Conducting Research Using NLSCY Data

Some thoughts and ideas

Rochelle Garner, PhD
Health Information & Research Division
Statistics Canada
Presentation

- Some research examples using NLSCY
- Things to Know (T2K) about the NLSCY
- Group discussion
  - What issues have you come across with respect to NLSCY?
  - Questions about conducting research with survey data in general?
Research using NLSCY

- Lots of research topic possibilities
- Cross-sectional vs. longitudinal
- Uni-level vs. multi-level
  - Children within families
  - Children within neighbourhoods
  - Children within classrooms
- If you’re going to do research with NLSCY, make use of all that it has to offer
Changes in Smoking Behaviors From Late Childhood to Adolescence: Insights From the Canadian National Longitudinal Survey of Children and Youth

Stefania Maggi  
Thompson Rivers University

Clyde Hertzman  
University of British Columbia

Tracy Vaillancourt  
McMaster University

Objective: To examine smoking behaviors in Canadian youth from late childhood to adolescence. By following participants from as young as 10 and 11 years, the authors proposed to identify distinct developmental pathways of smoking acquisition. Design: Growth mixture modeling was used to identify developmental trajectories of smoking among 10- to 17-year-old participants of the Canadian National Longitudinal Survey of Children and Youth. Main Outcome Measures: Developmental trajectories of trying smoking, smoking frequency, and smoking intensity. Results: Five developmental trajectories related to smoking frequency were identified, of which 2 were acquisition patterns that led to daily smoking at age 16–17, and 3 were experimentation patterns that led to nonsmoking at age 16–17. The largest variability in changes in smoking behavior over time was the reported level of smoking frequency. Conclusion: Analysis showed that there is more than 1 way in which Canadian children and adolescents acquire smoking behaviors over time. The authors were able to differentiate patterns of experimentation from patterns of acquisition. Whereas experimentation has been generally considered as 1 of the stages in the smoking acquisition process leading to regular smoking, these results indicate that experimentation can be described as a distinct process in itself.

Keywords: smoking acquisition, smoking experimentation, developmental trajectories of smoking, late childhood, adolescence
238

MAGGI, HERTZMAN, AND VAILLANCOURT

Figure 4. Longitudinal changes in the frequency of smoking between the ages of 10–11 and 16–17 years, reported by Canadian National Longitudinal Survey of Children and Youth respondents (n = 280). Frequency was measured on a scale of 0 to 4 where 0 = don’t smoke, 1 = a few times a year, 2 = at least once or twice a month, 3 = at least once or twice a week, 4 = every day. Five trajectories were identified: late infrequent experimenters (6.1%; those who reported smoking once or twice a year around 15–16 years of age), late frequent smokers (38%; those who gradually increased frequency of smoking, especially after the age of 14), early frequent experimenters (5.2%; those who reported smoking at least weekly between 10 and 15 years of age), early frequent smokers (34%; those who reported a sharp increase in smoking at around 12 years of age), and early infrequent experimenters (6.8%; those who reported smoking at least once or twice a year at age 10–11 only).

$5.030, p > .05$, or intensity, $\chi^2(1, N = 272) = .039, p > .05$, of smoking.

_Trying, frequency, and intensity of smoking._ Chi-square analysis revealed that the middle onset group was more likely to fall in the late frequent smokers trajectory, and that the early onset group $\chi^2(4, N = 392) = 11.432, p < .05$. That is, participants who tried smoking after the age of 12 were more likely to become daily smokers by the time they reached 16–17 years of age, whereas those who tried smoking before the age of 12 were more likely to smoke infrequently at age 14–15 and to have stopped smoking.
DOI 10.1007/s10964-006-9140-8

ORIGINAL PAPER

Adolescent Children of Adolescent Mothers: The Impact of Family Functioning on Trajectories of Development

V. Susan Dahinten · Jennifer D. Shapka · J. Douglas Willms

Received: 29 May 2006 / Accepted: 21 September 2006 / Published online: 10 January 2007
© Springer Science+Business Media, LLC 2007

Abstract This study drew on four cycles of longitudinal data from the Canadian National Longitudinal Survey of Children and Youth to examine the academic and behavioural trajectories of youth between 10 and 15 years of age as a function of maternal age at childbearing. The analyses controlled for several family characteristics and examined the mediating effects of three family functioning variables (maternal depression, and nurturing and rejecting parenting behaviours). Maternal age was related to academic competency in math (standardized Math scores), externaliz-
Correlates of Methylphenidate Use in Canadian Children: A Cross-Sectional Study

Alice Charach, MD, MSc, FRCPC\textsuperscript{1}, Hongmei Cao, MSc\textsuperscript{2}, Russell Schachar, MD, FRCPC\textsuperscript{3}, Teresa To, PhD\textsuperscript{4}

**Objectives:** This study aimed to estimate the prevalence of methylphenidate (MPH) use among Canadian children aged 2 to 11 years, from 1994–1995 to 1998–1999 and to identify sociodemographic, child, and parent–family correlates are identified.

Correlates of Methylphenidate Use in Canadian Children: A Cross-Sectional Study


<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample population (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 962 900</td>
</tr>
<tr>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Ritalin use</td>
<td>51 500</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1 971 400</td>
</tr>
<tr>
<td>Age groups</td>
<td></td>
</tr>
<tr>
<td>Preschool (2–5 years)</td>
<td>1 564 200</td>
</tr>
<tr>
<td>School age (6–11 Years)</td>
<td>2 298 700</td>
</tr>
<tr>
<td>Child behaviour</td>
<td></td>
</tr>
<tr>
<td>Parent-reported hyperactivity</td>
<td>309 700</td>
</tr>
<tr>
<td>Parent-reported aggression</td>
<td>301 100</td>
</tr>
<tr>
<td>Socio or demographic factors</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>3 154 200</td>
</tr>
<tr>
<td>Low income adequacy</td>
<td>690 200</td>
</tr>
<tr>
<td>Low maternal education</td>
<td>626 700</td>
</tr>
</tbody>
</table>
Multilevel Correlates of Childhood Physical Aggression and Prosocial Behavior

Elisa Romano, Richard E. Tremblay, Bernard Boulerice, and Raymond Swisher

Received December 1, 2003; revision received December 15, 2004; accepted January 10, 2005

The study identified independent individual, family, and neighborhood correlates of children’s physical aggression and prosocial behavior. Participants were 2,745 2–11-year olds nested in 1,982 families, which were themselves nested in 96 Canadian neighborhoods. Hierarchical linear modeling showed that the total variation explained by the three-level model was 20.83% for physical aggression and 17.57% for prosocial behavior. For both childhood behaviors, approximately 66% of this explained variance was between individuals and up to 30% was between families. The smallest amount of observed variation was between neighborhoods. Significant individual-level predictors common to both childhood behaviors were child’s sex and maternal hostility toward the target child. Specifically, boys had more mother-reported physical aggression and less prosocial behavior. Children who experienced greater-than-average maternal hostility (compared to siblings) were more physically aggressive and less prosocial. At the family level, significant common predictors were mother depressed mood and punitive parenting. Children had higher levels of physical aggression and lower levels of prosocial behavior in families where mothers had greater depressed mood and used more punitive parenting practices. At the neighborhood level, greater perceived problems and lower poverty level were associated with higher levels of physical aggression. Results are discussed with reference to past and future studies of multilevel effects on children’s socialization.

KEY WORDS: physical aggression; prosocial behavior; children; multilevel correlates.
BMC Public Health

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Effects of neighbourhood income on reported body mass index: an eight year longitudinal study of Canadian children


Lisa N Oliver (loliver@sfu.ca)
Michael V Hayes (mhayes@sfu.ca)

ISSN 1471-2458

Article type Research article

Submission date 8 May 2007

Acceptance date 14 January 2008

Publication date 14 January 2008

Article URL [http://www.biomedcentral.com/1471-2458/8/16](http://www.biomedcentral.com/1471-2458/8/16)

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance.
into 4 categories based on self-report: less than high school, high school, some postsecondary and postsecondary degree or diploma. Three analytic categories were created: less than high school, high school/some post secondary (omitted reference group), and post secondary degree/diploma.

Neighbourhood

Enumeration Areas, the smallest level of census geography in 1996 containing between 125 and 440 dwellings, were used as neighbourhood proxies. Neighbourhood low-income was assessed from the 1996 Census by calculating the proportion of the non-institutional population living below the low income cut-off. Proportions were divided into quintiles and the three middle quintiles were grouped resulting in three categories: ‘least poor’, ‘middle’ (omitted reference category) and ‘most poor’.

Whether the neighbourhood was urban or rural (coded 0=rural and 1=urban) was determined by a variable indicating if it was in a Census Metropolitan Area (CMA). A CMA consists of one or more municipalities that form an urban core greater than 100,000 residents [25].

While data were available for 2,229 children at Cycle 5 some were excluded because of missing data for PMK education, family structure, or because a postal code could
Research using NLSCY

- Always begin with a search of the literature
- Questionnaire content may spark some research questions
- Read the “future analysis” section in your favorite journal
  - There’s always a “next step” needed to answer a question
  - Your research could be that next step
Things to Know (T2K)
Things to Know (T2K)

- Child’s age
- Change to variables over time
- Carry forward variables
- Identifiers
- PMK Changes over time
- Skip patterns
- Same questions, different variables
- Multiple data files
T2K: Child’s Age

- NLSCY datasets contain two “age of child” variables
  - XMMMCQ01: Effective age of child
  - XMMMCbQ01: Age of child at date of interview
- Effective age used to determine which set of questions are administered for a particular child
- Your research question will determine which age variable is most appropriate
T2K: Change to Variable

- When there is a change to a variable
  - …in terms of wording
  - …in terms of response categories
  - …if variable is new

- This will be indicated by an extra lower-case letter in the variable name
  - E.g. FDRCdQ01
  - The lowercase “d” indicates that a change was made in Cycle 4 to this question
T2K: Carry Forwards

- Some variables are “carried forward” over the course of the survey
  - I.e. Although question is only asked once of the respondent, the response will be carried forward for that question longitudinally
    - Response does not change
  - E.g. Was the child ever breastfed?
- This has impact on ability to examine certain questions longitudinally
- Carry forwards are often indicated in the Codebook, but not always
Has the child ever been breast-fed?

1  Yes
2  No
6  Valid skip
9  Not stated

Coverage: Children aged 10 or 11 years old, if respondent is the biological parent (mother or father)
Note: This derived variable is based on questions MEDQ25 and MEDQ26. Data have been carried forward from previous cycles when available. If no information has been collected from the biological mother or father at the time of the interview in previous cycles, this variable value will be 'Not stated'. 
National Longitudinal Survey of Children and Youth, Cycle 6, 2004
Master file (Longitudinal)

Variable Name: FCHPQ1D
Collection Name: HLAQ1AE
Position: 358
Length: 1

Do you have any of the following long term conditions

...arthritis or rheumatism?

1 Yes
2 No
6 Valid skip
9 Not stated

Coverage: PMK with selected child aged 10 to 15 years old
Note: If firstint = 2 (Not the first interview for this adult in Cycle 6), data have been carried forward from previous cycles when available.

Variable Name: FCHPQ1E
Collection Name: HLAQ1AF
Position: 359
Length: 1

Do you have any of the following long term conditions

...back problems excluding arthritis?

1 Yes
2 No
6 Valid skip
9 Not stated

Coverage: PMK with selected child aged 10 to 15 years old
Note: If firstint = 2 (Not the first interview for this adult in Cycle 6), data have been carried forward from previous cycles when available.
T2K: PMK Changes

- In each cycle, the Person Most Knowledgeable (PMK) of the child provides information
  - This person is not necessarily the same at every cycle
- Different informants over time may have impact on your research
- Later cycles can use PPERSRUK and SPERSRUK
  - Earlier cycles may require PMK gender and date of birth information to track individuals
T2K: PMK Changes Cont’d

| Variable: BDMPbD27 | Position: 104 | Length: 1 |

### PMK Change

1. PMK is the same person for every cycle
2. PMK in cycle 2 is the SPOUSE in cycle 1
3. PMK in cycle 2 is another person
4. PMK and/or SPOUSE not surveyed in cycle 1
5. NOT APPLICABLE
6. DON'T KNOW
7. REFUSAL
8. NOT STATED

---

| Variable: CDMPbD27 | Position: 102 | Length: 1 |

### PMK Change

1. PMK is the same person for every cycle
2. PMK in cycle 3 is the SPOUSE in cycle 2
3. PMK in cycle 3 is another person
4. PMK and/or SPOUSE not surveyed in cycle 2
5. NOT APPLICABLE
6. DON'T KNOW
7. REFUSAL
8. NOT STATED

---
T2K: Identifiers

- Identifiers used in the NLSCY
  - PERSRUK: Unique identifier for child
  - CHILDID: Identifier unique to child only within a cycle
    - Used Cycles 1 through 3
  - FIELDRLUK: Unique identifier for household
  - PPERSUK: Unique identifier for PMK
    - Since Cycle 6
  - SPERSRUK: Unique identifier for spouse
    - Since Cycle 6
T2K: Skip Patterns

- Some questions in the NLSCY are only asked for a sub-group of respondents who are “screened through” on basis of characteristics or previous answers.
  - For these variables, there is often a response category (listed only in Codebook) called “Valid skip”.

- Meaning of “valid skips” will depend on the question and your research focus.
T2K: Skip Patterns Cont’d

- For example, valid skip may indicate…
  - Question only asked for certain age groups
  - Question only relevant to people who reported a particular characteristic
    - E.g. only youth who report smoking are asked how many cigarettes they smoke on a typical day

- Knowing the reason for the skip will determine how you treat the valid skips
T2K: Same question, Different variables

- Sometimes a topic will be covered in different ways in different age groups
  - Response categories may differ between age groups
  - There will be a “valid skip” category
- E.g. Alcohol consumption among youth
**Variable Name:** FDRCcQ06  
*Position:* 302  
*Length:* 1

Have you ever had a drink of alcohol?

1. Yes, at least one drink  
2. I have only had a few sips  
3. No  
6. Valid skip  
9. Not stated

**Coverage:** Respondents 10 to 11 years old

**Variable Name:** FDRCdQ6A  
*Position:* 303  
*Length:* 2

Which of the following best describes your experience with drinking alcohol?

01. I have never had a drink of alcohol  
02. I have only had a few sips  
03. I only tried once or twice (at least one drink)  
04. I do not drink alcohol anymore  
05. A few times a year  
06. About once or twice a month  
07. About 1-2 days a week  
08. About 3-5 days a week  
09. About 6-7 days a week  
96. Valid skip  
99. Not stated

**Coverage:** Respondents 12 to 17 years old
T2K: Multiple data files

- For most cycles, data from different questionnaires are saved in separate data files:
  - Child file (primary, secondary)
    - Early child development cohorts and longitudinal cohort
  - Self-complete questionnaire
  - Teachers and principals questionnaires
  - Youth questionnaires
T2K: Multiple data files

- Check questionnaires for content
- Be sure to check all pertinent data files before you decide that data for a particular subgroup is not available
  - E.g. Information on particular topic may move from <Self-Complete> dataset to <Youth> dataset at age 18
Contact Information

Rochelle Garner
Health Information & Research Division
Statistics Canada

rochelle.garner@statcan.gc.ca
Over to You

What are your research and analysis questions?

How can we help you?
Question #1: Weighting

- Shouldn't everyone from the original cohort who participated in at least one follow-up survey have a longitudinal weight?
  - Why does the use of weights extensively decrease the sample size on some variables (e.g., af1cq01)?
- Are there instances in which we have to formulate our own weights?
Question #1: Weighting

Cycle 1
n=22,381
Everyone has cross-sectional weight

NPHS sample
n=3896

NLSCY
n=18,935

Cycle 2

Cycle 1 longitudinal weights were “back applied” to only these kids

NLSCY
n=16,903
Question #2: Bootstrap weights

- I would like to better understand how bootstrap weights are constructed.
  - For example, the conference covers how to reconstruct the weights for specific conditions. How does this apply to bootstrap weights?
Question #3: Weights across multiple cycles

- Is there a procedure for using the longitudinal weights when you are pooling some respondents across cycles, as a result of their experiencing some event (so not just simply pooling two cycles together).
  - E.g. starting time 0 at the occurrence of an event

- I would like to combine age groups across cycles to produce a large enough longitudinal sample to use 'group based trajectory modeling'

- How to apply weights? Who is the target population?
Question #3: Weights across multiple cycles

C1  C2  C3  C4  C5  C6

A: 8 yrs

B: 10 yrs

C: 6 yrs
Question #3: Weights across multiple cycles

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>
Question #4: Census data

- Are there restrictions on the extent to which Census data can be used with the NLSCY?
- Are there noted problems with using these variables (i.e. neighbourhood characteristics) with the NLSCY (e.g., cluster sizes, restrictions on types of analyses)?
Question #5: NLSCY redesign

- I've heard that there will be a re-launch of the NLSCY, with a new sample drawn. Is there going to be some sort of consultative process on the survey design, or a forum for existing NLSCY researchers to make suggestions about content changes?
  - How much of the new survey will duplicate the existing questions, and how much will be new content?
  - What types of changes are planned?
  - Who will be making the new content decisions, and when?