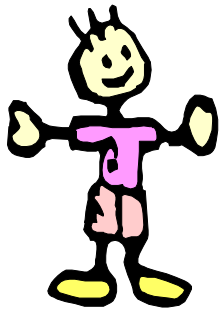
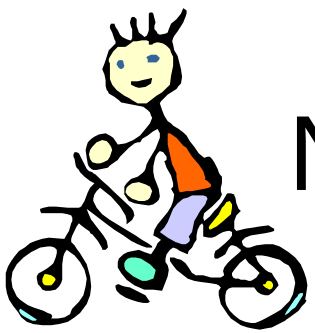
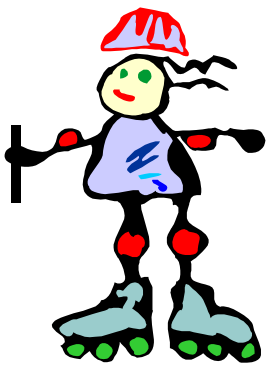
N L S C Y





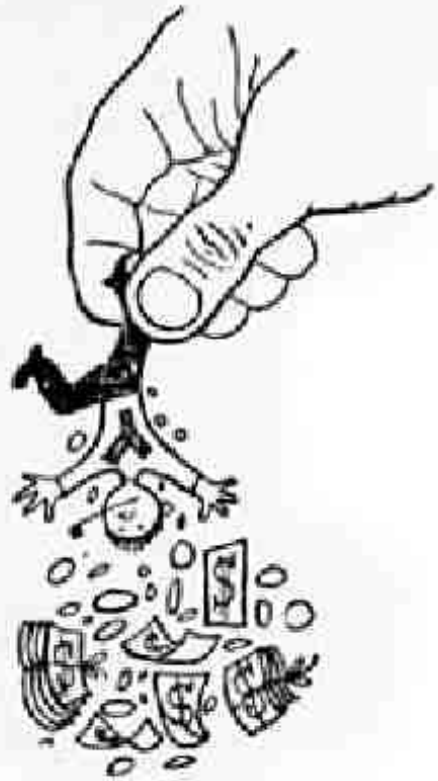
Overview of the National Longitudinal Survey of Children and Youth



Methodologist

- Not subject-matter expert
- We do the sampling thing
- Determine the sample size
- How the sample is selected
- Analysis done by the methodologist
  - •is to assess data quality
  - •adjust estimation methods accordingly

Picking people for a sample, can be an expensive thing to do.

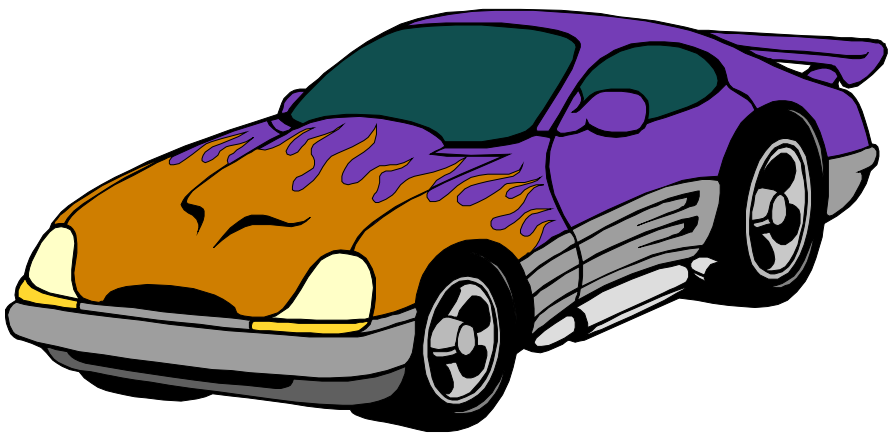


The job of a methodologist is to get the least expensive sample that will still produce the best estimates possible.

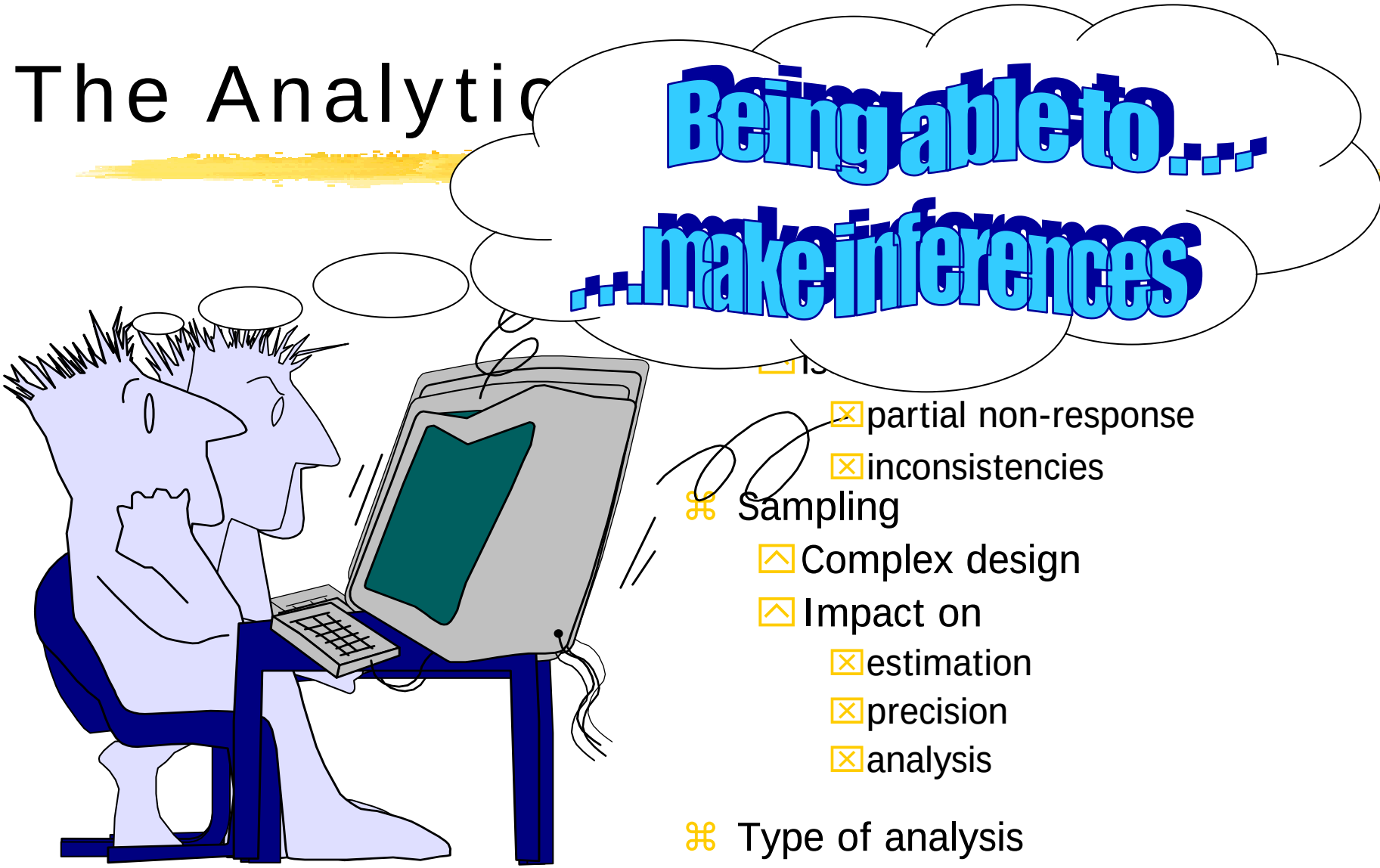
The methodologist helps determines the survey vehicle.

Goal of the Presentation

- The survey vehicle has an impact on the analysis.
- Today we're going to describe the vehicle in order to facilitate the analytical process.



The Analytic

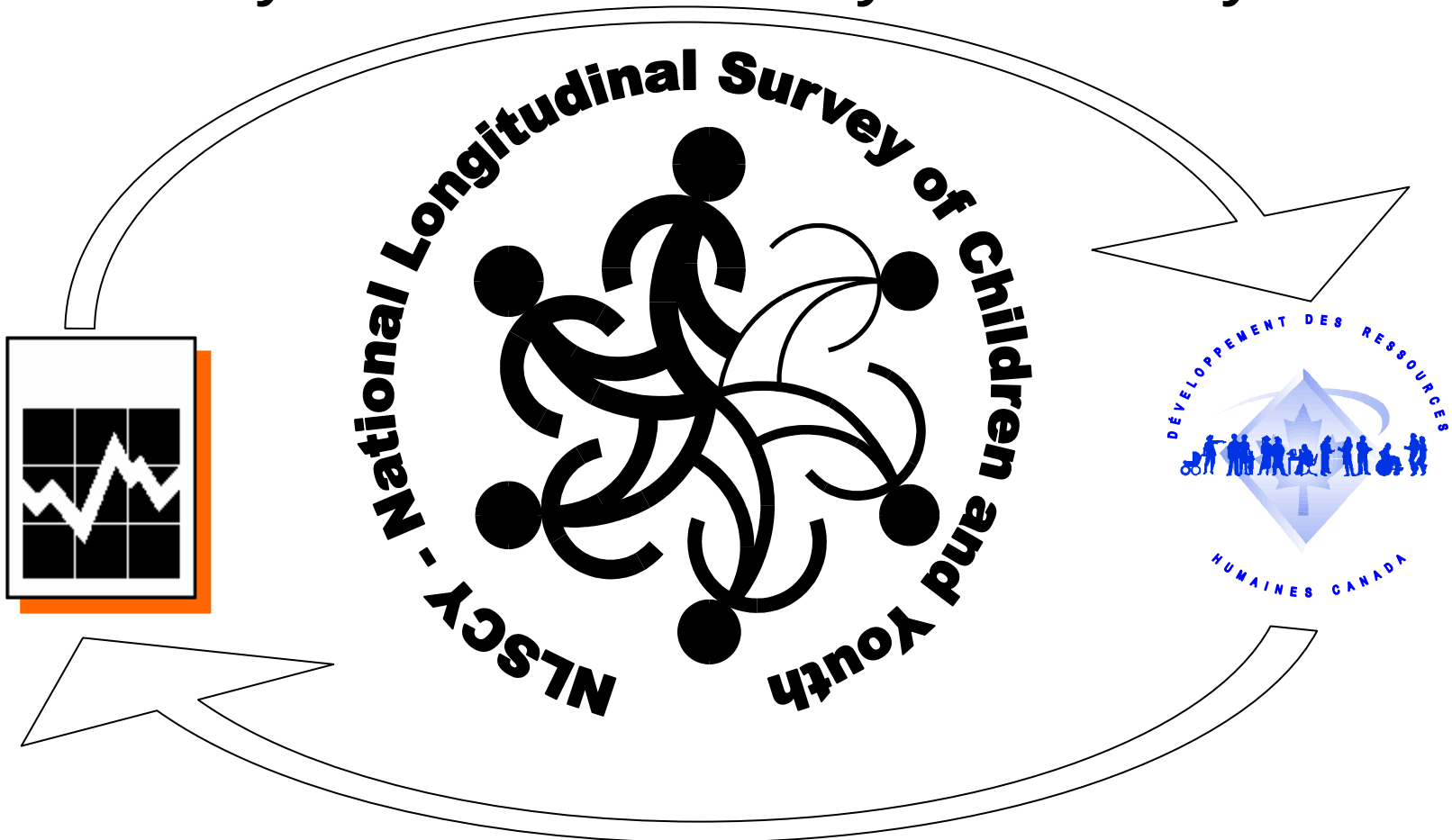


- Is
  - partial non-response
  - inconsistencies
- Sampling
  - Complex design
  - Impact on
    - estimation
    - precision
    - analysis
- Type of analysis
  - longitudinal, cross-sectional
  - other and mixed


National Longitudinal Survey of Children and Youth
NLSCY

NLSCY - Overview

The survey is almost entirely funded by HRDC



NLSCY - Overview

- Complex data structure
  - the lives of children are complex
  - dual child/household structure
  - new content in each cycle
  - some changes in old content
- Other constraints
  - limit on quantity of information
  - limited resources

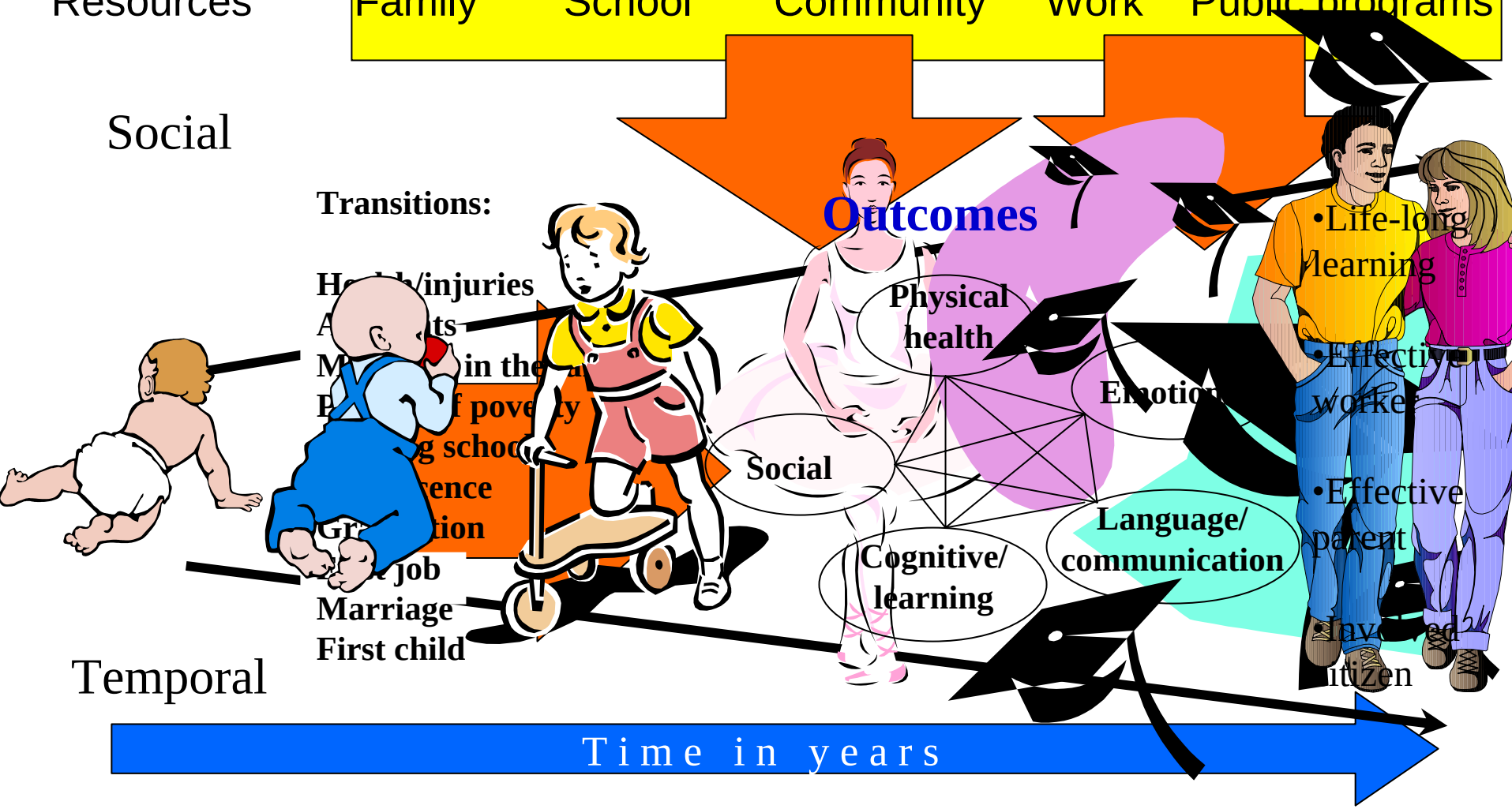
Analytical Framework

Context

Resources

Family School Community Work Public programs

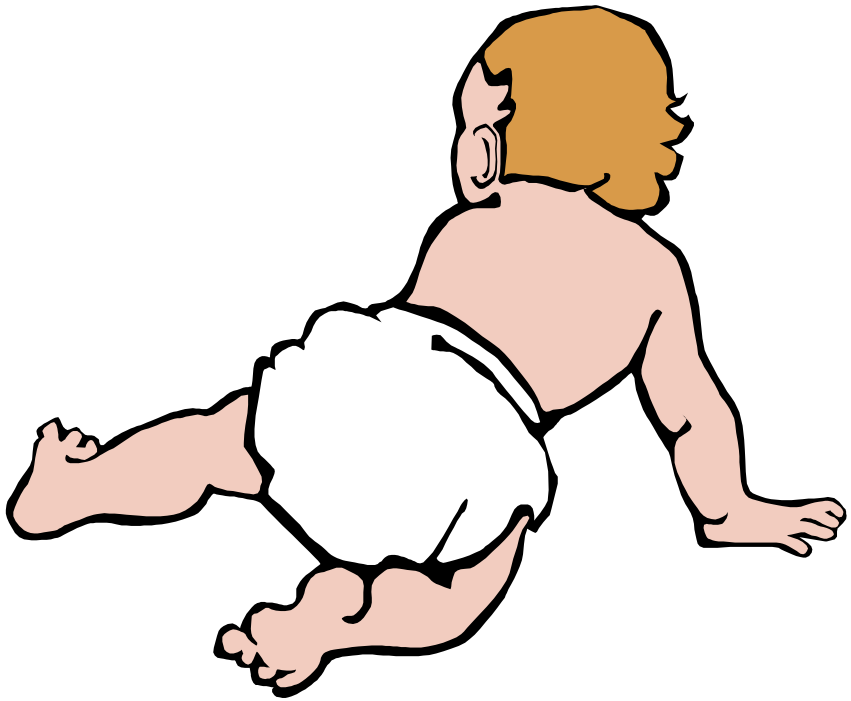
Social



Temporal

Time in years

Unit of Analysis



- The child
- Sources of information
  - Person most knowledgeable about the child (PMK)
  - Teacher
  - School principal
  - Child himself/herself
    - cognitive measures
    - self-administered

Unit of Analysis

- Caution
- Other types of Analysis
  - Weights are designed for the child
  - Concepts like family are characteristics of the child
    - Not a domain for estimation

Statements like . . .

The NLSCY estimates the number of children who live in families with a certain characteristic …

NLSCY Content

Child (depending on age)

- socio-demographic
- health
- perinatal information
- development (motor, social and physical)
- temperament
- academic performance
- education
- literacy
- extracurricular activities
- work experience
- socialization
- relationship with parents
- family history and legal custody of children
- child care
- behaviour
- self-esteem
- cigarettes, alcohol, drugs
- vocabulary assessment
- math test
- reading comprehension test
- sexual activity and loving relationship

Parents

- socio-demographic
- education/literacy
- labour market
- income
- health
- social support
- parental involvement at school
- parents' aspirations for child's education

Family

- demography of members
- relationships between members of household
- family functioning
- household
- neighbourhood

School

- number of students
- discipline problems
- school atmosphere
- resources
- characteristics

Teachers

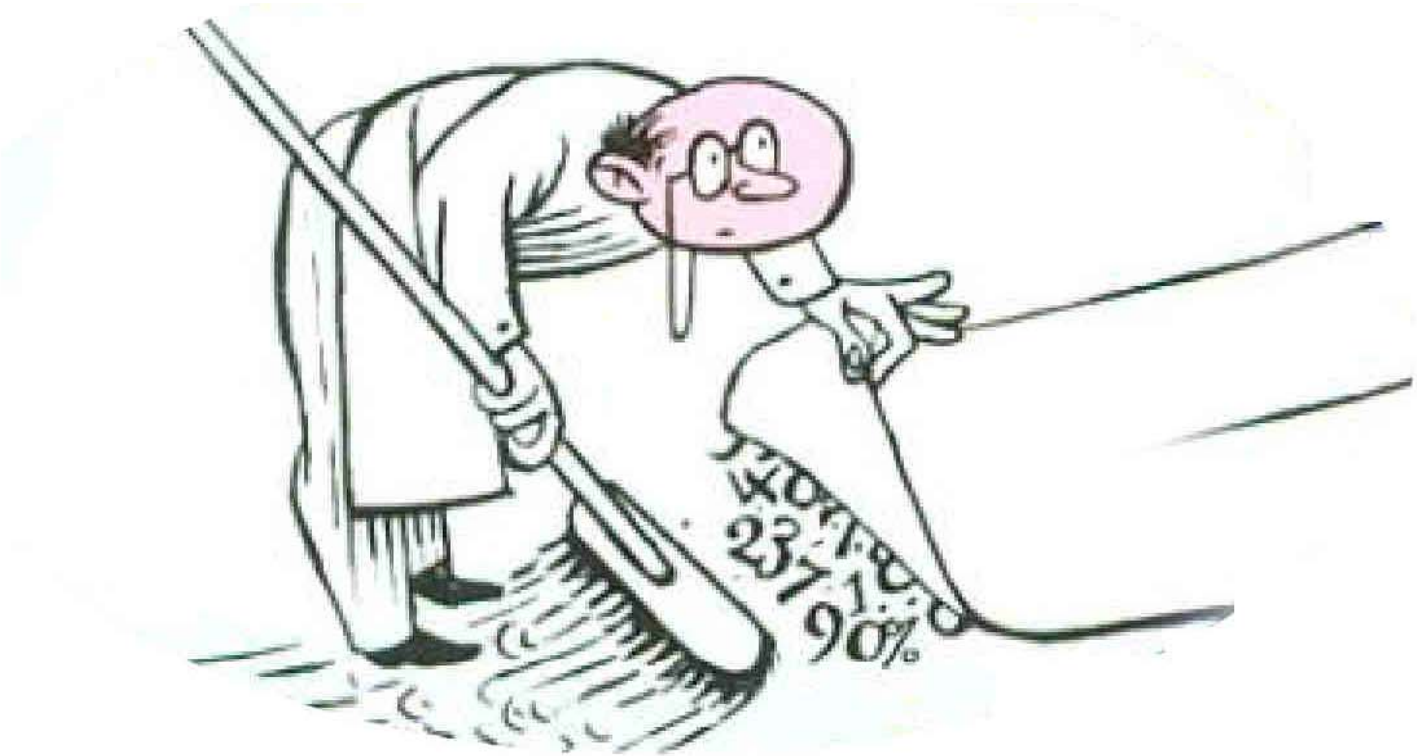
- teaching practices
- demography
- qualifications

Principal

- demography
- qualifications

Note: Minor changes are made in the content from one cycle to the next.

In the NLSCY, you will find data are less processed.

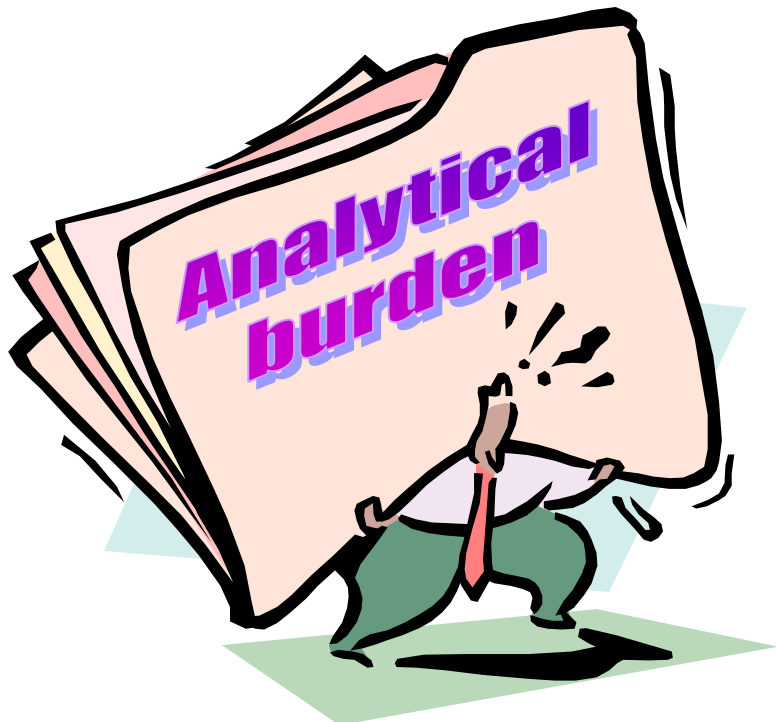


Investment for the NLSCY

- Focus on derived variables
  - scales, cognitive measures
  - transition measures
- Non-response adjustment
  - total non-response
- Processing of financial data
  - family income
  - personal income
- Dissemination within reasonable time

Data and Analysis

- Changes in some longitudinal variables
  - improving the concept
  - different respondent
    - means a conceptual difference
  - different response
    - different PMK
    - response error
- Unprocessed responses
  - verification and consistency
  - partial responses



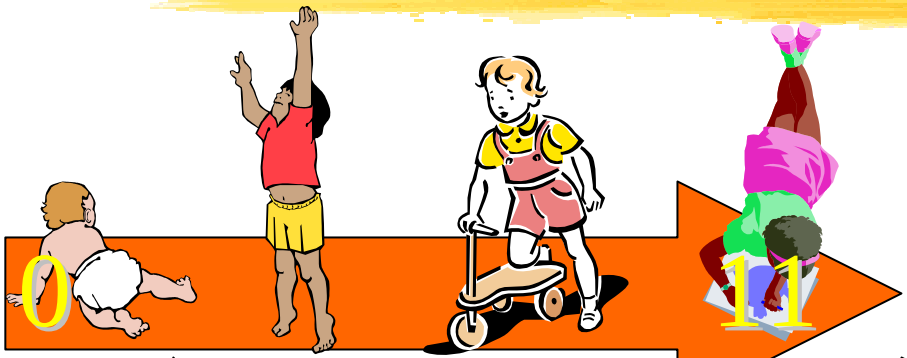
Partial Non-response

- Respondent units are those which answered the key questions.
  - Not necessarily all the questions.
  - Some variables will include non-response, identified by:
    - not stated
    - don’t know
    - refusal
  - Sometimes an entire block is missing.

What Should You Do About Partial Non-response?

- Assess the extent of the partial non-response.
- Determine the impact on your analysis
  - By examining the variables related to the variable of interest
  - See if the missing responses can form a separate category
  - Decide to do non-response processing
    - reweight for each variable to take partial non-response into account. Can be very tedious.
    - document your processing procedures

NLSCY Data Collection Strategy



0 1 2 Early Years 13

0 1 2 3 4 15

0 1 2 3 4 5 6 Released summer 2003 17

0 1 2 3 4 5 E.Y. 8 19

Sample counts

Cycle 1		Cycle 2		Cycle 3		Cycle 4	
Age	Sample	Age	Sample	Age	Sample	Age	Sample
						0	1795
						1	2213
				0	1736	2	2726
				1	6390	3	4390
		0	1962	2	1594	4	1398
		1	2192	3	2029	5	1680
0	1867	2	1799	4	1976	6	1721
1	2083	3	1855	5	1531	7	1651
2	1505	4	1426	6	1536	8	1187
3	1453	5	1271	7	1049	9	1132
4	1382	6	1313	8	1382	10	1077
5	1270	7	1116	9	941	11	1007
6	1211	8	1146	10	1241	12	927
7	1181	9	1023	11	843	13	928
8	1252	10	1193	12	1259	14	956
9	1211	11	1056	13	872	15	916
10	1278	12	1202	14	1256	16	976
11	1210	13	1068	15	906	17	881

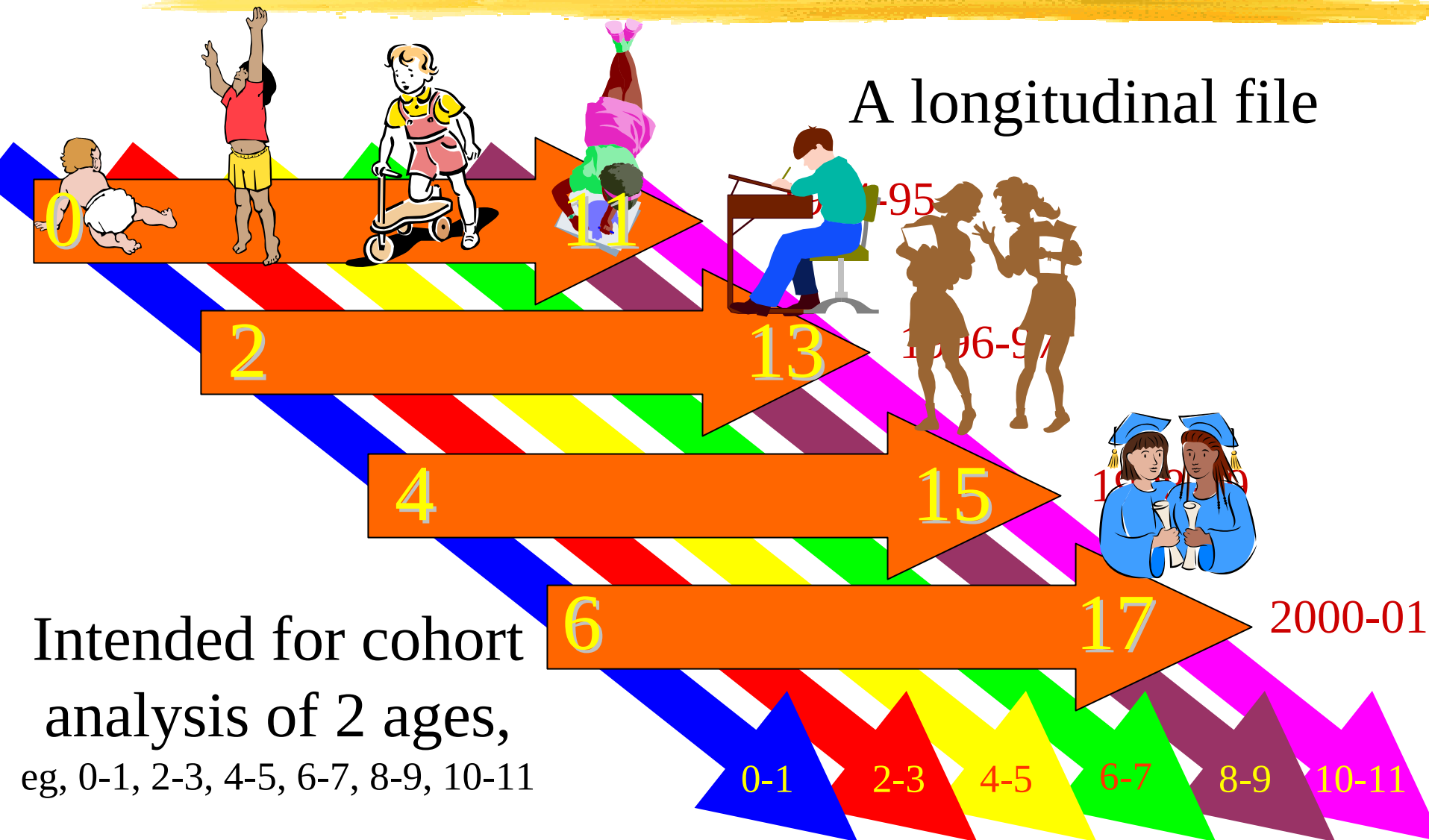
Issues

- CROSS-SECTIONAL ANALYSIS
  - Limitations due to the age of the sample
    - part of the sample was not selected for cross-sectional estimates
    - inherent complexity in the sample design to meet divergent needs
    - coverage problems
      - no update of the sample to reflect changes in the population (e.g., immigration); only the sampling weights have been adjusted to reflect changes
      - the older the cohort gets, the more difficult it is to adjust the sampling weights properly

Issues

- CROSS-SECTIONAL ANALYSIS
  - Limitations due to the nature of the survey
    - Some concepts were defined for the purposes of longitudinal analysis
    - Problems with sample erosion
    - Conditioning bias
  - Interpretation of the results
    - Impact on the effectiveness of estimation methods
    - Making inferences
    - Greater potential with the supplementary samples that have been added

One Survey but actually many datasets



Intended for cohort analysis of 2 ages,

eg, 0-1, 2-3, 4-5, 6-7, 8-9, 10-11

Sample counts

Cycle 1		Cycle 2		Cycle 3		Cycle 4	
Age	Sample	Age	Sample	Age	Sample	Age	Sample
0	1867	2	1799	4	1976	6	1721
1	2083	3	1855	5	1531	7	1651
2	1505	4	1426	6	1536	8	1187
3	1453	5	1271	7	1049	9	1132
4	1382	6	1313	8	1382	10	1077
5	1270	7	1116	9	941	11	1007
6	1211	8	1146	10	1241	12	927
7	1181	9	1023	11	843	13	928
8	1252	10	1193	12	1259	14	956
9	1211	11	1056	13	872	15	916
10	1278	12	1202	14	1256	16	976
11	1210	13	1068	15	906	17	881

Issues

- LONGITUDINAL ANALYSIS
  - Limitations due to sample erosion
    - sample shrinkage problems
    - representation (coverage) problems
    - Swiss cheese problems
  - Conditioning bias
  - Interpretation of results
    - impact on effectiveness of estimation
    - inferences

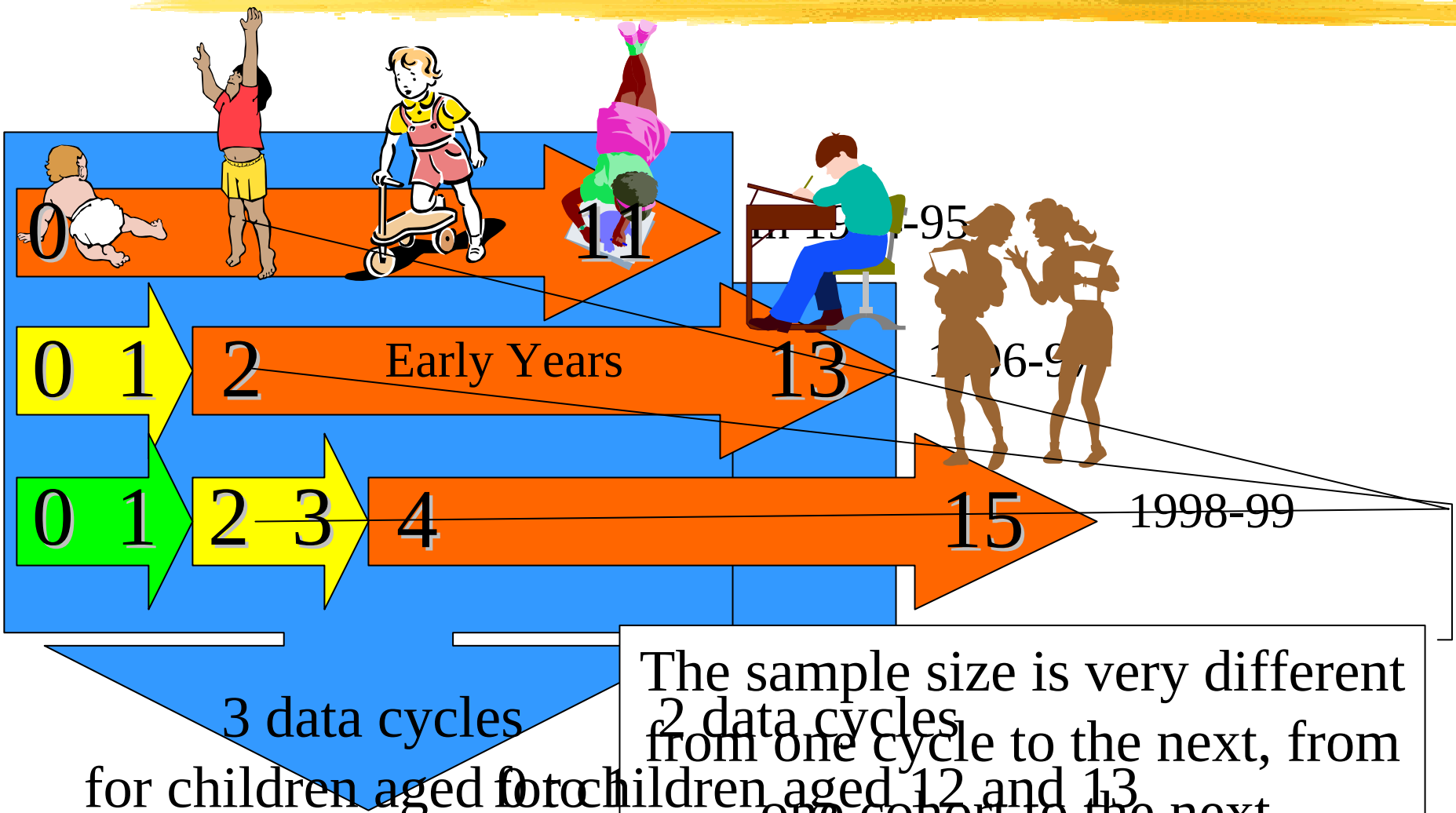
Issues

- MIXED ANALYSIS (longitudinal and cross-sectional)
  - Pay attention to the differences in the population targeted by the two types of analysis
  - Sample sizes vary a lot for these two types of analysis
  - Pay attention to the conclusions drawn from these analyses
  - The problems mentioned earlier can take different forms depending on the type of analysis

Dissecting NLSCY Data

Cross-sectional Data

Repeated Surveys



3 data cycles
for children aged 0 to

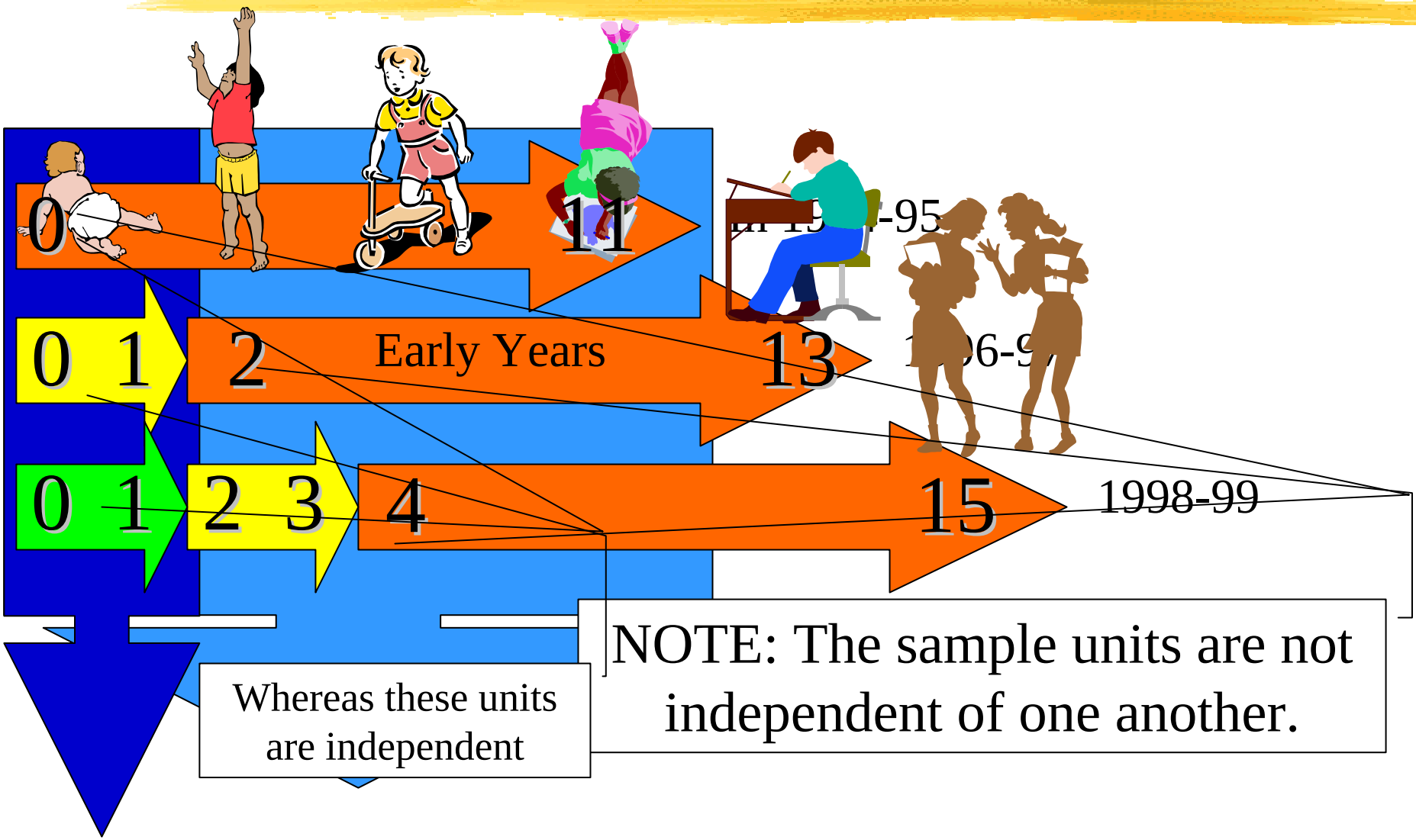
2 data cycles
for children aged 12 and 13

The sample size is very different from one cycle to the next, from one cohort to the next.

Dissecting NLSCY Data

Cross-sectional Data

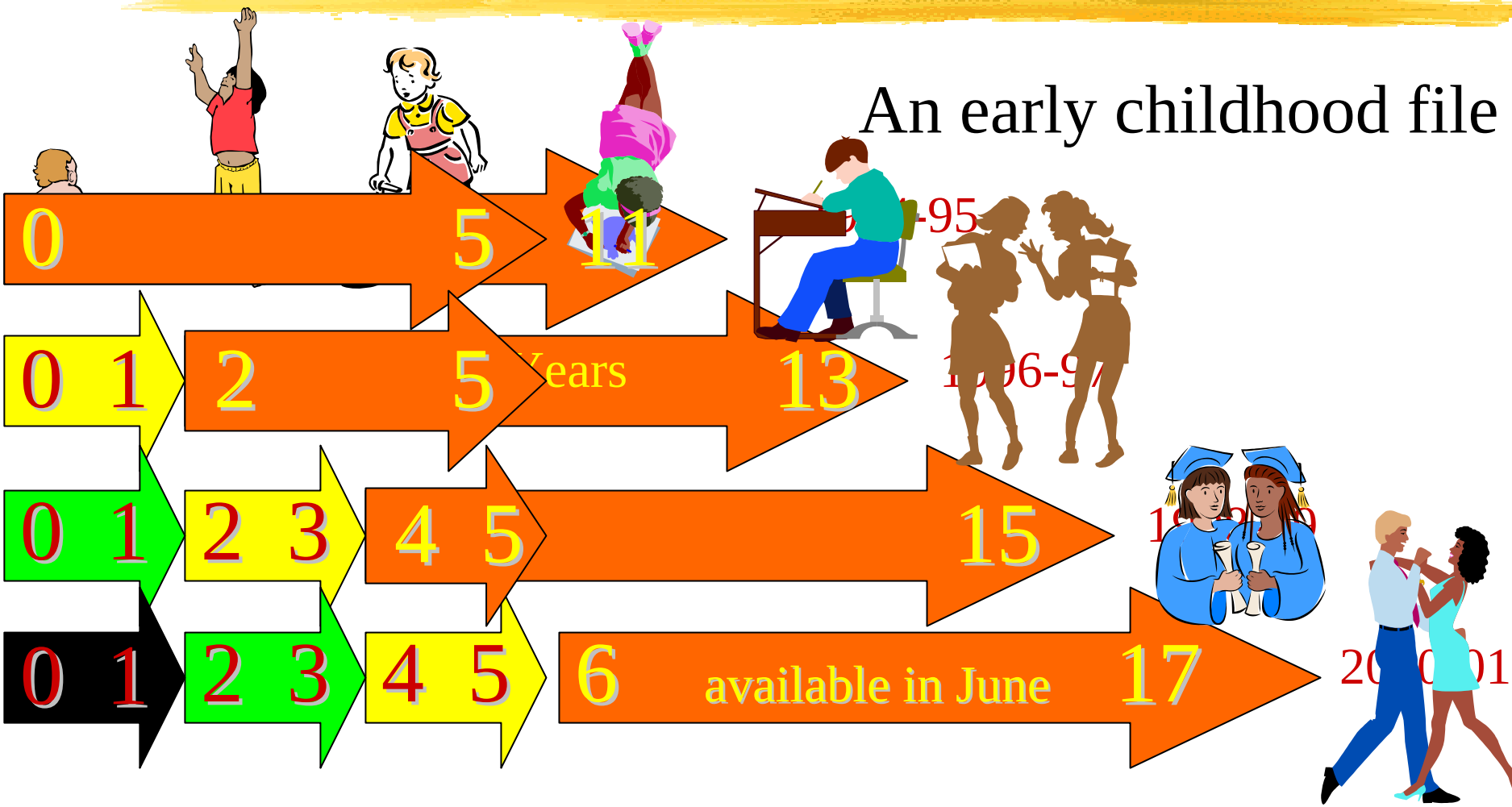
Repeated Surveys



Issues

- CROSS-SECTIONAL ANALYSIS (REPEATED)
  - Same limitations as noted earlier
  - The sample overlaps from one cycle to the next.
  - Independence or interdependence of samples
    - There is sample interdependence when the sample is made up of the same respondents
    - Involves a covariance factor
    - Sample independence is possible only for certain domains (e.g., children aged 0-1)

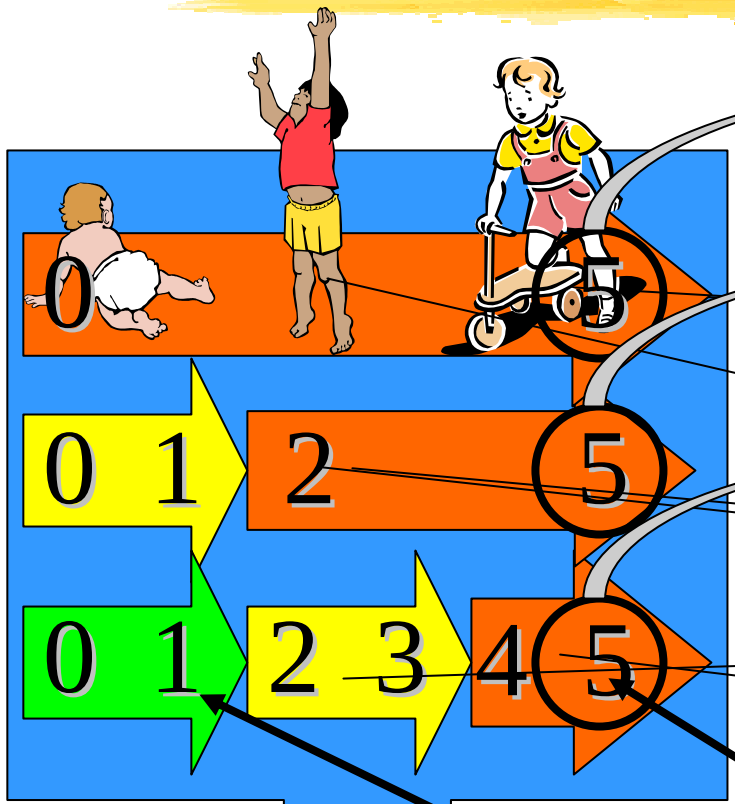
One Survey but actually many datasets



Dissecting NLSCY Data

Early Years

Cross-sectional Data



12,333 1,793

10,465 1,453

20,210 5,420

12,333 children aged 0 to 5 in Cycle 1 in a cross-sectional study.

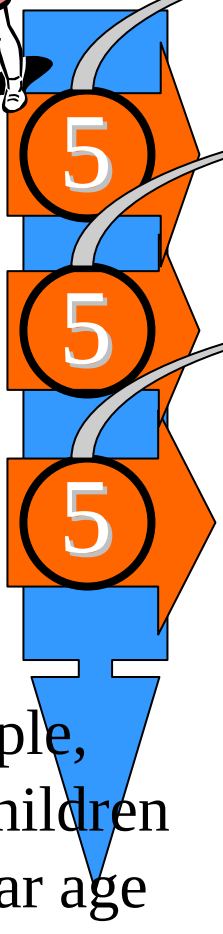
10,465 2-year-olds

5,420 five-year-olds

The sample size is very different from one cycle to the next, from one cohort to the next.

Or, once again, repeated analysis of the three data cycles

Analysis of Overlapping Domains



Born in 1989

Born in 1991

Born in 1993

For example, analysis of children at a particular age

Analysis of five-year-olds.

- There are actually 8,666 five-year-olds in the 3 cycles.
- The reference period becomes an analytical variable.
- Inference for a prescribed population is in context.

1,793

1,453

5,420

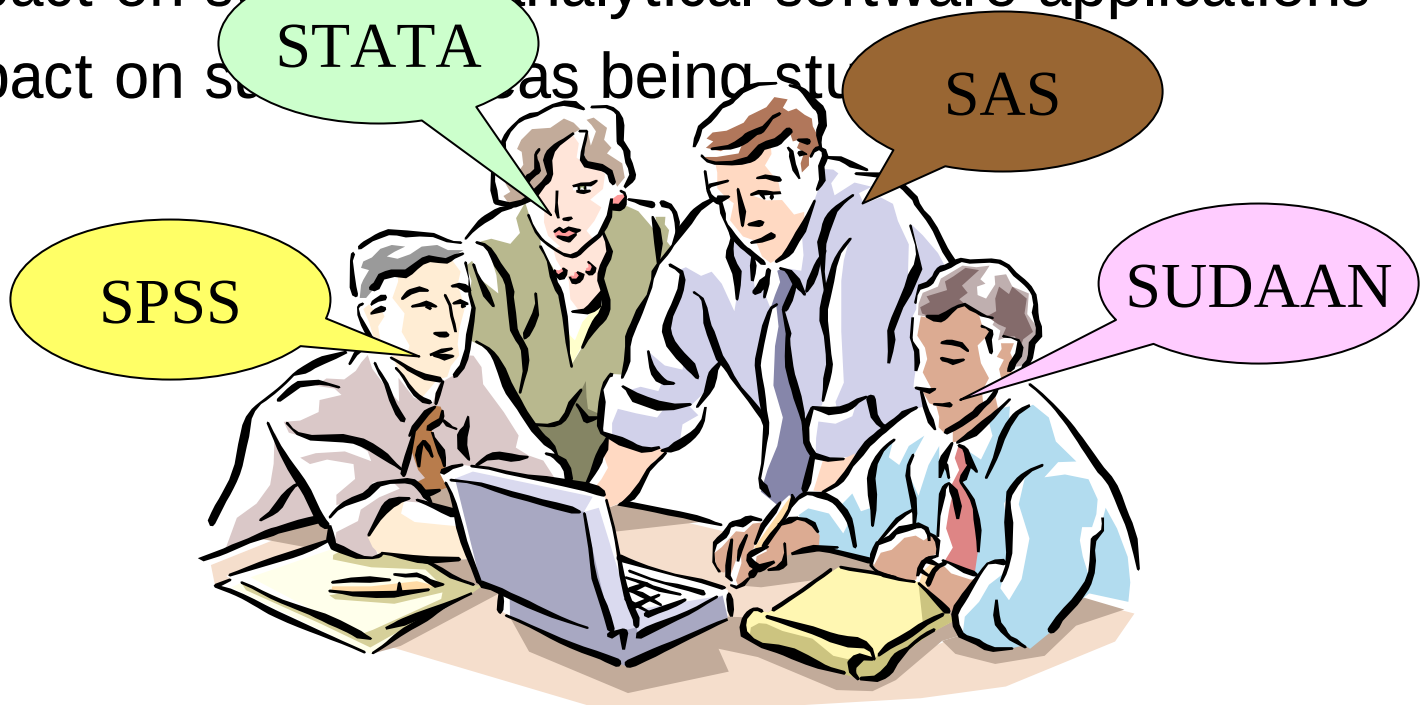
8,666

Issues

- OTHER – OVERLAPPING ANALYSIS
  - Pay attention to the definition of the subject area being studied
    - If the same respondent unit appears in the analytical sample more than once
      - Are they independent?
      - Do they impair the validity of the conclusions?
    - Relative weight of each unit in the analysis
    - Has the unit of analysis changed?
  - Pay attention to the conclusions drawn from the analyses

Where Does the NLSCY Sample Come From?

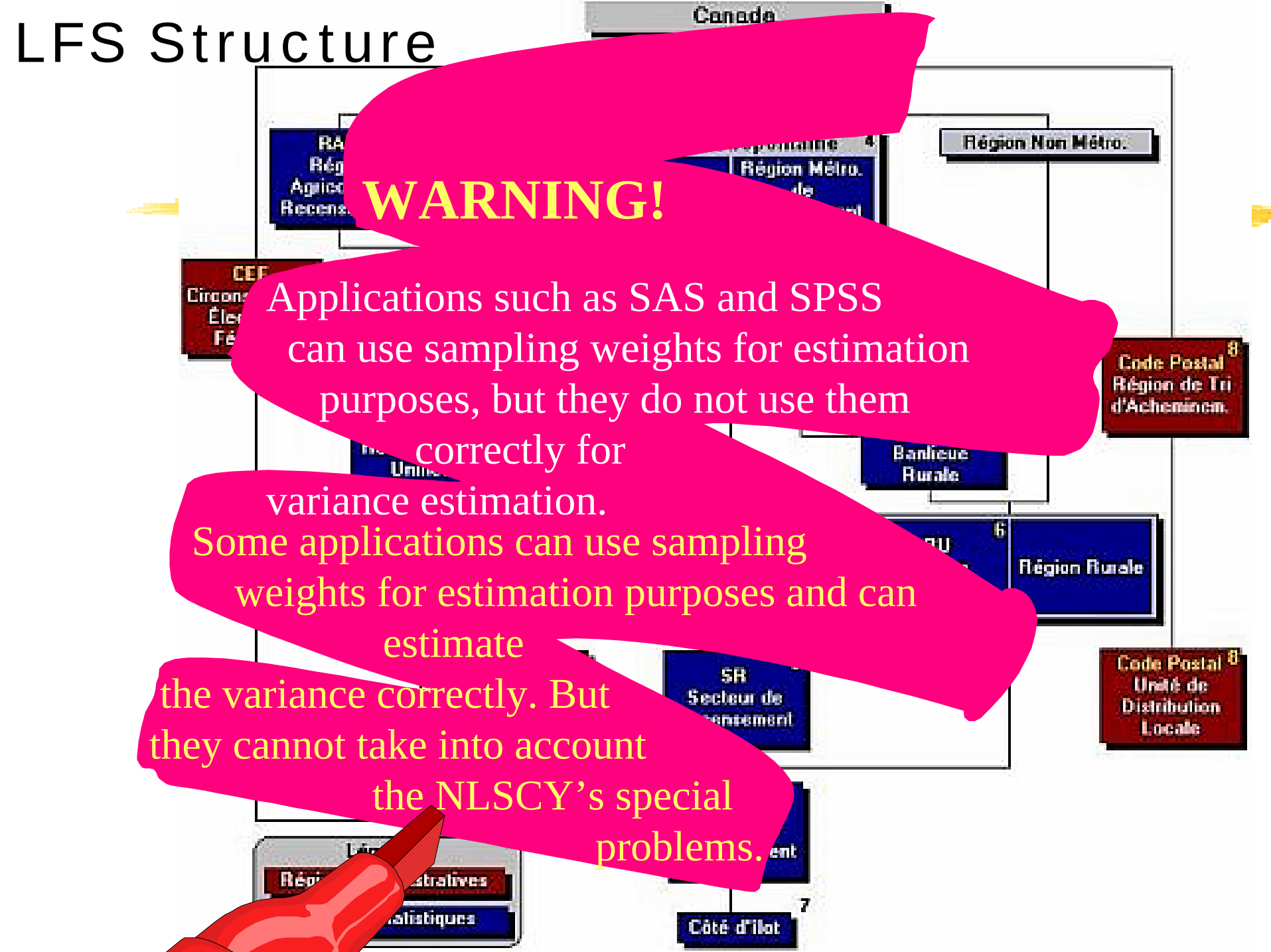
- Why ask the question?
  - Issues concerning analysis
    - impact on statistical/analytical software applications
    - impact on s as being stu



The NLSCY Sample

- A large part of the sample comes from the LFS
  - geographic stratification
  - multi-stage, with the primary sampling unit being a geographic cluster
  - with a systematic sample of households
  - After the LFS interview, households are identified as containing or not containing the population of interest (children)
  - The in-sample unit is the child (not the household)
- Constitutes the initial frame of children selected in 1994
- Main source of newborns and cross-sectional samples

LFS Structure



Where Does the NLSCY Sample Come From?

- Other reasons for asking the question
  - Issues concerning the type of analysis that can be done
    - cross-sectional, longitudinal, other
  - Issues concerning interpretation of the results
    - impact on the effectiveness of estimation methods
    - inferences

Implications for Analysis

- The structure favours analysis for geographic areas
- The advantage is reflected in a gain in operational efficiency
- Target the analysis to take advantage of this structure
- Loss of effectiveness for other subject areas
- Can use a larger sample for the same cost
- Some estimation methods can improve efficiency

The NLSCY Sample

- Sample size changes substantially from one initiative to another
  - between the Early Years (EY) cohorts and the longitudinal cohort
  - from one cycle to the next
  - because the sample size was increased with exclusively cross-sectional units.

The NLSCY Sample

- Renewal of the sample with EY cohorts is not consistent
  - focus is on new age 0-1 cohorts
  - specific additions such as EY (five-year-olds)
- This is a consequence of having analytical goals that vary from one initiative to the next and one cycle to the next

The NLSCY Sample

- For large supplementary samples
  - we used the Birth Registry
    - geographic and demographic stratification
    - simple random sample of geographic clusters
    - with a subsample of children in each cluster
  - it's the option used for the samples of one-year-olds and five-year-olds in Cycles 3 and 4.

Information Products

- Applied Research Branch documents
- Applied Research Bulletin - *Special Edition on Child Development* - Fall 1999
- "Investing in Children: Ideas for Action" - 1999
- NLSCY's "Developments" newsletter - occasional
- *Vulnerable Children*, book by Doug Willms - 2000
- Web Site: www.hrdc-drhc.gc.ca/nlscy-elnej

Who Uses NLSCY Data?

- Canadian Council on Social Development - *Progress of Canada's Children,* annual report on children (Web site: www.ccsd.ca)
- Vanier Institute of the Family - *Transition* and *Families and Health*
- *Canadian Living* - December 1996 and March 2000 issues
- *"The Early Years - Reversing the Real Brain Drain"* - Fraser Mustard and Margaret McCain - 1999
- Some provinces use the data to evaluate their programs for children
- National Children's Agenda (federal/provincial/territorial) uses the data to develop indicators

A Key Goal of Analysis...

- Be able to derive facts from the data
- A scientific sample design allows for:
  - estimates
  - inferences
  - a degree of certainty
- The analysis is judged on its ability to:
  - back up its statements
  - prove its validity

The Survey Vehicle

We have seen that the survey is loaded with information

The Survey Vehicle

NLSCY

We can lighten the load by targeting our analysis

The Survey Vehicle

NLSCY

We can see what’s possible and what’s not

The Survey Vehicle

NLSCY

We can greatly improve the survey's effectiveness
by taking advantage of the way it's constructed

The Survey Vehicle

NLSCY

When we adjust our analytical approach, the vehicle becomes lighter and more manoeuvrable

The Survey Vehicle

NLSCY

And we know we're not the only ones doing analysis

The Survey Vehicle

How to find stuff about the NLSCY

www.statcan.ca/english/ssds/44500.htm

www.statcan.ca/english/ips/data/89f0078x1e1999003.htm

www.hrdc-drhc.ca/sp-ps/arb-dgra/nlscy-elnej/home.shtml

My cordinates

André Cyr
Room 2500, Main Building, Statistics Canada
Tunney`s Pasture, Ottawa, Ontario
K1Y 4P2

andre.cyr@statcan.ca
Tel: (613) 951-1452