as grouping level

EU-DAP results (I)

- EU-DAP results are consistent (and still better) with Palmer analysis
- A program effect does exist when comparing students who received EU-DAP with those who did not
- The effects is found for analyses using
 - **individuals** as the unit of analysis
 - the multilevel model with **class** defined as level 2
 - in the multilevel model with **school** defined as level 2 results appear slightly weaker

EU-DAP results (II)

outcome	Model with 9 centers (RIGLS binomial 1st order MQL)		
	OR (95%CI) unadjusted	OR (95%CI) adjusted for center prevalence of daily smoking fixed	
		center school student	center class student
ALO smoking	0.82 (0.74-0.91)	0.88 (0.68-1.13)	0.88 (0.71-1.08)
Regular smoking	0.76 (0.67-0.88)	0.80 (0.59-1.09)	0.85 (0.65-1.10)
Daily smoking	0.67 (0.57-0.80)	0.68 (0.48-0.97)	0.74 (0.55-0.99)
ALO drunkenness	0.72 (0.62-0.84)	0.75 (0.58-0.96)	0.74 (0.60-0.92)
Regular drunkenness	0.62 (0.47-0.82)	0.66 (0.45-0.97)	0.65 (0.46-0.92)
ALO cannabis	0.68 (0.56-0.83)	0.79 (0.59-1.05)	0.77 (0.61-0.98)
Regular cannabis	0.65 (0.50-0.84)	0.81 (0.57-1.14)	0.77 (0.57-1.03)
ALO drugs	0.76 (0.64-0.89)	0.87 (0.66-1.13)	0.85 (0.67-1.09)

Conclusions (I)

- **Assuming that the multilevel analysis with class as level 2 yields the correct estimates, we can conclude that**
 - **School level analysis under-estimates the effect**
 - **Loss of power**
 - **ICC larger at the school level**
 - **Individual level analysis over-estimates the effect**

Conclusions (II)

- **There is a class effect in the model**
 - **“class level attitude”**
 - **program exposure**
 - **How the pupils received the program**
 - **How they interacted with the program**
- **The school effect is less clear**