Institute for Fiscal Studies

How social scientists have used biological data: some case studies using ALSPAC

Claire Crawford

Why are social scientists interested in biological data?

- As outcomes:
  - Growing interest in determinants of health behaviours and outcomes
    - e.g. socio-economic differences in a range of health outcomes ,or engagement in risky behaviours (smoking, drinking, drug use), especially among young people
  - Potentially more accurate than self-reported measures
    - e.g. recall error or deliberate mis-reporting out of embarrassment or to “look cool”
- As sources of exogenous variation (Mendelian randomisation):
  - Identifying the causal impact of particular characteristics or behaviours is challenging because they are not randomly allocated
  - Growing awareness of genetic influences → potential use of genetic markers as instruments for these characteristics or behaviours
    - e.g. body mass or substance abuse

Why ALSPAC?

- Rich biological and social information from multiple time points
  - Started following mothers during pregnancy, so can consider ante-natal/in utero factors; cohort members now in their early 20s
  - Has followed up cohort members and their families regularly using self-report questionnaires and via clinic sessions
  - Is at the forefront of linkage to administrative data, both biological and socio-economic (e.g. education and health records)
- Potential downside: not nationally representative

Case study 1: biological data as an outcome

- Month of birth differences in engagement in risky behaviours
  - Well known that children born at start of academic year tend to do better in exams than those born at end of academic year
    - e.g. Crawford et al (2007, 2011) for England
  - Growing awareness of differences in non-cognitive skills . . .
    - e.g. Crawford et al (2011); Chen et al (2013); Muhlenweg (2010)
  - . . . and engagement in risky behaviours too
    - e.g. Argys & Rees (2006); Crawford et al (2011); Landerso et al (2013)

What can biological data add?

- Most studies tend to rely on self-reported measures of engagement in such behaviours
  - e.g. How many cigarettes do you usually smoke per day? When was the last time you got drunk? Have you ever taken drugs?
- Responses vary depending on mode of interview, who else is in the room when they respond; may also be subject to recall bias
- Biological data offers the possibility of confirming responses
  - e.g. cotinine levels for smoking behaviour

Ongoing work: Crawford, Greaves & Parey (2013)

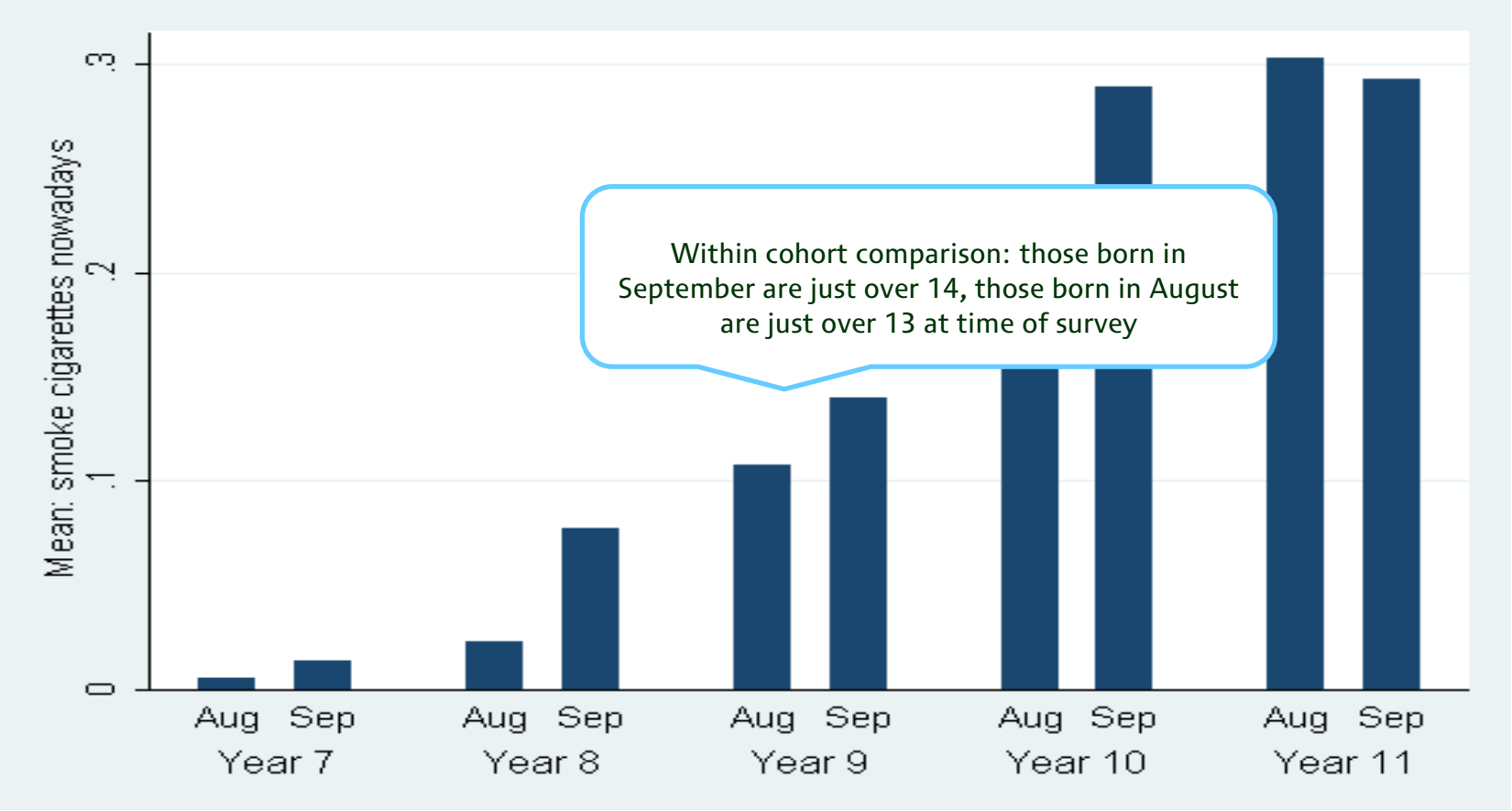
- Smoking, Drinking and Drug Use survey (SDD)
  - Random sample of secondary school pupils (approx. 6,000/ year) in England, Wales, Scotland; clustered at school level
  - Repeated cross section: biannual between 1982 and 1998 and annual since then (1988 onwards available on ESDS)
  - Questionnaire with emphasis on smoking and drinking
  - Supplemented with cotinine sample and diary
- ALSPAC
  - Approx. 14,000 kids born 1 April 1991 to 31 December 1992 in Avon
    - Split across 3 academic cohorts -> 2 discontinuities in year group
  - Multiple questionnaires and biological samples (not yet available) for both cohort members and their parents (especially mothers)
    - Plus rich background characteristics and links to educational attainment

Summary statistics from SDD (1988-99 England and Wales only)

	ALL	YEAR 7	YEAR 8	YEAR 9	YEAR 10	YEAR 11
SMOKING FREQUENCY						
EVER	0.42	0.15	0.30	0.43	0.55	0.65
ONCE	0.17	0.11	0.18	0.20	0.19	0.18
PAST	0.10	0.03	0.07	0.11	0.13	0.14
< 1 CIG PW	0.05	0.01	0.03	0.05	0.07	0.08
1-6 CIG PW	0.03	0.00	0.01	0.03	0.05	0.06
> 6 CIG PW	0.07	0.00	0.01	0.04	0.11	0.18
SHARE OF FRIENDS						
ALL	0.02	0.01	0.01	0.02	0.04	0.05
MOST	0.14	0.02	0.06	0.14	0.21	0.27
HALF	0.12	0.03	0.06	0.11	0.17	0.21
A FEW	0.35	0.22	0.35	0.42	0.40	0.35
NONE	0.37	0.72	0.51	0.31	0.18	0.11
COTIDINE SAMPLE						
UNDETECTED	0.34	0.38	0.37	0.35	0.34	0.29
PASSIVE	0.55	0.60	0.59	0.58	0.51	0.49
HIGH	0.11	0.02	0.04	0.08	0.15	0.22
CIGS IN DIARY	0.16	0.03	0.07	0.13	0.23	0.31

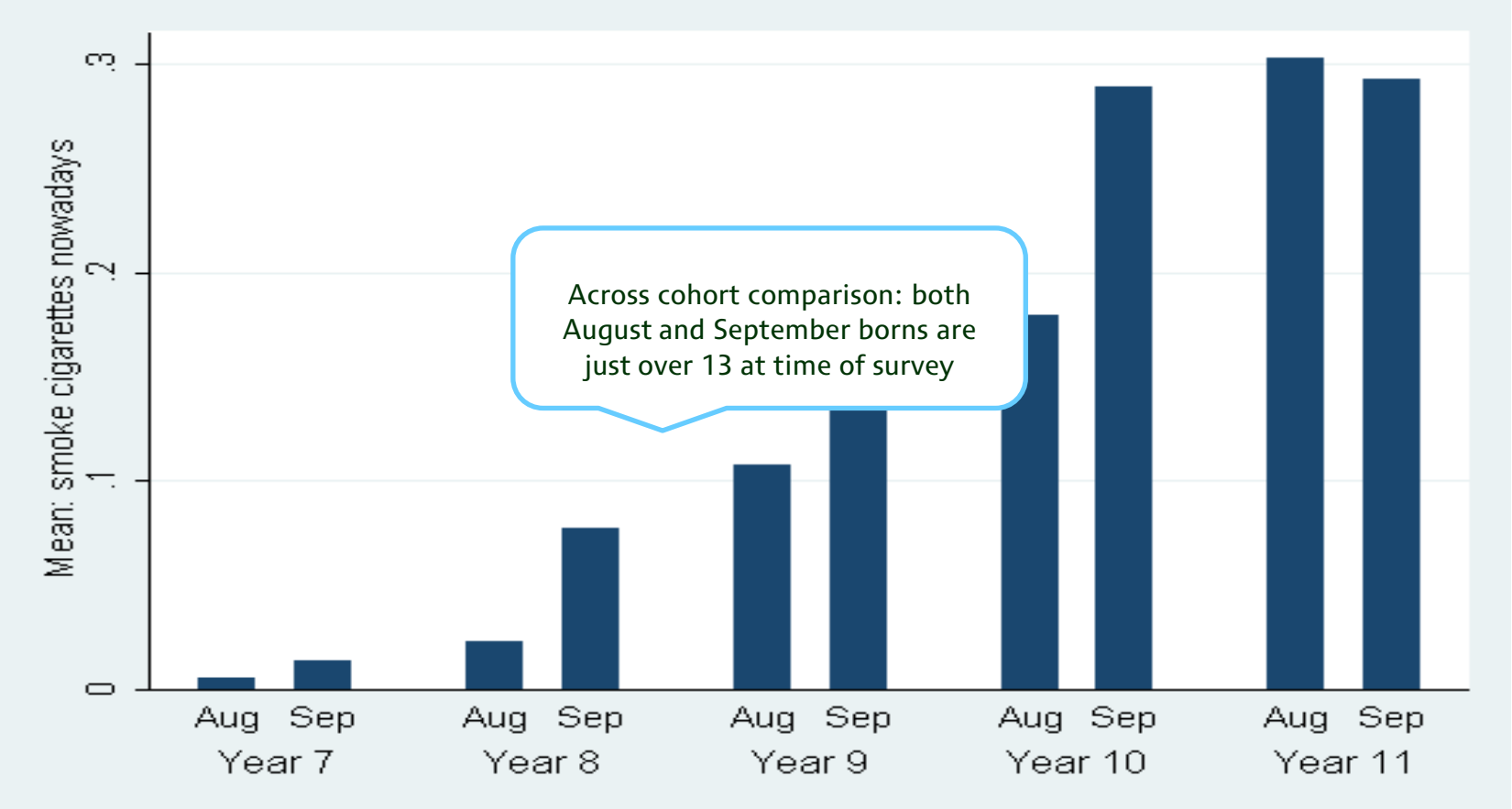
Source: Crawford, Greaves & Parey (2013)

The importance of accounting for age . . .



Source: Crawford, Greaves & Parey (2013)

The importance of accounting for age . . .



Source: Crawford, Greaves & Parey (2013)

Case Study 2: a source of exogenous variation

- Link between child adiposity (fat mass) and educational attainment by researchers from Bristol using ALSPAC data
  - von Hinke Kessler Scholder, Davey Smith, Lawlor, Propper & Windmeijer (2011), *Genetic Markers as Instrumental Variables*, CMPO Working Paper No. 11/274
- Use genetic variation in two SNPs – *FTO* and *MC4R* – as instruments for adiposity/fat mass
  - Underlying reasoning is that variation in SNPs is random (or at least unrelated to educational attainment) but also predictive of adiposity

Genetic predictors of fat mass

FTO	MC4R	Frequency	Mean fat mass
TT	TT	20.8%	98.398
TT	CT/CC	15.6%	98.998
TA	TT	27.4%	99.963
TA	CT/CC	20.4%	100.654
AA	TT	9.2%	100.990
AA	CT/CC	6.6%	102.378

Source: von Hinke Kessler Scholder, Davey Smith, Lawlor, Propper & Windmeijer (2011), *Genetic Markers as Instrumental Variables*, CMPO Working Paper No. 11/274

- Strong relationship between risk alleles and fat mass
  - Those with neither have lowest fat mass; those with both the highest

But are such predictors strong enough?

- Find a significant negative relationship between fat mass and educational attainment that becomes positive and insignificant once they use genetic markers to instrument for adiposity

- Suggests some caution may be required:
  - Are differences in genetic markers able to explain enough variation in the risk factors of interest to make results economically important?
  - Are existing sample sizes large enough to detect significant results?

References

- Argys & Rees (2006), Searching for peer group effects: a test of the contagion hypothesis, *The Review of Economics and Statistics,* Vol. 90, pp. 442–458.
- Crawford, Dearden & Meghir (2007), *When You Are Born Matters: The Imapct of Date of Birth on Child Cognitive Outcomes in England*, CEE Discussion Paper 93.
- Crawford, Dearden & Greaves (2011), *Does when you are born matter? The impact of month of birth on children's cognitive and non-cognitive skills in England*, IFS Briefing Note 122.
- Crawford, Greaves & Parey (2013), *Peer Composition and Risky Behaviour: Evidence from the School Starting Age Discontinuity*, forthcoming IFS Working Paper.
- Chen, Fortin, Oreopoulos & Phipps (2013), *Young in Class: Implications for Inattentive/Hyperactive Behaviour of Canadian Boys and Girls* (available at: www.cream-migration.org/uploads/N_Fortin_Mar_2013.pdf).
- von Hinke Kessler Scholder, Davey Smith, Lawlor, Propper & Windmeijer (2011), *Genetic Markers as Instrumental Variables*, CMPO Working Paper No. 11/274.
- Landerso, Nielson & Simonsen (2013), School Starting Age and Crime, Aarhus University Economics Working Papers 2013-3.
- Muhlenweg (2010), Young and innocent: international evidence on age effects within grades on victimization in elementary school, *Economics Letters*, Vol. 109, pp. 157–60.