THOUGHTS ON HANDLING ROUTINELY-COLLECTED ELECTRONIC HEALTH RECORDS

(or attempting to analyse observational data intended for continuity of care and performance – what could possibly go wrong!!?)

Adam J. Streeter

Research Fellow – Medical Statistics

My background:

PLYMOUTH UNIVERSITY PENINSULA CLINICAL TRIALS UNIT

- School-based trials
- ERS (Institute of Health & Communities)
- Oncology

PenCLAHRC

UNIVERSITY OF EXETER WITH PLYMOUTH UNIVERSITY

- Collaboration on RCD (CPRD)

CPRD – a brief history: (other RCD databases are available)

the 80s
- Developed by GP, Dr Alan Dean, in Essex for day-to-day management of his surgery

1987
- set-up as venture capital company, VAMP, in 1987, recruited other practices

1993
- Take-over by Reuters, but Dept. of Health - independent custodian for benefit of public health – General Practice Research Database

1994
- OPCS (now ONS) took over management on behalf of DoH

2012
- Became the Clinical Practice Research Datalink (CPRD)

Lawson, D. H., Sherman, V., & Hollowell, J. (1998). The General Practice Research Database. *QJM: An International Journal of Medicine*, *91*(6), 445–452
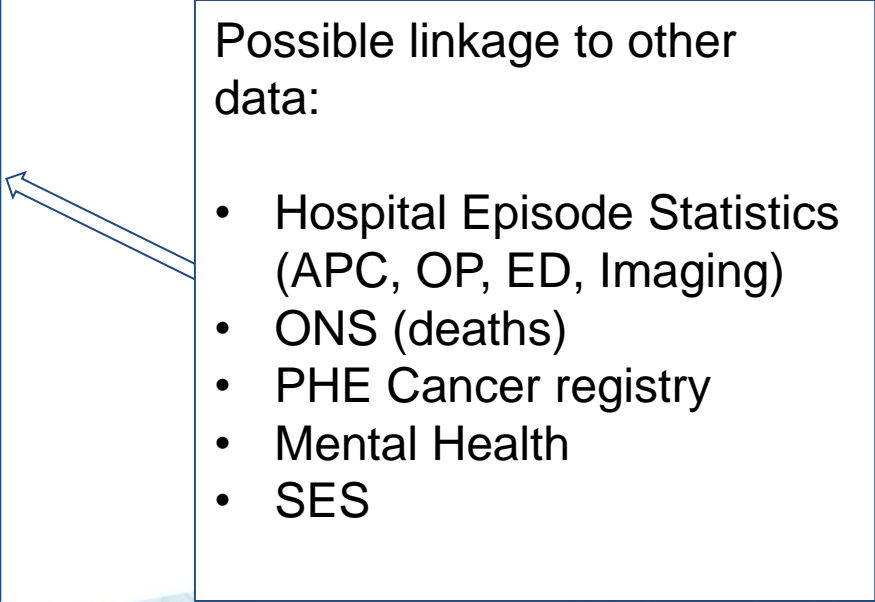
CPRD – what is it?

Files linked by *Patient ID* and/or *Consultation ID* fields:

- Patient – demographics, registration
- Consultations - date
- Clinical - diagnosis
- Immunisation
- Therapy - prescriptions
- Practice – demographic, UTS, LCD
- Referrals – secondary care
- Test results - laboratory
- Staff – description, type

Dictionaries

Possible linkage to other data:

- Hospital Episode Statistics (APC, OP, ED, Imaging)
- ONS (deaths)
- PHE Cancer registry
- Mental Health
- SES

My background:

REVIEW

Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review

Adam J. Streeter[a,b], Nan Xuan Lin[a,c], Louise Crathorne[d], Marcela Haasova[e], Christopher Hyde[f], David Melzer[g], William E. Henley[a,*]

[a]Health Statistics Group, Institute of Health Research, University of Exeter Medical School, University of Exeter, St. Luke's Campus

Confounding bias:

- Self-controlled case only
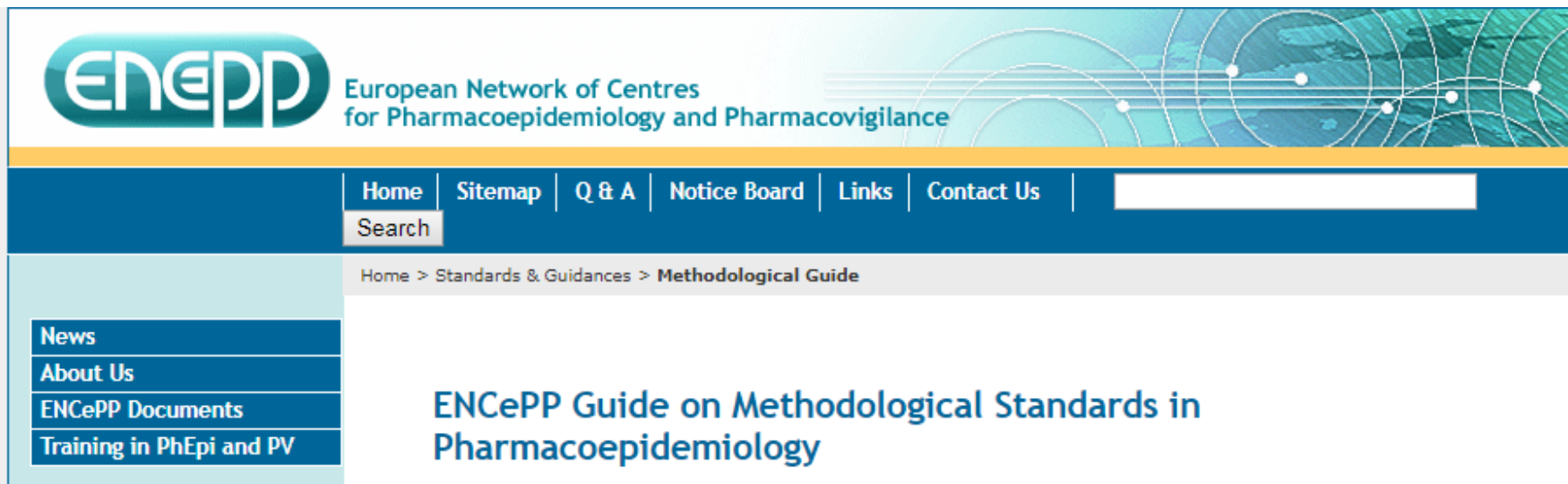- New-user active comparator
- Cohort

Study design

Confounding adjustment (measured only)

Confounding adjustment (incl. unmeasured)

- Propensity scores
- Marginal structural models
- G-methods

- Prior event rate ratio
- Instrumental variables
- Difference-in-differences
- Regression discontinuity
- Sensitivity analysis

Other resources:

ENePP — European Network of Centres for Pharmacoepidemiology and Pharmacovigilance

Home | Sitemap | Q & A | Notice Board | Links | Contact Us
Search

Home > Standards & Guidances > **Methodological Guide**

News
About Us
ENCePP Documents
Training in PhEpi and PV

ENCePP Guide on Methodological Standards in Pharmacoepidemiology

RECORD
REporting of studies Conducted using Observational Routinely-collected Data

Pharmacoepidemiology extension on the way!
RECORD - PE

Case study:

- Study into arthritis funded by major pharma

- Clinical PI applied for CPRD, downloaded data

- Unexpected results!

Case study:

- Study into arthritis funded by major pharma

- Clinical PI applied for CPRD, downloaded data
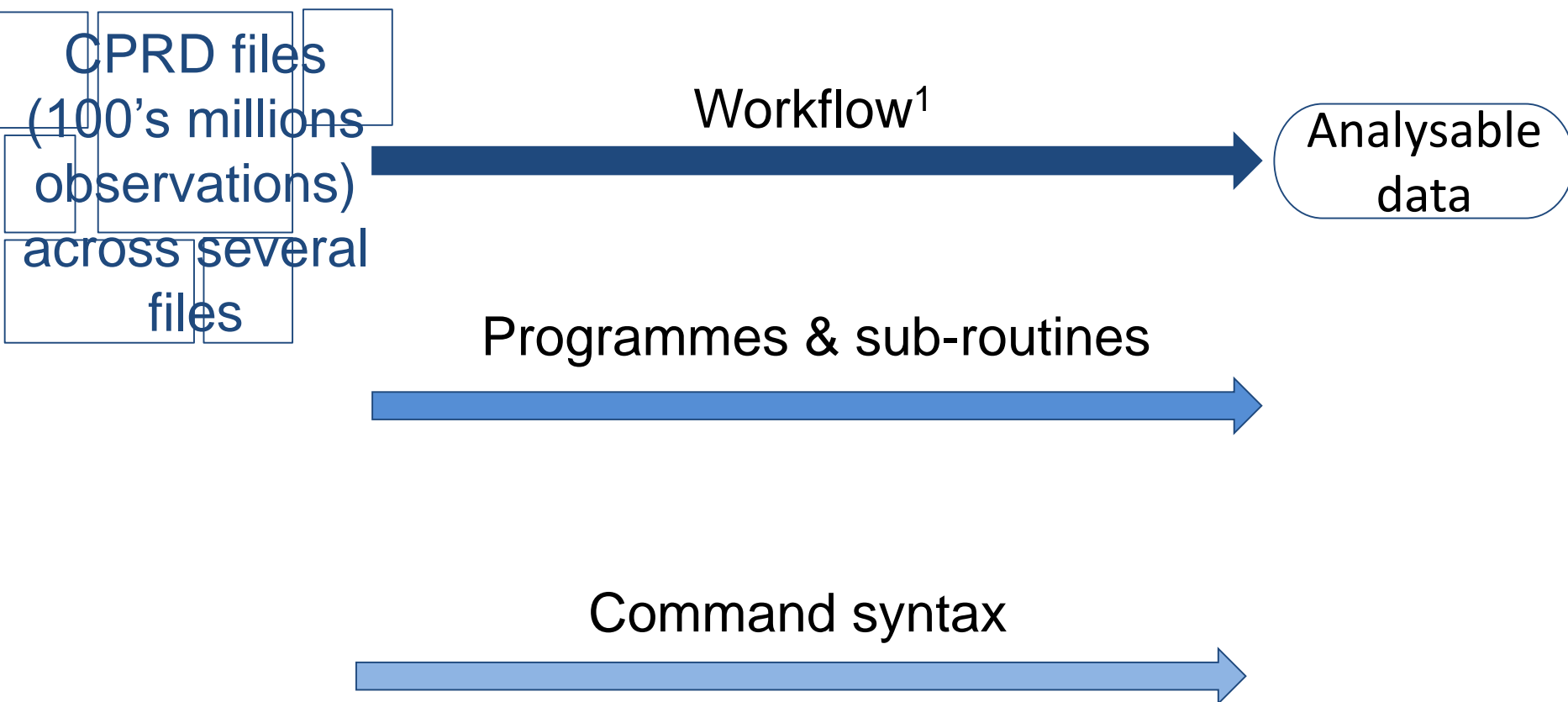
- Unexpected results!

But - after data permissions had expired - statistician found:

- Medcodes for condition were incomplete

- Not all the data downloaded (PI was worried about storage space!)

What lessons can be learnt?

➢ PI at fault? More training for PI?

➢ Closer liaising with statistician

➢ Statistician role for querying & downloading from database

➢ Availability of IT infrastructure assumed – Not reflected in ISAC protocol

Own view of CPRD

CPRD files (100's millions observations) across several files

Workflow[1] → Analysable data

Programmes & sub-routines

Command syntax

1. Long, J. S. (2009). *The workflow of data analysis using Stata*. Stata Press.

Why worry?

- ➢ We want the right answer!! (i.e. no errors)

- ➢ More regulation in CTUs than RCD studies

- ➢ More programming with RCD than trial data

- ➢ Good programming practice currently v. big issue in UKCRC, scrutiny by MHRA

- ➢ Replicability, audit trails

Replicability

- Directory paths & specific folders

- Document changes

- Version control (use version ctrl software?)

- Data signature – checksum

Programming resources

- Write programs for people, not computers.

- Let the computer do the work.

- Make incremental changes.

- Don't repeat yourself (or others).

- Plan for mistakes.

- Optimize software only after it works correctly.

- Document design and purpose, not mechanics.

- Collaborate.

Best practices for scientific computing. *PLoS Biology* (2014), *12*(1), e1001745.

Programming resources

❑ Statistican involved in data download?

❑ Unmeasured confounding & study design

❑ Guidelines?

❑ Programming good practice!

❑ Share code

Thank you

Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *Journal of Clinical Epidemiology* (2017), *87*, 23–34. Streeter, A. J., Lin, N. X., Crathorne, L., Haasova, M., Hyde, C., Melzer, D., & Henley, W. E.