The inheritance of opportunity in 19th century Canada

Luiza Antoine, Guelph
Kris Inwood, Guelph
Chris Minns, LSE
Fraser Summerfield, St Francis Xavier

What are we doing?

- How early labour market experience is influenced by father's occupation
- Using 1871 and 1891 census samples linked to 1881 complete count data
- Examine young adults in 1881 & 1891 against the heads of their families/households ten years earlier
- Occupation the variable of interest

Relevant literatures

- Intergenerational mobility UK and US (Ferrie and Long, EJ 2007; Long EREH 2013; Long and Ferrie, AER 2013); Argentina (Perez, JEH forthcoming)
- “First jobs” and parental networks speak to core of economic sociology (Granovetter 1975; Montgomery, AER 1991)
- Dissipation of entry effects in immigrant families? (Inwood, Summerfield, Minns, EREH 2016)
- Anglophone vs Francophone Canadian (Albouy CJE 2008; Green, MacKinnon and Minns, JEH 2006)

First jobs?

- Corak and Piriano (JOLE 2011) document high share of young Canadian men who work in their father's firm
- In contrast to placements patterns for English apprentices pre-1800 (Leunig, Minns, Wallis, JEH 2011)
- How much occupational inheritance in the late 19thC, and what explains it?
- Inheritance of first job is one aspect of intergenerational mobility

Evidence: linked census records

- 1st stage: SVM classification on time invariant characteristics with true links, minimizes false matches, leaves many ambiguous (Antonie et al 2014)

- 2nd stage: disambiguation of multiples with persisting co-residence of family members (Richards 2013)

- assess sample size, error rate and bias at each stage

- 7% samples of 1871 and 1891 census enumerations linked to1881 ‘complete count’

- 1871-1881 linked set takes those aged 16-25 yrs in 1871
- 1881-1891 linked set takes those aged 6-15 yrs in 1881

Data Challenges

- Imprecise recording
  - Spelling of names - transcription and/or enumeration
  - Inconsistency of age reporting
  - Incomplete and missing data

1871 Census

ID	Last name	First name	Gender	Age	Birthplace	Marital status
r1	Bagg	**Addia**	female	**12**	Ontario	single
r2	**Pritchard**	Thomas	male	14	Ontario	single
r3	**Bambridge**	**M**	male	**36**	England	married

1881 Census

ID	Last name	First name	Gender	Age	Birthplace	Marital status
r4	Bagg	**Adelia**	female	**23**	Ontario	single
r5	**Pretchard**	Thomas	male	24	Ontario	single
r6	**Bambrige**	**Martin**	male	**45**	England	married

Data Challenges

- Small number of attributes leads to extensive duplication
- Records end up being extremely similar

1871 Census

ID	Last name	First name	Gender	Age	Birthplace	Marital status
r1	Barns	Mary	female	11	Ontario	single
r2	Barns	Mary	female	9	Ontario	single
r3	Barns	Mary	female	8	Ontario	single
r4	Barns	Mary	female	12	Ontario	single
r5	Barns	Mary	female	10	Ontario	single
r6	Barns	Mary	female	10	Ontario	single

1881 Census

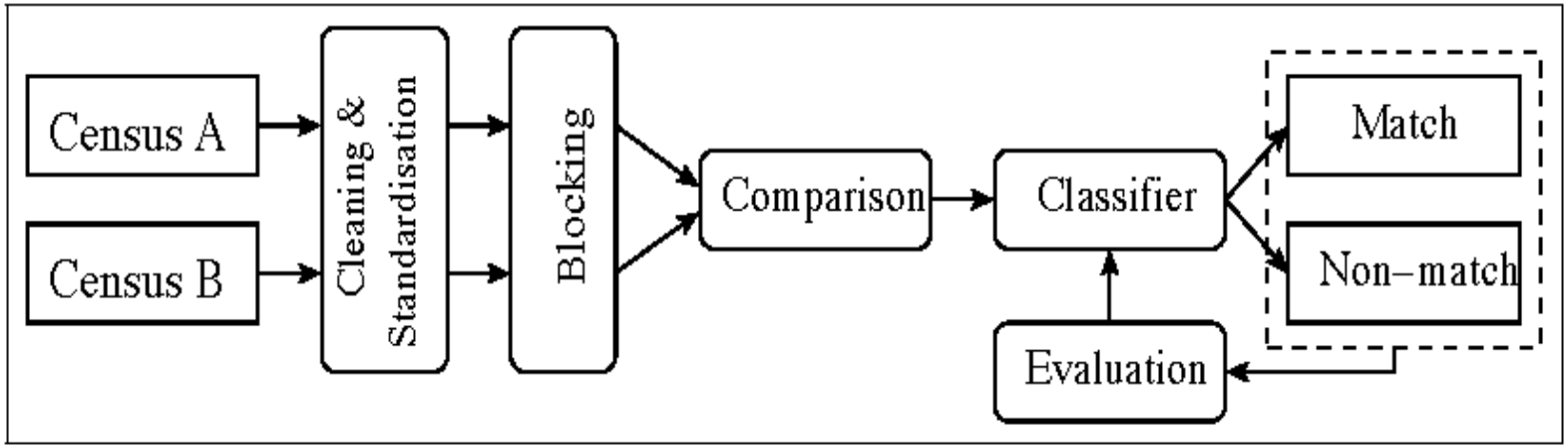
ID	Last name	First name	Gender	Age	Birthplace	Marital status
r7	Barns	Mary	female	20	Ontario	single
r8	Barns	Mary	female	22	Ontario	single

Classification

Our matching problem is a binary classification problem.

- Each pair of records represented by a feature vector $\phi_{(a,b)}$
- Classification model - Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel (LIBSVM package)
- Classes: match (positive class) and non-match (negative class)

Record Linkage System



Feature Construction and Comparison

- We use the following attributes to generate features that reflect record-pair similarity:

Tag	L	F	GD	AGE	BP	MS
Attr.	last name	first name	gender	age	birthplace	marital status
Type	string	string	binary	integer	categorical	categorical

- Social science analysis restricts linking features to time-invariant characteristics

  Ignore: location, occupation, family context and relationship, religion, etc.

Evaluation

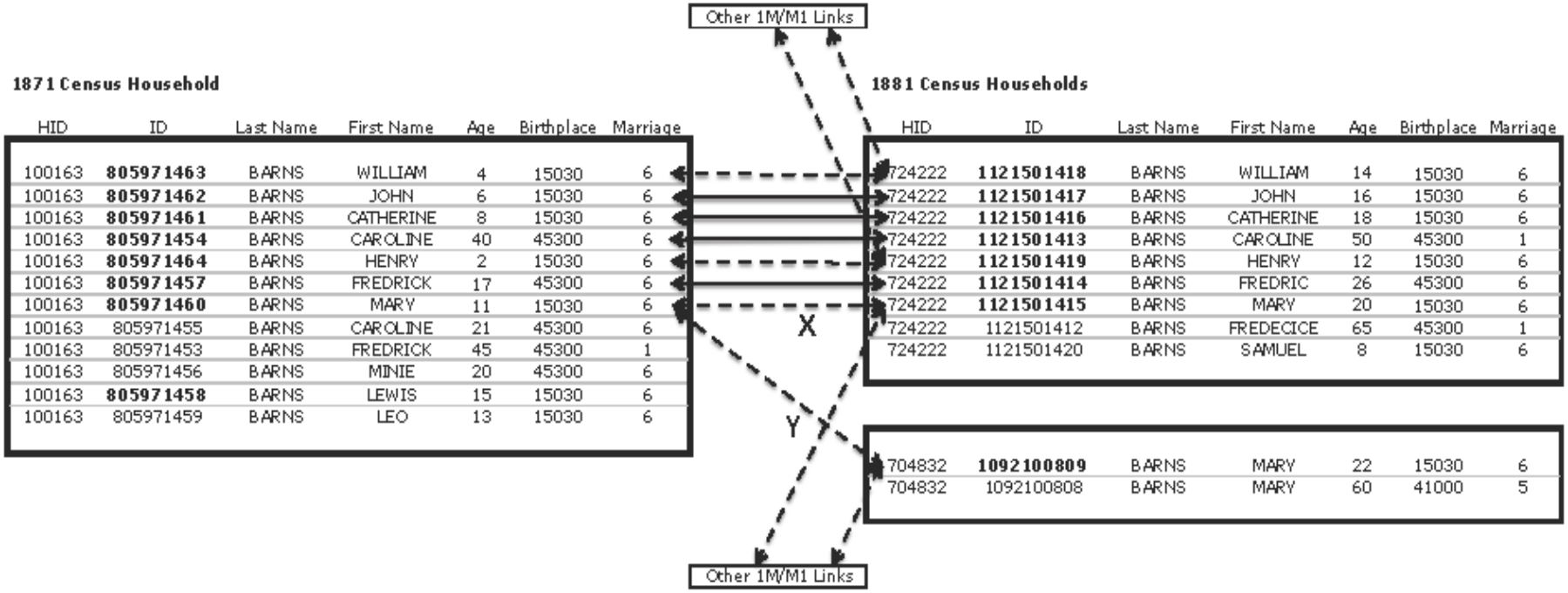
- Historical sensibility implies minimization of false positive errors

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = \mathrm{Recall} \quad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}} = 1 - \mathrm{Precision}$$

- Full Linkage System Evaluation Estimates - 5 Fold Cross Validation - Mean (std. dev.)

	True Links	TP	FP	FN	TPR	FPR
Mean	1,708.6	684.8	36.0	1,023.8	40.1%	5.0%
Std. Dev.	45.1	38.4	9.6	24.1	1.5%	1.3%

1871-1901 - Disambiguate with family information



1871 Census Household

HID	ID	Last Name	First Name	Age	Birthplace	Marriage
100163	**805971463**	BARNS	WILLIAM	4	15030	6
100163	**805971462**	BARNS	JOHN	6	15030	6
100163	**805971461**	BARNS	CATHERINE	8	15030	6
100163	**805971454**	BARNS	CAROLINE	40	45300	6
100163	**805971464**	BARNS	HENRY	2	15030	6
100163	**805971457**	BARNS	FREDRICK	17	45300	6
100163	**805971460**	BARNS	MARY	11	15030	6
100163	805971455	BARNS	CAROLINE	21	45300	6
100163	805971453	BARNS	FREDRICK	45	45300	1
100163	805971456	BARNS	MINIE	20	45300	6
100163	**805971458**	BARNS	LEWIS	15	15030	6
100163	805971459	BARNS	LEO	13	15030	6

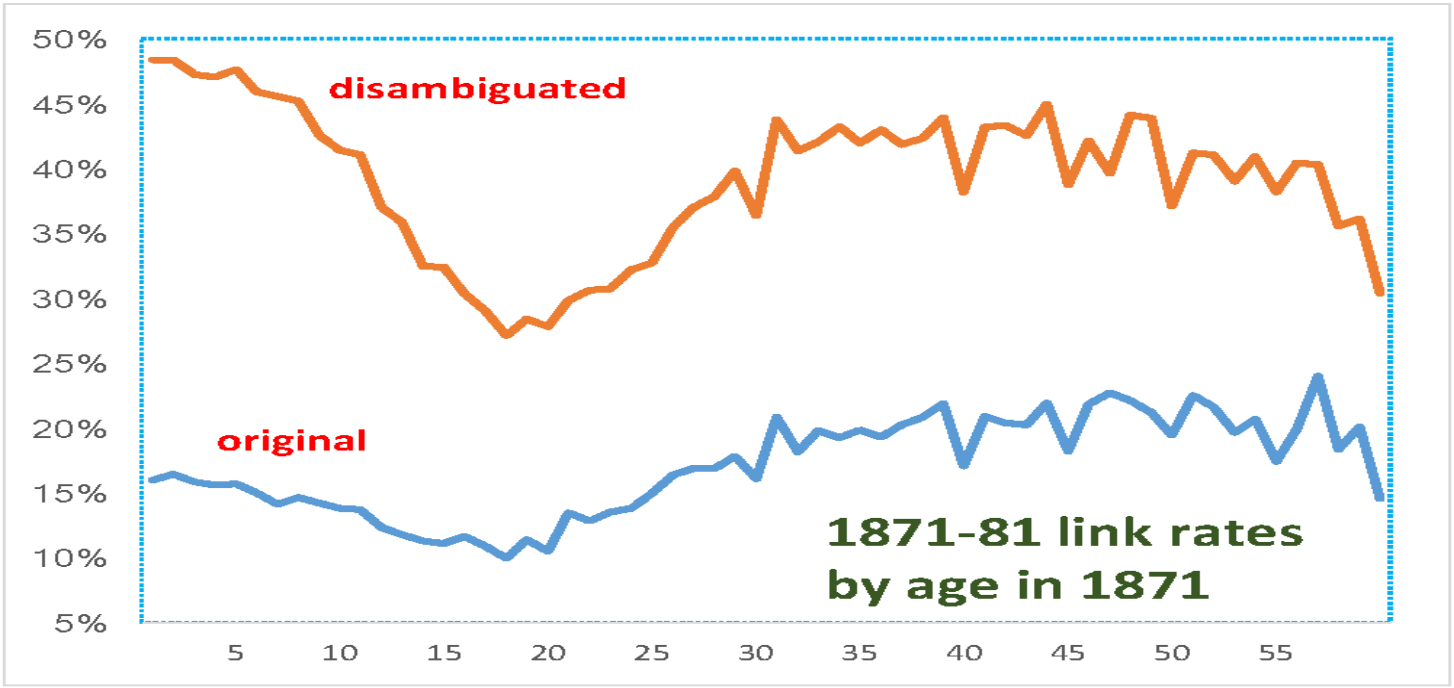
1881 Census Households

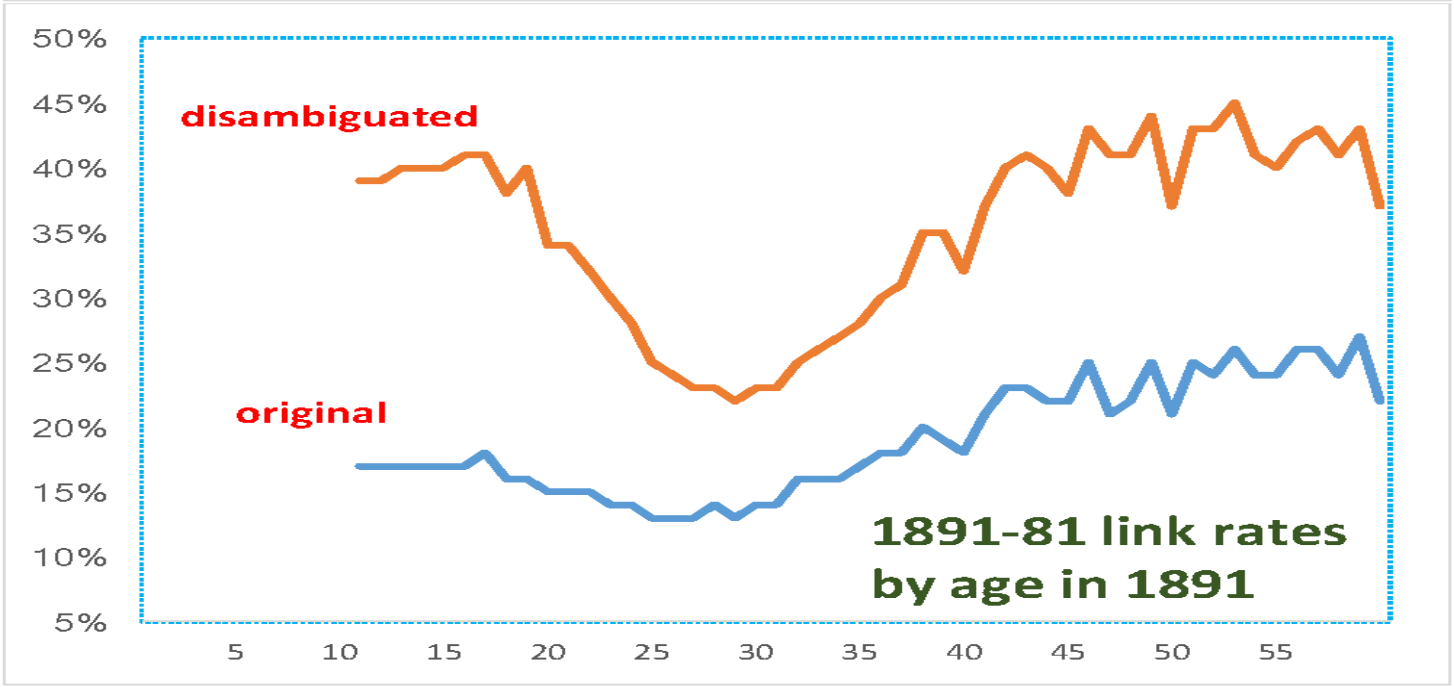
HID	ID	Last Name	First Name	Age	Birthplace	Marriage
724222	**1121501418**	BARNS	WILLIAM	14	15030	6
724222	**1121501417**	BARNS	JOHN	16	15030	6
724222	**1121501416**	BARNS	CATHERINE	18	15030	6
724222	**1121501413**	BARNS	CAROLINE	50	45300	1
724222	**1121501419**	BARNS	HENRY	12	15030	6
724222	**1121501414**	BARNS	FREDRIC	26	45300	6
724222	**1121501415**	BARNS	MARY	20	15030	6
724222	1121501412	BARNS	FREDECICE	65	45300	1
724222	1121501420	BARNS	SAMUEL	8	15030	6
704832	**1092100809**	BARNS	MARY	22	15030	6
704832	1092100808	BARNS	MARY	60	41000	5

1871-1901 - Disambiguate with family information

- We aim to disambiguate multiple links (i.e. a record linked to several records)
- We use a set similarity measure when comparing households that have ambiguous links and we disambiguate based on this score

Decade	Number	Linkage Rate
1871-1881	1,103,713	31.84%
1881-1891	1,209,865	28.28%
1891-1901	1,265,998	26.44%


50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
disambiguated
original
1871-81 link rates
by age in 1871
5
10
15
20
25
30
35
40
45
50
55


50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
disambiguated
original
1891-81 link rates
by age in 1891
5
10
15
20
25
30
35
40
45
50
55

Average link rates by other characteristics				
	1871-1881		1881-1891	
	original	original + disambiguated	original	original + disambiguated
Canada	0.15	0.38	0.18	0.35
BC			0.11	0.14
Manitoba			0.14	0.24
New Brunswick	0.19	0.37	0.26	0.39
Nova Scotia	0.18	0.42	0.24	0.42
Ontario	0.15	0.41	0.18	0.35
PEI			0.26	0.43
Quebec	0.13	0.35	0.17	0.33
Unorg Territor			0.10	0.18
Gender				
female	0.14	0.36	0.17	0.33
male	0.17	0.41	0.19	0.37
Ethnicity				
french-cdn	0.12	0.34	0.16	0.32
other	0.16	0.40	0.19	0.36

Average link rates by other characteristics

	1871-1881		1881-1891	
	original	original + disambiguated	original	original + disambiguated
Religion				
Catholic	0.12	0.34	0.16	0.32
other	0.18	0.41	0.20	0.37
Literacy				
can read	0.18	0.38	0.19	0.36
not read	0.14	0.40	0.16	0.29
Birthplace				
born Canada	0.15	0.39	0.19	0.36
born other	0.17	0.34	0.18	0.27
Marital status				
single	0.13	0.38	0.17	0.38
married/div	0.19	0.39	0.20	0.32

Marginal impact of individual characteristics on the odds of establishing a link

- simple logistic regression
- original versus original + disambiguated
- 1871-81 vs 1881-91

Does disambiguation increase/decrease bias in the sense that the odds ratio moves decisively (.2 or more) away/towards 1.0, or does it not change?

1871 - 1881

	Odds Ratio	z		Odds Ratio	z
Ontario					
male	1.20	9.18		1.26	16.04
single	0.60	-15.03		0.60	-19.97
old	0.69	-9.31		0.53	-20.17
young	0.91	-2.85		1.22	7.45
french	0.80	-4.29		0.35	-25.65
histatus	0.84	-1.99		0.68	-5.75
cdnbp	0.96	-1.41		1.36	16.02
Quebec					
male	1.35	12.72		1.28	14.71
single	0.74	-7.47		0.72	-11.43
old	0.75	-5.90		0.54	-16.27
young	0.87	-3.46		1.21	6.73
french	0.64	-15.04		1.01	0.65
histatus	0.92	-0.92		0.73	-4.44
cdnbp	1.74	10.74		1.55	11.42
Atlantic					
male	1.29	8.86		1.30	11.54
single	0.64	-9.68		0.67	-10.53
old	0.82	-3.49		0.59	-11.23
young	0.99	-0.28		1.14	3.34
french	0.61	-9.70		0.51	-17.50
histatus	0.91	-0.85		0.69	-3.78
cdnbp	2.09	13.03		1.69	12.97

Odds ratios for establishing a link, 1891 all

	Original links		Original + Disambiguated	
	Odds Ratio	z	Odds Ratio	z
Ontario				
male	1.10	5.23	1.20	12.39
single	**0.94**	**-2.58**	**1.38**	**16.53**
old	**1.79**	**20.74**	**2.01**	**27.95**
young	0.83	-7.49	0.88	-6.23
illit	0.66	-9.61	0.62	-13.77
fren	**1.03**	**0.59**	**0.51**	**-15.03**
histatus	1.04	0.65	0.90	-2.25
cdnbp	**1.03**	**1.39**	**1.53**	**22.96**
Quebec				
male	1.34	12.31	1.32	14.93
single	**1.11**	**3.27**	**1.66**	**20.29**
old	1.74	15.88	1.62	15.71
young	0.74	-9.28	0.85	-6.25
illit	0.81	-7.10	0.79	-10.20
fren	**0.68**	**-13.1**	**1.01**	**0.21**
histatus	0.87	-1.92	0.76	-4.72
cdnbp	1.96	13.05	2.02	15.98
Atlantic				
male	1.27	9.18	1.30	11.29
single	**0.97**	**-0.78**	**1.42**	**11.06**
old	1.44	9.39	1.51	11.45
young	0.89	-3.33	0.87	-4.41
illit	0.82	-5.20	0.78	-7.15
fren	0.65	-8.75	0.56	-13.81
histatus	1.06	0.75	0.96	-0.49
cdnbp	1.83	11.41	1.82	13.58

Manitoba				
male	1.07	1.12	1.05	1.00
single	**1.03**	**0.38**	**1.25**	**3.59**
old	2.66	9.40	2.60	9.77
young	0.88	-1.71	0.91	-1.60
illit	0.79	-1.73	0.62	-3.92
fren	1.01	0.11	0.85	-1.77
histatus	0.89	-0.59	0.87	-0.92
cdnbp	**2.04**	**10.78**	**2.83**	**18.75**
West				
male	0.90	-1.34	0.86	-2.42
single	0.92	-1.08	0.94	-0.96
old	**2.50**	**7.49**	**3.05**	**10.25**
young	1.27	3.21	1.30	4.12
illit	0.18	-12.5	0.12	-16.55
fren	0.29	-3.67	0.21	-4.99
histatus	1.07	0.40	1.10	0.64
cdnbp	**2.28**	**12.04**	**2.77**	**17.12**

Question: Does disambiguation increase or decrease bias at the margin?

Answer: It depends on the period, region and individual characteristic

Conclusions:

1. By and large linked data are not fully representative but extent of bias varies link with method, characteristic, region and period
2. Adolescents, whom we examine, have the lowest linkage rates, with or without disambiguation
3. 'Price' in terms of added bias of increasing linked sample size through disambiguation is tolerable
4. Bias generally is smaller *within* demographic groups >> our focus on 'first job' is an example of disaggregate approach

Occupational matrix, 1871-81

Heads, 1871 / Youths, 1881	White-collar	Farm	Skilled /semi-skilled	Unskilled	Total
White-collar	296 (.48)	215 (.05)	269 (.18)	97 (.10)	877
Farm	83 (.14)	3222 (.75)	285 (.19)	253 (.25)	3843
Skilled/ semi-skilled	171 (.28)	371 (.09)	713 (.48)	241 (.24)	1496
Unskilled	64 (.10)	471 (.11)	207 (.14)	424 (.24)	1166
Total	614	4279	1474	1015	

Occupational matrices, 1881-91

Youths, 1881 \ Heads, 1871	White-collar	Farm	Skilled /semi-skilled	Unskilled	Total
White-collar	530 (.53)	354 (.08)	436 (.23)	86 (.09)	1406
Farm	87 (.08)	3222 (.70)	183 (.10)	221 (.22)	3713
Skilled/ semi-skilled	256 (.26)	552 (.12)	1102 (.58)	374 (.38)	2284
Unskilled	123 (.12)	507 (.11)	190 (.10)	305 (.31)	1125
Total	996	4635	1911	986	

Measures of occupational inheritance

- M – sum of off diagonal shares
- U – upward mobility out of unskilled to skilled, white-collar
- D – downward mobility out of white-collar to skilled, unskilled
- OF – movement off the farm

Improved measures of occupational inheritance

- Deming and Stephan (1940) algorithm to adjust marginal frequencies
- M’, U’, D’, OF’ after controlling for difference in prevalence of jobs

The *Altham Statistic*: a summary measure of occupational inheritance

- "Distance" between matrices P & Q

$$d(P,Q) = \left[\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{l=1}^{r}\sum_{m=1}^{s} |log\left(\frac{p_{ij}p_{lm}q_{im}q_{lj}}{p_{im}p_{lj}q_{ij}q_{lm}}\right)|^2\right]^{1/2}$$

Using the Altham Statistic

- d(P,Q) compare 2 populations (distance in mobility between the two)
- d(P,J), d(Q,J) : compare each population to matrix of 1s (distance from independence)
- If $d(P,Q) \neq 0$ and $d(P,J) > d(Q,J)$, more mobility in Q than in J

Inheritance Comparisons 1

	(i) Full linked sample			
	M	U	D	OF
1871-81	.37	.33	.38	.25
1881-91	.40	.47	.38	.31
	M’	U’	D’	OF’
1871-81 adjusted	.40	.44	.38	.32

Inheritance Comparisons 2

	(ii) Anglophone			
	M	U	D	OF
1871-81	.37	.24	.40	.27
1881-91	.40	.41	.40	.33
	M’	U’	D’	OF’
1871-81 adjusted	.41	.34	.39	.34
	(iii) Francophone			
	M	U	D	OF
1871-81	.38	.40	.49	.26
1881-91	.39	.55	.36	.27
	M’	U’	D’	OF’
1871-81 adjusted	.41	.47	.56	.32

Inheritance Comparisons 3

	(ii) Anglophone			
	M	U	D	OF
1871-81	.37	.24	.40	.27
1881-91	.40	.41	.40	.33
	M’	U’	D’	OF’
1871-81 adjusted	.41	.34	.39	.34
	(iv) Migrants			
	M	U	D	OF
1871-81	.36	.36	.38	.25
1881-91	.40	.47	.38	.31
	M’	U’	D’	OF’
1871-81 adjusted	.39	.49	.31	.29

Altham Statistics 1

	d(P,J)	G2	d(Q,J)	G2	d(P,Q)	G2
1871-81	20.6	3044***				
v. 1881-91			21.5	4000***	5.8	48.2***

Altham Statistics 2

	d(P,J)	G2	d(Q,J)	G2	d(P,Q)	G2
Anglo 1871-81	21.6	904***				
v. Franco 1871-81			19.0	744***	8.6	39***
v. Migrant 1871-81			21.4	1425***	4.4	16*
Anglo 1881-91	20.6	1501***				
v. Franco 1881-91			24.4	448***	6.8	9.5
v. Migrant 1881-91			22.2	2049***	3.8	19*

Altham Statistics 3

	d(P,J)	G2	d(Q,J)	G2	d(P,Q)	G2
Canada 1871-81	20.6	3043***				
v. US 1850-80			12.6	333***	11.0	133***

Inheritance in particular occupations

- How much inheritance in major occupations?
- Relative to “expected inheritance?”

- Compare

(1) $\frac{Head\ occ_i \cup Youth\ occ_i}{Youth\ occ_i}$ and (2) $\frac{Head\ occ_i}{\sum Head\ occ_i}$

Occupational inheritance 1871-81 (N>100)

	(1) $\frac{Head\ occ_i \cup Youth\ occ_i}{Youth\ occ_i}$	(2) $\frac{Head\ occ_i}{\sum Head\ occ_i}$	(1) / (2)
Woodsman	0.537	0.018	29.3
Boot & shoe maker	0.340	0.020	16.7
Day labourer	0.296	0.039	7.5
Carpenter	0.295	0.039	7.5
Blacksmith	0.271	0.019	14.5
Labourer	0.257	0.069	3.7
Dealer	0.252	0.029	8.8

Occupational inheritance 1881-91 (N>100)

	(1) $\frac{Head\ occ_i \cup Youth\ occ_i}{Youth\ occ_i}$	(2) $\frac{Head\ occ_i}{\sum Head\ occ_i}$	(1) / (2)
Woodsman	0.426	0.012	35.6
Boot & shoe maker	0.458	0.020	23.1
Day labourer	0.255	0.017	14.6
Carpenter	0.256	0.042	6.1
Blacksmith	0.202	0.018	11.4
Labourer	0.249	0.072	3.5
Dealer	0.206	0.037	5.6

Conclusions

- Mostly descriptive tour of data so far…
- But significant group differences in inheritance/persistence
- Francophones more upward and downward movement
- Migrants more upward movement
- Sure looks like Canada had less movement than US… so far

Conclusions

- Canada saw less movement into non-paternal occupation than the US?
- Intergenerational persistence in specific occupations was a thing.
- Corak and Piriano (2011): 5.6% of adults employed in same firm as father, v. .041% with full random assignment
- Our ratios for select occupations are lower than that

Next steps

- multinomial logit regressions of second-generation occupations.
- occupational task data as another way to think about distance?
- Inter v. intra generational movement
- Track 16-25 year olds in 1881 to 26-35 in 1891: M = .18, U = .18, D = .23, OF = .12

Thank you!

Chris (@chris__minns), Fraser, Kris, and Luiza