

CANCER AND THE HEALTHY IMMIGRANT EFFECT: PRELIMINARY ANALYSIS USING THE CENSUS COHORT

Ted McDonald, Mike Farnworth, Zikuan Liu
Department of Economics
University of New Brunswick

CRDCN conference – October 3, 2013



Background

- Cancer is the leading cause of death in Canada, with on average 21 new cases diagnosed every hour. About one in four Canadians will die from the disease
- Davis, Donovan, and Herberman (2007) estimate that less than 10% of cancers are the result of a mutation in, or the operation of, a particular gene. The rest are related to individual behaviors, environment, and the interaction of genes and these dimensions
- There has been a huge volume of research conducted on the contextual determinants of cancer
 - Demographic: age, sex, race, immigrant status, marital status
 - Socioeconomic: education, household and personal income, occupation, employment status, poverty
 - Geographic: urban/rural, neighborhood effects, environment, access to services
 - Health behaviors: smoking, diet, alcohol

Data sources for cancer research

- Previous work has used a variety of data sources and each type of data has strengths and limitations
- Case control studies such as chart reviews
 - Small samples, often non-random selection
- Survey data such as CCHS (Canada) and NHIS (US)
 - Self reported
 - Cross-sectional
 - Small numbers of cancers even in large surveys
 - Prevalence rather than incidence

- Administrative cancer registry data such as CCR (Canada) and SEER (US)
 - Gold standard for data on cancer incidence
 - Only basic demographic information is reported – sex, age, place of residence, and race/ethnicity in the US
 - No information on the broader at-risk population, which must come from area level population data from other data sources
 - No data on individual socioeconomic status – area level data such as income quintile from the Census or other data source must be linked via patient postal code
 - Place of residence is as of the date of diagnosis

- **Linked administrative/survey datasets**
- **US SEER-National Longitudinal Mortality Study linked dataset that includes individual data on socioeconomic status.**
 - Clegg et.al. (2009): Lower cancer incidence rates with higher education levels for men; for women, cancer incidence rates were lower for university degree holders than other education levels
- **Canadian IMDB-CCR-CMD immigrant linked dataset**
 - McDermott et.al (2011): immigrants from most regions of birth have lower standardized incidence rates for most forms of cancer
 - No native-born, immigrants who arrived 1980-98 only, no socioeconomic information

The 1991 Canadian Census Cohort

- A new dataset assembled by Statistics Canada is now available through a pilot program in the Canadian RDC network
- It is based on individuals aged 25+ who completed the 20% long form of the 1991 Census of Canada, linked to:
 - Canadian Cancer registry from 1984-2003 [2008 next year]
 - Canadian Mortality database from 1991- 2006 [2008 next year]
 - Tax file data on location of residence and marital status 1986-2008
- Individual level data are linked using probabilistic linkage methods (Peters et.al, Int.J. Epid, 2013)

Information in the dataset

- From the Census we can observe:
 - Educational attainment, field of study as of 1991
 - Personal and household income from different sources as of 1991
 - Occupation as of 1991
 - Immigrant status, country of birth, year of arrival, age at arrival
 - Visible minority status, aboriginal status
 - Mother tongue and language proficiency as of 1991

- From the CCR we can observe:
 - Date of diagnosis of cancer
 - Cancer site(s)
- From the CMD we can observe:
 - Date of death
 - Cause of death
- From the tax file information we can observe:
 - Six digit postal code of residence for *each* year in which a tax return was filed from 1986-2008
- 2.7 million individuals could be linked to at least one tax return, including ~500,000 immigrants. 247,000 cases of cancer diagnosed in this cohort from 1991.

Limitations of this linked dataset

- Probabilistic linkage is incomplete (80% of records could be linked)
- Some people do not file income tax and so are lost to follow-up
- No information on health behaviors
- No update of socioeconomic information after 1991
- No information on individuals who were not in Canada as of the 1991 census date (newer immigrants, returning Canadian residents)

Preliminary Research Questions

- How does the likelihood of being diagnosed with cancer vary by immigrant status, year of arrival, and country of birth?
- => Is there evidence of a healthy immigrant effect for cancer?
- => Is such an advantage lost with time in Canada?
- Immigrants on average are better educated and much more likely to reside in larger urban areas

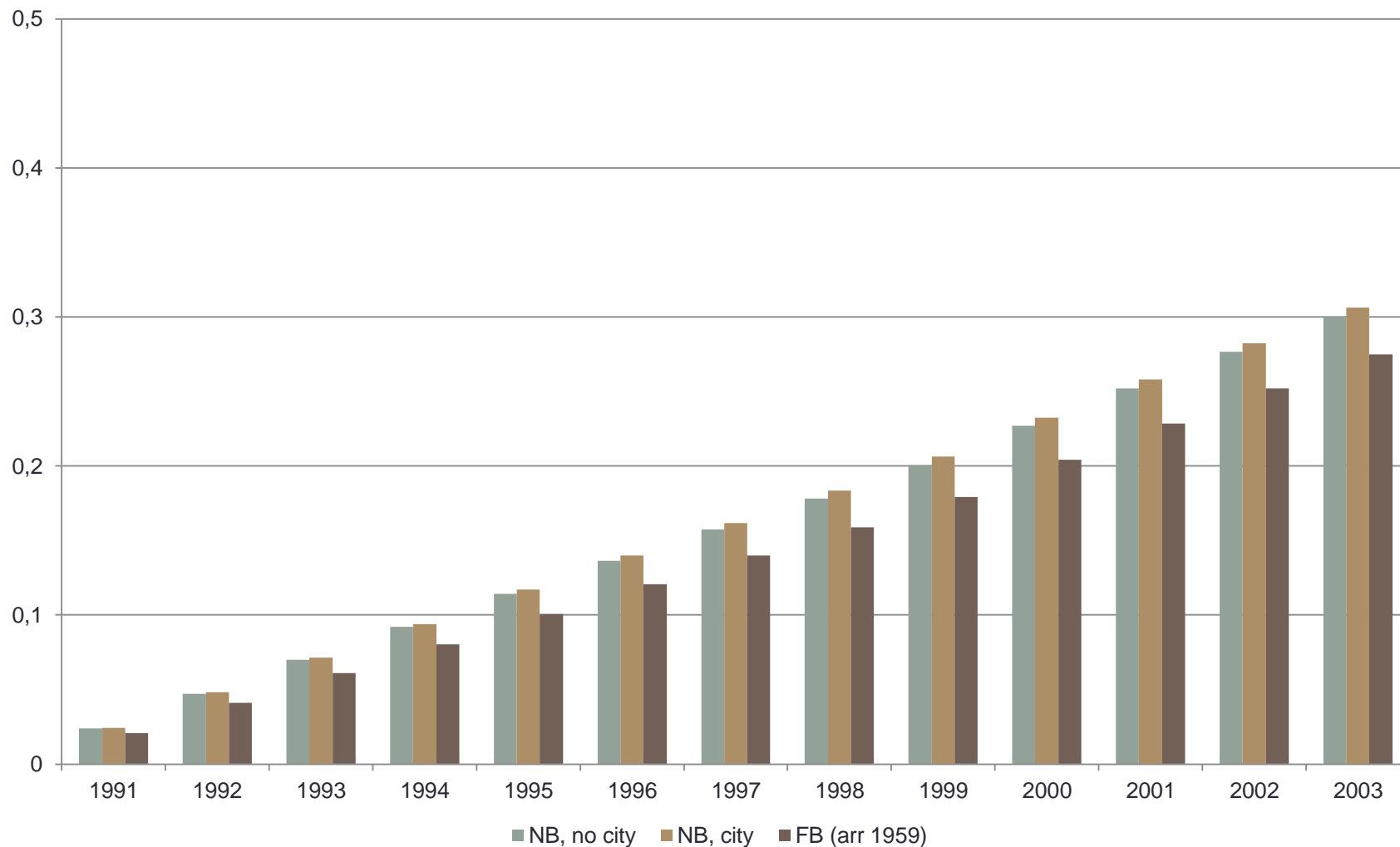
Methods

- Discrete time Logistic duration model
 - Time until diagnosis of cancer
 - Calendar years 1991-2003 as discrete intervals (so up to 13 person-years of observation per individual)
 - Dependent variable – (0/1) diagnosed with cancer in a given year, conditional on being in the sample during the year and not previously diagnosed with cancer
 - Censoring: death, no tax return filed, end of sample period
 - Clustering of individual-specific error terms

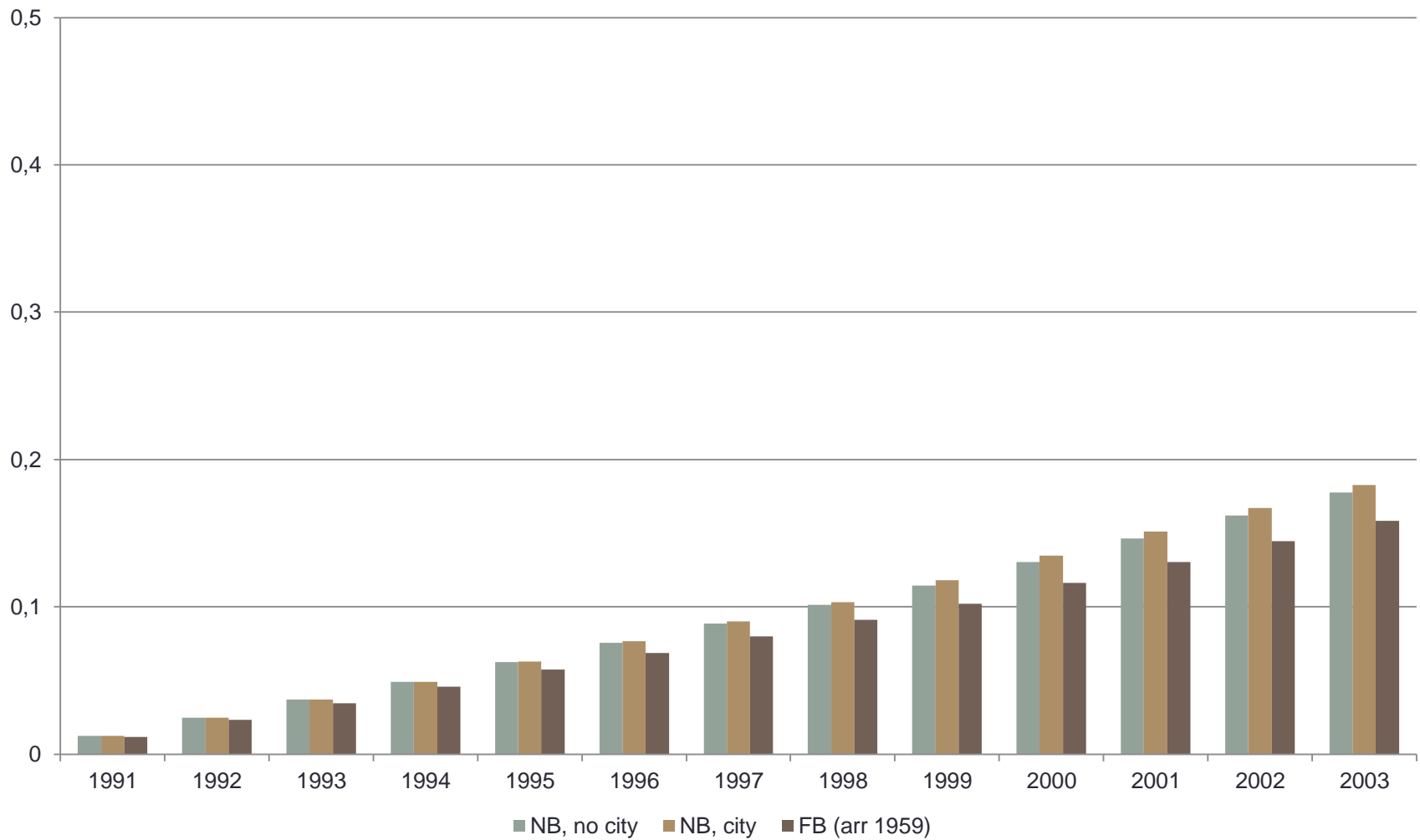
- Sample: aged 25-79 in 1991 in the Census cohort and not previously diagnosed with cancer
- Flexible specification of year/age/birth cohort effects
- Time invariant: education level, visible minority status, immigrant status; country of birth, period of arrival in Canada for immigrants
- Other time-varying covariates: resides in a large city, years in Canada if an immigrant (interacted with period of arrival)

Descriptive Statistics – age/time paths by birth cohort

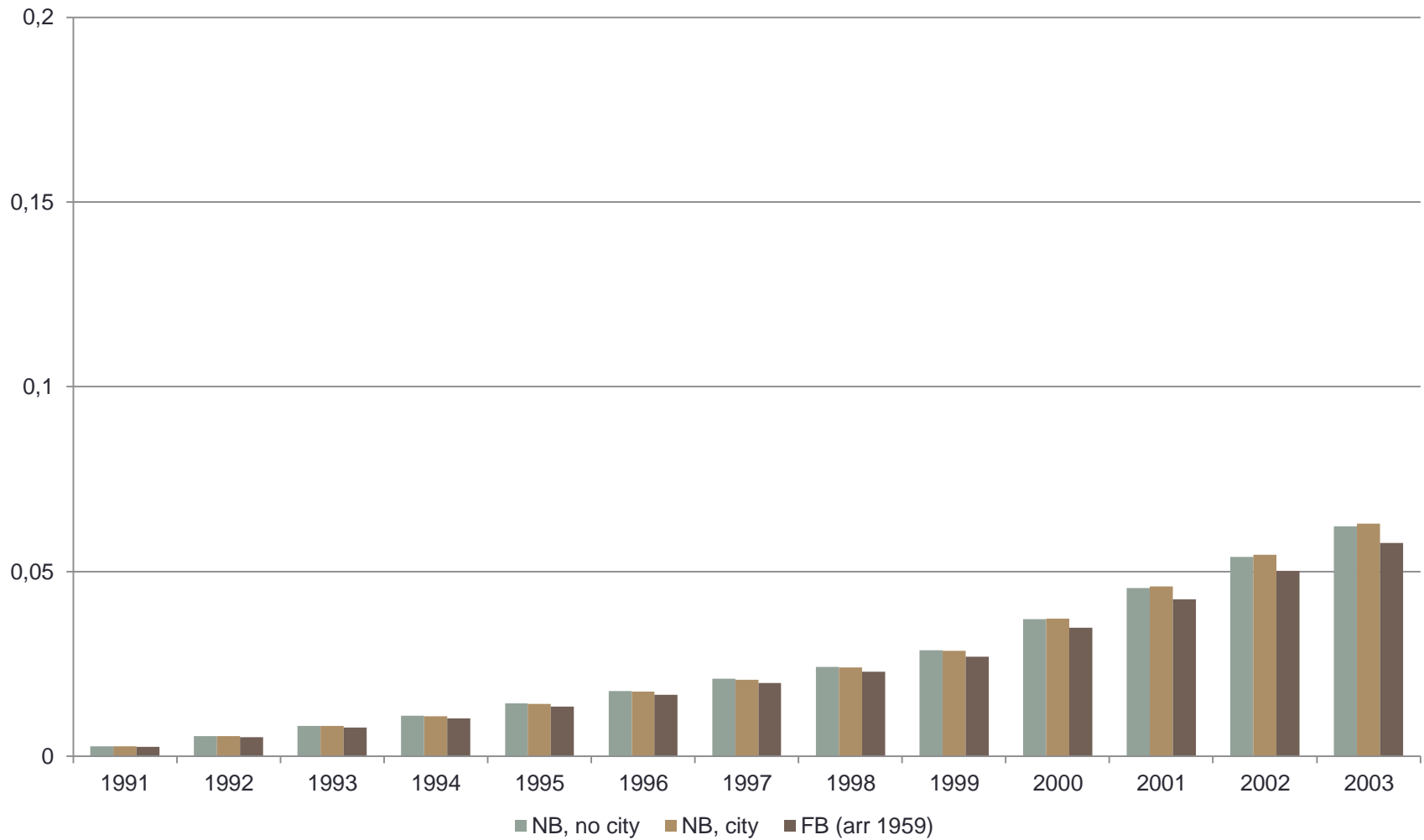
1a: Men Born in 1925 – failure rate



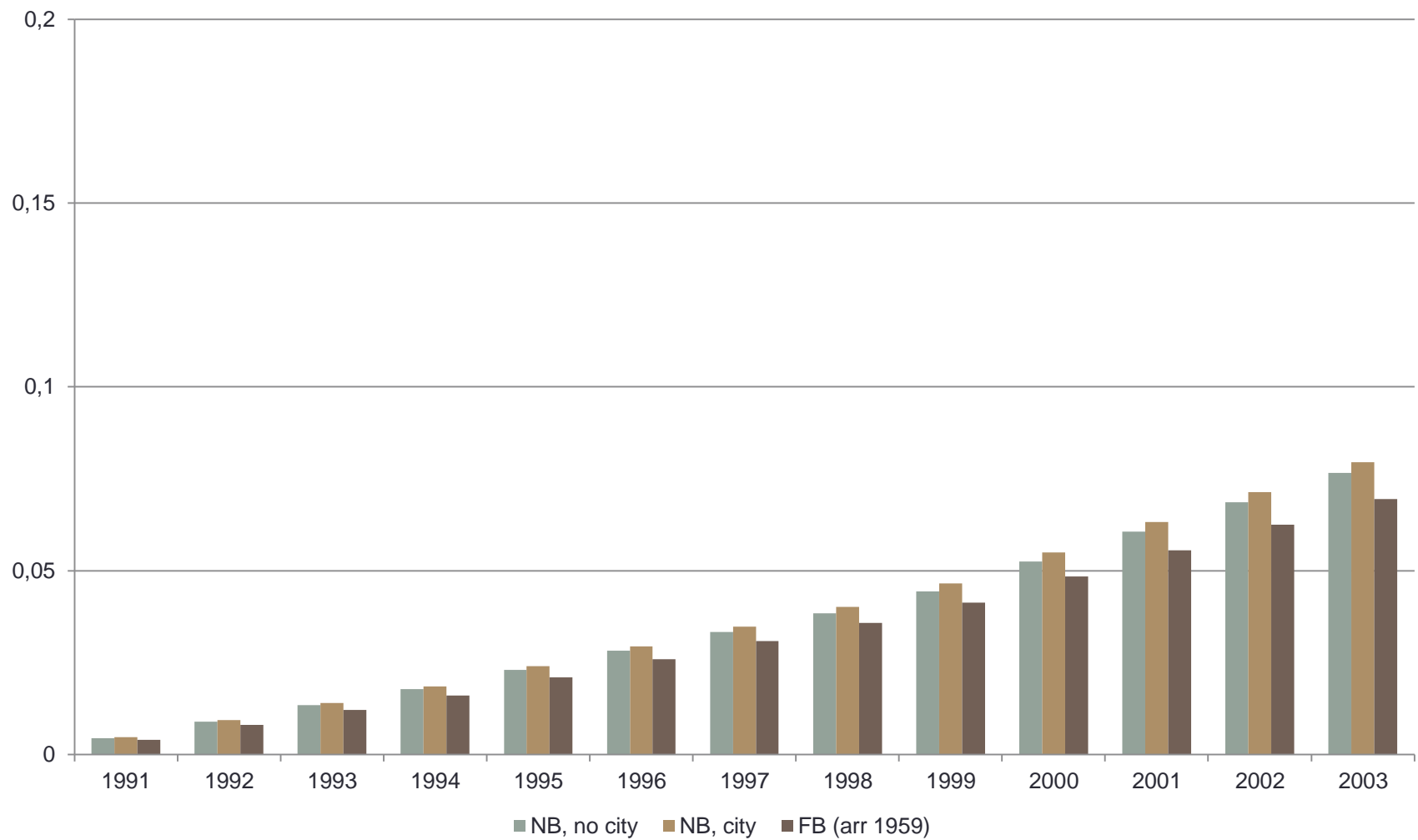
1b: Women Born in 1925 - failure rate



1c: Men Born in 1945 - failure rate



1d: Women Born in 1945 – failure rate



Full Duration model: odds ratios of selected covariates

	MEN		WOMEN	
	OR	p-val	OR	p-val
Not a large city	1.000		1.000	
Large city (pop>500,000)	1.052	0.000	1.058	0.000
Less than secondary school	1.000		1.000	
Secondary school	0.950	0.000	0.961	0.000
Trades certificate or diploma	0.937	0.000	0.968	0.000
Bachelor's degree	0.859	0.000	0.929	0.000
Higher degree	0.817	0.000	0.962	0.000
Born in Canada, white	1.000		1.000	
Born in Canada, other race	0.870	0.000	0.910	0.053
# individuals	1.23m		1.23m	
# observations	14.3m		14.5m	
Pseudo-Rsq	0.117		0.047	

Odds ratios - region of birth

	MEN		WOMEN	
	OR	p-val	OR	p-val
Born in UK/Ireland/Aus/NZ	1		1	
Born in the US	0.961	0.473	1.056	0.662
Born in Europe	0.905	0.052	0.849	0.002
Born in Western Asia/Mideast	0.693	0.000	0.783	0.004
Born in other Asian countries	0.628	0.000	0.709	0.000
Born elsewhere	0.839	0.001	0.771	0.000

Odds ratios – arrival cohort and ysm (UK born)

MEN	year arrived						
	<1930	1930-39	1940-49	1950-59	1960-69	1970-79	1980-90
YSM							
1-10							0.768
11-20						0.851	0.821
21-30					0.926	0.937	0.845
31+	1.104	1.077	1.002	1.005	0.986	0.945	
WOMEN	year arrived						
	<1930	1930-39	1940-49	1950-59	1960-69	1970-79	1980-90
YSM							
1-10							0.906
11-20						0.956	0.916
21-30					0.944	0.943	0.857
31+	1.172	0.936	1.051	1.034	0.976	0.897	

Conclusions

- Evidence of a healthy immigrant effect is found for cancer incidence among certain immigrants to Canada
 - The gap is widest for immigrants from East and South Asia;
 - The gap is wider for more recent male arrivals to Canada, and some evidence it narrows with additional years in Canada
 - Differences in SES, region of residence between immigrants and non-immigrants do not explain the immigrant cancer gap
- Canadian born visible minorities also have significantly lower cancer incidence, so advantage appears to persist into the second generation

Next steps

- Estimation of duration models for specific types of cancer
- Unobserved heterogeneity and robustness checks
- What are immigrant characteristics actually reflecting in terms of causes of cancer?
 - Link group-specific information on health behaviors from CCHS and other health survey datasets