

# Data Structure of the Intergenerational Income Database

---

► **Catherine Haeck, Ph. D.**

Professor

Economics department



Groupe de recherche sur le  
**CAPITAL HUMAIN**  
**ESG** UQÀM

## ► What is the Intergenerational Income Database (IID)?

---

- The IID is a database that links administrative tax data for children and their parents
- It has a longitudinal structure that allows to follow both parents and children into adulthood
- It allows researchers to study intergenerational phenomena such as transmission of earnings from parents to children, income mobility, etc.

## ► Who is in the IID?

There are six cohorts of kids (and their parents)

	Kids year of birth																						
Cohort name	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
1982	63			66																			
1984			65			68																	
1986					67			70															
1991										72			75										
1996															77			80					
2001																				82			85

## ► Conditions to be in the IID

---

- Child birth cohort
- When the child was 16 to 19 years old
  - ✓ parent and child lived together at least one year in Canada
  - ✓ and filed at least one tax return that same year
- Parent may be a biological parent or not

*IID Coverage : 70% of 16 to 19 years old (use weights)*

## ► Who is NOT in the IID

---

- Other birth cohorts (for now)
- Immigrant children whose parents stayed in their home country
- Immigrant children who arrived after 16 to 19 years old
- Children not living with their parents as of 16 years old

## ► How is the data made available?

---

- The data comes into several pieces that need to be put together to form the longitudinal profile of the parent-child couple
- Here, we present the structure (and some of the content) of the data sets made available in the RDCs

## ► Label used by Statistics Canada

Kids year of birth	Cohort name	Panel
1963-66	1982	A
1965-68	1984	A
1967-70	1986	A
1972-75	1991	B
1977-80	1996	B
1982-85	2001	B

## ▶ Two main folders to find in your account

---

Path to main folders may resemble:

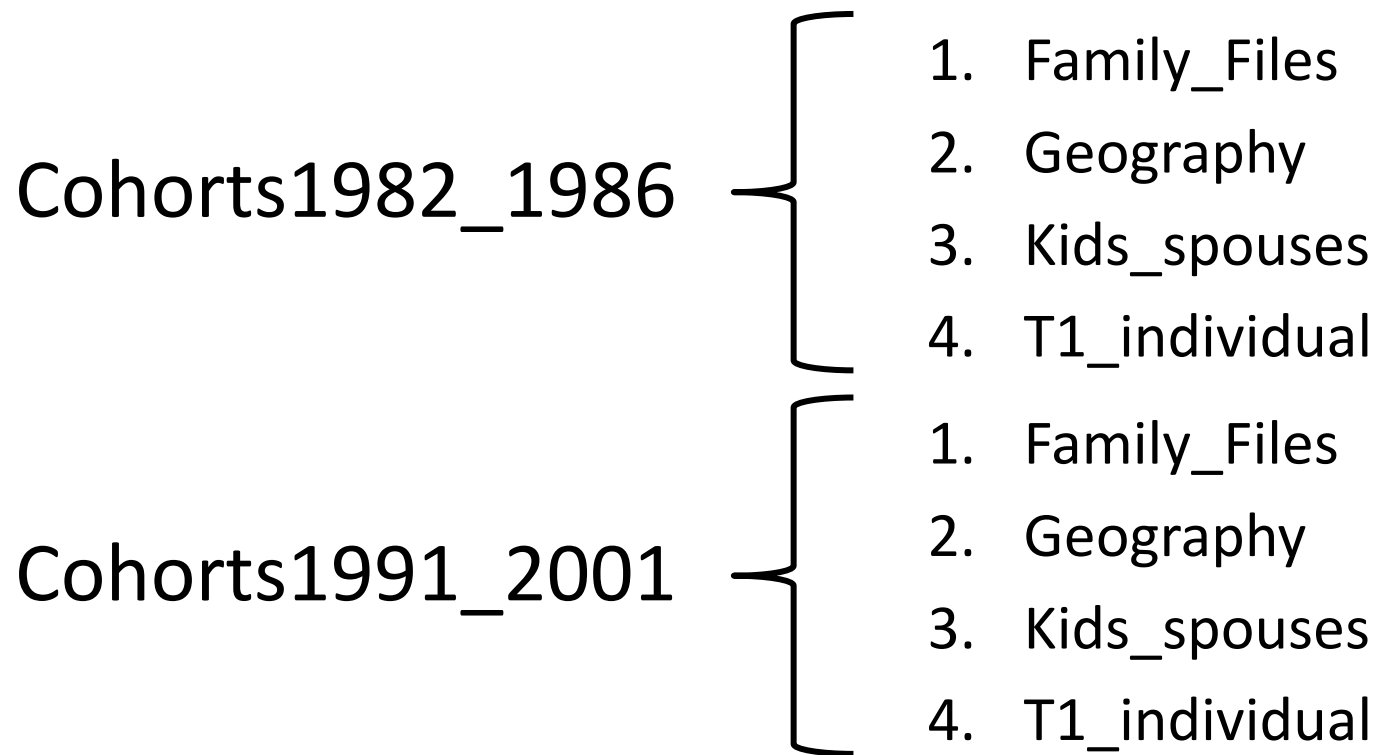
BD/IID\_BDMIR\_2015/IID\_BDMIR\_2015\_v1/data\_donnees/data/stat\_en

Data are zip in two files:

1. cohorts1982\_1986.zip
2. cohorts1991\_2001.zip



► Folder structure is identical for both cohorts



## ▶ 1. Family\_Files

---

- The family file was created from Statistics Canada's T1 Family File (T1FF)
- The T1FF is a dataset of T1 records from the Canada Revenue Agency that matches members of each tax filer's family
- The IID uses the child record (1 per child), with parents' identification and information on family structure

## ► 1. Family\_Files

---

Cohorts1982\_1986 {

- IID\_1982\_family\_f1\_v1.dta
- IID\_1984\_family\_f1\_v1.dta
- IID\_1986\_family\_f1\_v1.dta

Cohorts1991\_2001 {

- IID\_1991\_family\_f1\_v1.dta
- IID\_1996\_family\_f1\_v1.dta
- IID\_2001\_family\_f1\_v1.dta

## ► 1. Family\_Files – Main variables

---

IID\_**YEAR**\_family\_f1\_v1.dta

- *kcasenum* – kid ID
- *xcasenum* – (**d**)dad, (**m**)mom ID (bio or not)
- fpcode – postal code
- *yrl* – year link
- *xyob* – (**k**)kid, (**d**)dad, (**m**)mom year of birth (YYYY-MM\_DD)
- *xlsex* – (**k**)kid, (**d**)dad, (**m**)mom gender
- *numkid* – number of kids in the family

## ► 2. Geography

---

- Contains the census geography
- **YEAR** : 1982 to 2014

Cohorts1982\_1986 ➡ IID\_**YEAR**\_A\_geo\_f1\_v1.dta

Cohorts1991\_2001 ➡ IID\_**YEAR**\_B\_geo\_f1\_v1.dta


## ► 2. Geography – Which census year?

Fiscal year	Census Geo
1982-1986	1981
1987-1991	1986
1992-1996	1991
1997-2001	1996
2002-2006	2001
2007-2011	2006
2012-2014	2011

For a given fiscal year  
the census geography  
available is always the  
closest previous  
census year

## ► 2. Geography - Variables

---

- *casenum* - ID  Variable to use to link Geo with T1 files
- *pr* – province
- *cma81* – census metro area (census year : 1981)
- *cd81* – census division (census year : 1981)
- *csd81* – census subdivision (census year : 1981)
- *geo\_linked* – 1 if successfully linked (=0 if missing postal codes)

As of census 2011, geo files also include :

- *er11* - economic region (census year : 2011)

### ▶ 3. Kids\_spouses

---

- Contains T1 fiscal information on the spouses of the kids
- **YEAR** : 1981 to 2014

Cohorts1982\_1986 ➡ IID\_**YEAR**\_A\_ksp\_f1\_v1.dta

Cohorts1991\_2001 ➡ IID\_**YEAR**\_B\_ksp\_f1\_v1.dta



### ▶ 3. Kids\_spouses – Main variables

---

- *kcasenum* – ID kid
- *coh* – cohort identification year
- *kspbirthdate* – spouse birth date (YYYY-MM-DD)
- *kspsex* – spouse's gender
- + all T1 variables for the spouse

## ▶ 4. T1\_individual

---

- Contains T1 fiscal information for the kid, the dad and the mom
- **YEAR** : 1978(**81**) to 2014

Cohorts1982\_1986 ➡ IID\_**YEAR**\_A\_t1\_f1.dta

Cohorts1991\_2001 ➡ IID\_**YEAR**\_B\_t1\_f1.dta

## ► 4. T1\_individual – Variables

- *casenum* – ID
- *coh* – 82, 84, 86, 91, 96, 01
- *birthdate* – YYYY-MM-DD
- *type* – mother/father/child
- *lang\_cd* – FR/EN
- *sex* – gender
- *marital\_status* – (unstated, married, common-law, widowed, divorced, separated, single)
- ... + many others

IID Codebook  
Contains Complete  
Definition

## ► 4. T1\_individual – variable: *total\_inc*

From 1982 to present:

- Canada/Quebec Pension Plan Benefits
- Capital Gains/Losses Calculated
- Dividends, Taxable Grossed Up
- Earning from T4 Slips
- Interest and Investment Income
- Old Age Security Pension
- Other Employment Income
- Other Income
- Pension and Superannuation Income
- Rental Income
- Net Self-employment Income:
- Net Business Income
- Net Commission Income
- Net Farming Income
- Net Fishing Income
- Net Professional Income
- Employment Insurance Benefits

From 1986 to present:

- Alimony or Maintenance Income
- GST (page 1 of tax form) and FST Credit.

From 1988 to present:

- Limited Partnership Income Line 122. Prior to 1988, LTPI was included in Net Business Income, Net Rental Income, or Other Income.
- Registered Retirement Savings Plan Income. Prior to 1988, T4RSP was included in Other Income.

From 1992 to present:

- Net Federal Supplements
- Social Assistance Payments
- Workers' Compensation Payments

From 1996:

- Guaranteed Income Supplements, which is a component of Net Federal Supplements is available as a separate variable.

Also from 1982 to 1992 Family Allowance Received was included in the calculation of Total Income as defined by the Canada Revenue Agency, and from 1982 to 1987, Expenses, Allowable was subtracted from Total Income as defined by the Canada Revenue Agency.

## ► 4. T1\_individual – other variables (examples)

---

- *earn\_t4* : CRA's total earnings from T4 (wages, salaries, and commissions, before deductions. Excludes self-employment income.
- *ui\_ben* : Employment Insurance Benefits
- *charity* : Gross charitable donations claimed
- *cpp\_qpp\_ben* : Canada Pension Plan and Quebec Pension Plan benefits (retirement, disability, certain children's benefits, etc.)
- *childcare\_exp* : Calculated amount of child care expenses *allowed* as a deduction
- *sic\_code* : Standard Industrial Classification Code (Panel B only)
- *tuit\_educamt\_calc* : Total amount of tuition and education amounts claimed as credit by the client
- *mental\_dsblt\_cd* : 0/1 Mental disability, condition certified Form T2201 (Panel A only)

## ► IID Potential

---

- The potential for research using the IID is extremely high
- The number of observations allows researchers to look at finer subgroups
- The variety of questions investigated using the IID will depend on the variety of researchers using the IID

## ► What is the GRCH doing?

---

- We measure the transmission of income from parents to children
- We have mapped the geographical patterns and also the trends over time
- Our long term goal is to understand some of the mechanisms leading to our observed findings

## ► We create five non-overlapping cohorts

Cohort 84 is merged with cohorts 82 and 86

Kids year of birth		63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
Cohort name		63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
1982-84		63			66																			
1984-86						67			70															
1991											72			75										
1996																77			80					
2001																					82			85



## ► How we proceed using Stata

---

### Start from Family file (master do file)

- Add parent income (from T1)
- Add kid income (from T1)
- Add spouse income (from T1 and kids spouse data)
- Add geo (from geo data)
- Adjust for CPI
- Generate income rank
- Generate age at birth

## ► We proceed with cohorts defined using globals

```
89 // Now we go through a loop in order to build the rest of the main file
90 glo cohorts "1963 1967 1972 1977 1982"
91 foreach cohort in $cohorts {
92     cap log close
93     glo cohort `cohort`
94     if `cohort`==1963 {
95         glo minYOB 1963
96         glo maxYOB 1966
97         glo minT1Age 25 // Define the youngest age at which we'll look for T1 files
98         glo maxT1Age 39 // Define the oldest age...
99         glo cohortFile1 1982
100        glo cohortFile2 1984
101        glo cohortName 82_86
102    }
103
104    if `cohort`==1967 {
105        glo minYOB 1967
106        glo maxYOB 1970
107        glo minT1Age 25 // Define the youngest age at which we'll look for T1 files
108        glo maxT1Age 39 // Define the oldest age...
109        glo cohortFile1 1984
```

## ► Papers from the GRCH on mobility

---

- “Intergenerational Mobility between and within Canada and the United States”, Marie Connolly, Miles Corak, and Catherine Haeck
- “Social Mobility Trends in Canada: Going up the Great Gatsby Curve”, Marie Connolly, Catherine Haeck, and David Lapierre
- Visit us at : <http://grch.esg.uqam.ca>

# ESG UQÀM



Groupe de recherche sur le  
**CAPITAL HUMAIN**  
**ESG UQÀM**