A Tale of (More Than ?) Two Cohorts – from Canada

By **Dr. John Frank, Scientific Director,**
CIHR-Institute of Population & Public Health

Professor, Dept. of Public Health Sciences,
University of Toronto

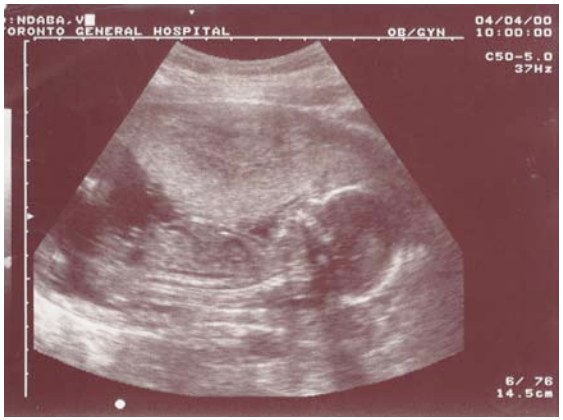
Senior Scientist, Institute for Work and Health,
Toronto

November 16 -19, 2005

CIHR IRSC

Canada

Why study gene-environment interactions?

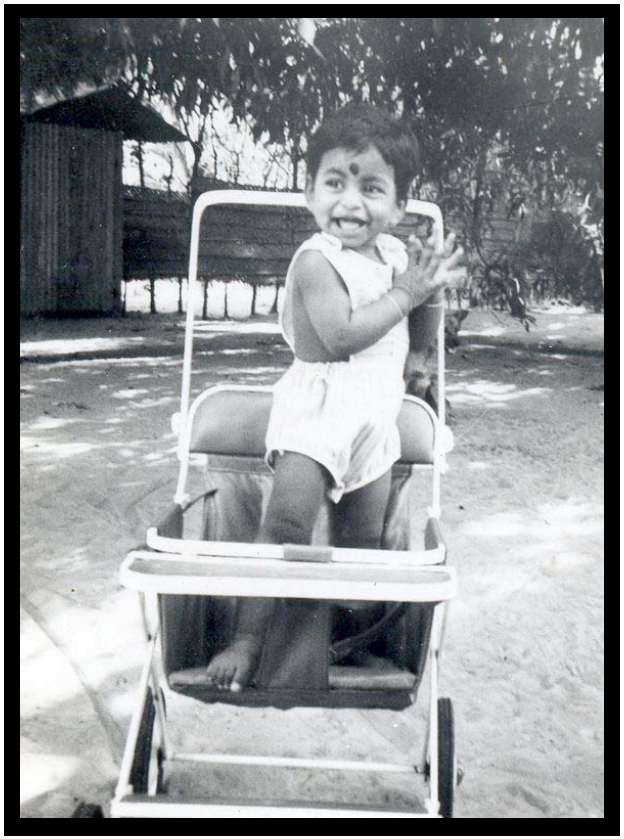


- Most disease burden is jointly determined by interaction of individual genetic endowments and complex sequence of environmental factors
- These gene-environment interactions require decades to fully manifest over the life course
- Diseases and conditions of later life occur in some and not others because of intense interactions between particular genetic constitutions and particular sequence of social and physical environments

CIHR IRSC

Canada

Why study gene-environment interactions? cont'd



- BUT…little is known about underlying causes of these conditions and why they are now increasing in frequency – for e.g. asthma
- Requires study of these sequential events in large numbers of people over time, on whom baseline genetic and repeated environmental exposures are taken, to:
  - understand the causal pathways; and,
  - develop disease prevention strategies

Studying Genetic and Environmental Contributions to Disease Causation: An Uneven Playing Field

Measurement Attribute	Genetic Exposure Measures	Environmental Exposure Measures
Time-varying?	No – one sample per lifetime is enough (unless gene expression arrays are used)	Yes – new samples needed whenever exposure changes
Data Collection Costs	Cheap (on a sample)	Expensive (real-time assays)
Sample Storage (for later analysis)	Easy (buccal swab, buffy coat)	Difficult (e.g. air/water/diet samples)
Data Analysis Costs	Getting cheaper by the day	Getting Costlier (as awareness of chemical/physical/biological complexity increases)
Overall Ease & Cost of Accurate Ascertainment	**Easy / Cheap**	**Difficult / Costly**

Comparison of "Huge, Data-Thin" Cohorts (e.g. U.K. BioBank) And "Small, Data-Thick" Cohorts (e.g. Southampton)

Cohort Attribute	Huge – Thin	Small – Thick
Cost Per Subject due to:	Low (e.g. < $500. / data-wave)	High (if > $1,000. / data-wave)
Sample Size due to choice of:	500,000+	< 30,000
Exposures	Cheap-to-collect/store measures – e.g. genetic	Expensive, balanced mix of environmental and genetic measures
Outcomes	Cheap-to-collect administrative data – e.g. hospitalizations for diagnoses/deaths (dichotomous) → ↑ **SS.**	Expensive, directly measured bi-chemical physiologic, imaging, functional outcomes (often continuous) → ↓ **SS.**
Leading "Exposure-Measure Bias"	Large environmental exposure error >> genetic factor errors	"Better balanced errors" for environmental versus genetic factors
Leading to:	**Biased main effects and interaction results**	**Less biased results**