Database use in Pharmacoepidemiology

Gillian Hall – October 2014

Disclosures

- No funding was received for this presentation
- I have worked extensively with a number of data resources
- I am on the advisory committee of one data resource

ISPE Database Guidance

ISPE Database SIG:

‘Guidelines for good database selection and use in pharmacoepidemiology research’

Pharmacoepidemiology and Drug Safety 2012; 21: 1–10

ISPE Database Guidance

Aims to assist in:

- selection and evaluation of a resource
- use of a data resource
- review of database studies
- provide a check list of factors to consider

Motivation for ISPE guidance

Variation in resources:

- healthcare system
- reason for data collection
  - Clinical – electronic medical record
  - Financial – claims / payment system

Motivation for ISPE guidance

- analysis not always by specialist teams
- linkage between resources
- different concerns about confidentiality
- number and variety of resources

Guidance for Industry and FDA Staff

Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data

U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)
May 2013

http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM243537.pdf

‘Investigators should demonstrate a complete understanding of the electronic healthcare data source and its appropriateness to address specific hypotheses’

FDA, May 2013

Guidance sections

1. Selection of a database
2. Use of multiple data resources
3. Extraction & analysis of the study population
4. Privacy and security
5. Quality and validation procedures
6. Documentation

1. Database selection

- Is the appropriate population covered?
  - Size
  - Coverage
  - Representativeness

1. Database selection

Safety of a new diabetes treatment:

- UK and Netherlands primary care databases included all diabetic patients
- Italy only some treatment given in primary care

1. Database selection

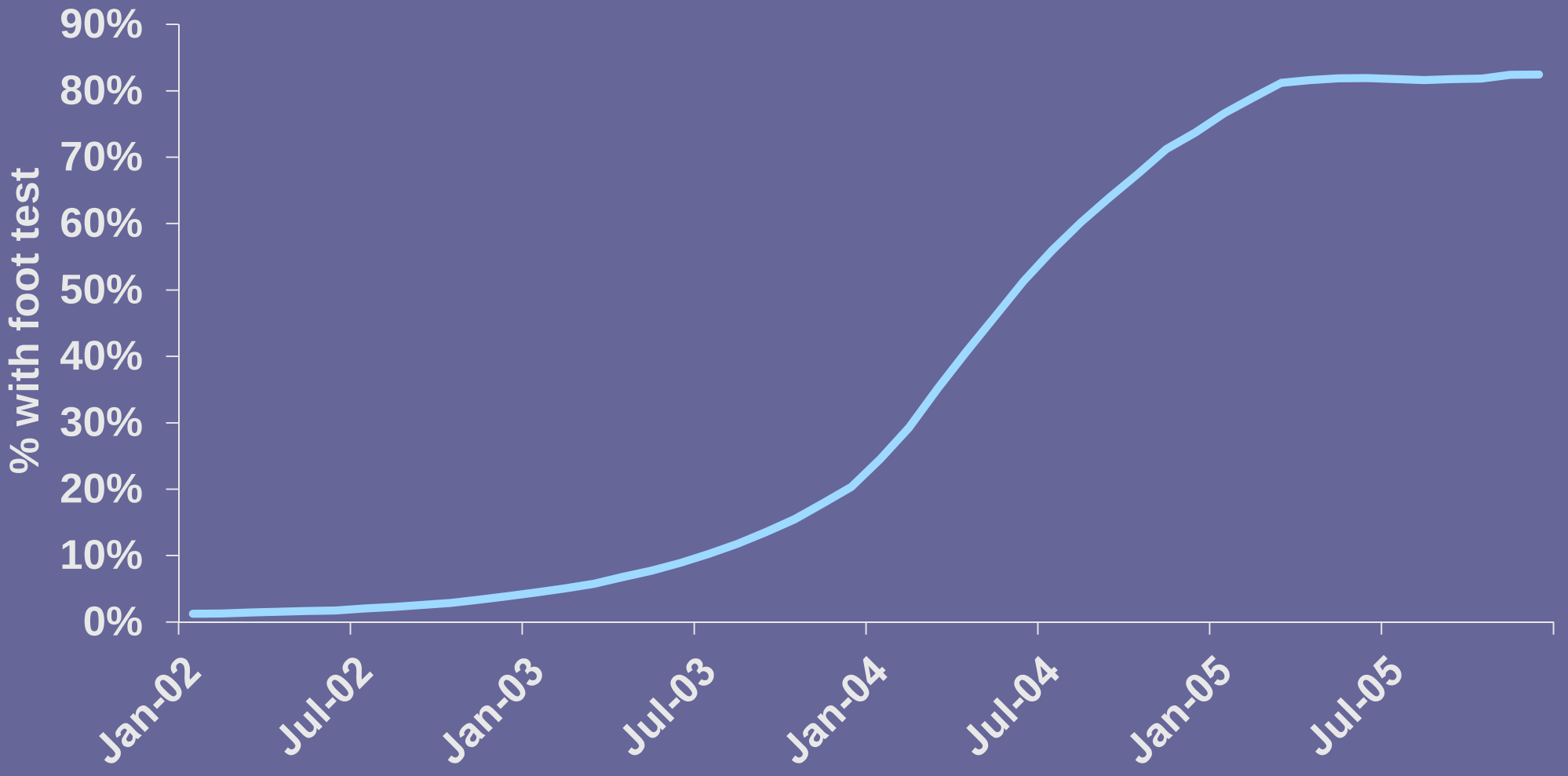
- Are study variables captured?
  - Exposure
  - Outcomes
  - Inclusion criteria
  - Exclusion criteria
  - Potential confounders

1. Database selection

- Are study variables captured?
  - On every patient
  - Consistently
  - Without bias
  - In appropriate detail
  - Accessible for research (including coding)

% diabetic patients with peripheral neuropathy testing in previous15 months

(THIN data Blak et al ICPE 2010)



‘Getting closer to a cure for what ails health care’

‘Government Plans Major Health Care Reform’

1. Database selection

- Time to release of data
  - Canadian provincial data >12 months
  - UK hospital episodes statistics 18 months

1. Database selection

- Expertise on the database / healthcare system is key

1. Database selection

- Can data from one health care system provide evidence of safety in another.

1. Database selection

- No database can answer every question
- Some questions can’t be answered by database studies…

  …but these analyses can sometimes contribute to an understanding of the issue.

1. Database selection

Atypical antipsychotics - cardiac arrhythmias

- Sertindole versus risperidone
- Sertindole:
  - no longer available
  - the possible association with arrhythmias well reported = potential confounding

Thomas et al (2010). Acta Psychiatrica Scandinavica

1. Database selection

Long-term follow-up of patients exposed to a treatment used in end-stage renal disease.

- Uncommon condition
- Patients often unable to work
- Long-term follow-up

= specialist renal database (USRDS)

2. Multiple Resources

- To increase:
  - numbers
  - representativeness
  - breadth of patient information

2. Multiple Resources

- To increase breadth of patient information

e.g. UK primary care databases

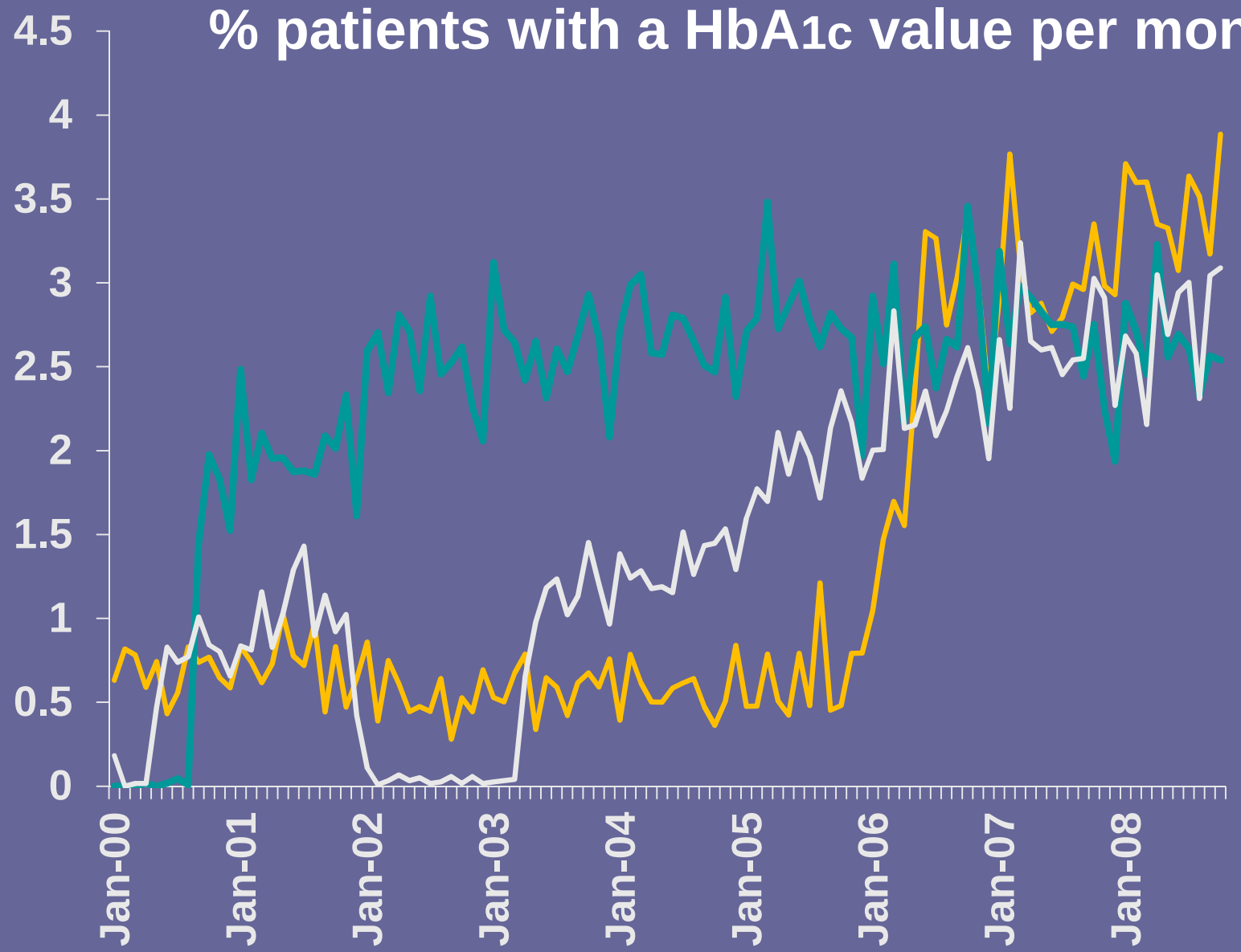
  - Laboratory data
  - Inpatient episodes
  - Outpatient episodes
  - Mortality data
  - Cancer registers

2. Multiple Resources

UK primary care study of cause of death

- Cause recorded in 63.4% of deaths
- Studies of acute potentially fatal conditions will miss cases
- Death certificates returned (92.8%) but slow and expensive.

Hall GC. (2009) *Pharmacoepidemiol Drug Saf*


% patients with a HbA1c value per month
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05
Jan-06
Jan-07
Jan-08

2. Multiple Resources

Issues to consider:

- Is reliable person-matching possible?
- Are data compatible in metrics, policy and terminology?
- Can duplicates be identified?
- Should a central or distributed system be used?

3. Extraction and analysis

- First step often to extract a specific study population
- All study variables extracted for this population
- Final analysis

3. Extraction and analysis

- A detailed specification is key

  = documentation of every step of the extraction and analysis based on the final protocol

3. Extraction and analysis

A detailed specification:

- Default dates:
  - DOB – default to 01/07/yyyy
  - other data:
  - if day missing, default to 15/mm/yyyy
  - if month and day missing…

3. Extraction and analysis

Main aim of a detailed specification

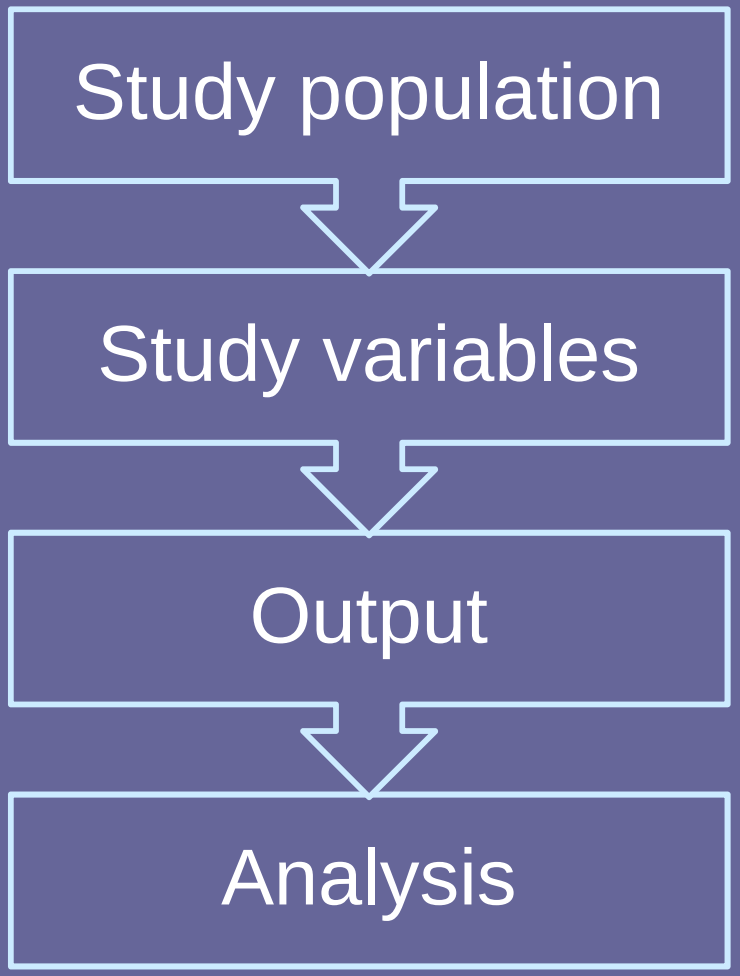
- A documented agreed detailed strategy for all stakeholders: epidemiologists, statisticians, programmer and tester

3. Extraction and analysis

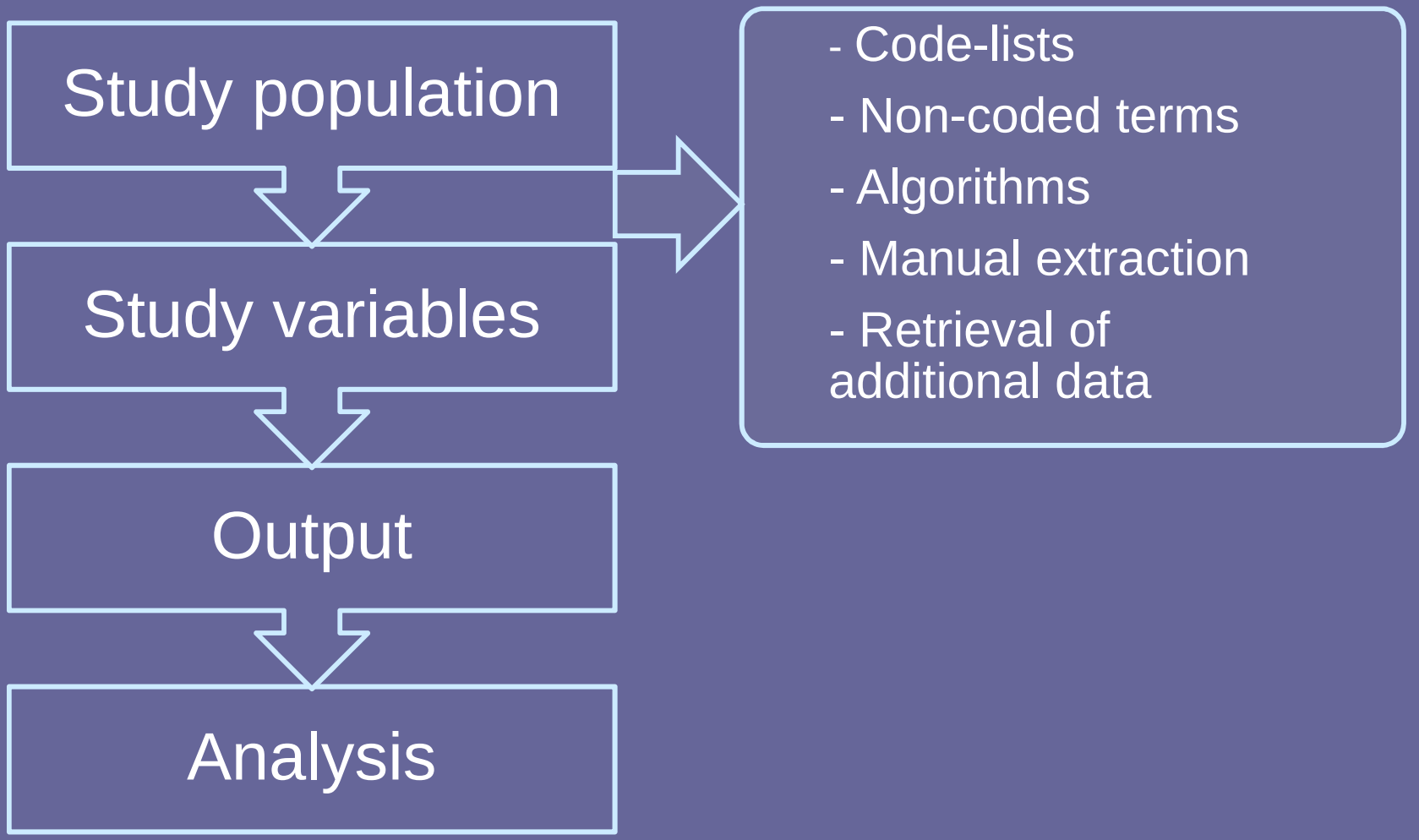
A detailed specification

- Ensures that principle researcher, programmer and tester all working to the same plan
- That multiple resources are all working to the same plan

3. Extraction and analysis



3. Extraction and analysis



3. Extraction and analysis

Code lists:

- Straight forward - common clearly defined diagnoses
- More challenging – rare conditions
  - ill-defined disease
  - difficult diagnosis
- Clinical expertise can be valuable
- May need to use a combination of codes or update lists during the study

3. Extraction and analysis

Code lists:

Should be study specific

E.g. in diabetes would you include:

- Type 1 and 2?
- Gestational diabetes?
- Management codes?

3. Extraction and analysis

- Often an heuristic process

4. Privacy and security

- Compliance with privacy and security policy
  - Regular privacy reviews
- Limited use of identifying information
- Masking patient identifiers
- Managing data in a secure environment

4. Privacy and security

- Situations that might increase the potential for revealing patient identification:
  - a new database
  - collection of additional data
  - multiple resources
  - narrative data

4. Privacy and security

- Whose responsibility is it to ensure privacy?

5. Quality and validation

- Two aspects:
  - overall database
  - study specific elements

5. Quality and validation

In both cases:

- Accuracy and completeness at the source
- Data integrity after extraction

5. Quality and validation

Overall database:

- Documentation on established databases
  - Literature + supplier
- What about a new resource?

5. Quality and validation

Validation of PainDB - electronic versus written notes

- 88% entries matched
- 55% patients, all visits recorded
- 9% entries on Pain DB, not in written notes

Hall et al (2008) *Anaesthesia*

5. Quality and validation

Study population checks:

- Study variables
- Assumptions
- Algorithms
- Code lists

5. Quality and validation

Study variables:

- Which should be validated?
  - Primary exposure and outcome
  - Other variables
- Is there a mechanism?
- Is published validation relevant?
- What accuracy and completeness is acceptable?

5. Quality and validation

Missing data:

- Missing data methods used

Assumption and algorithms

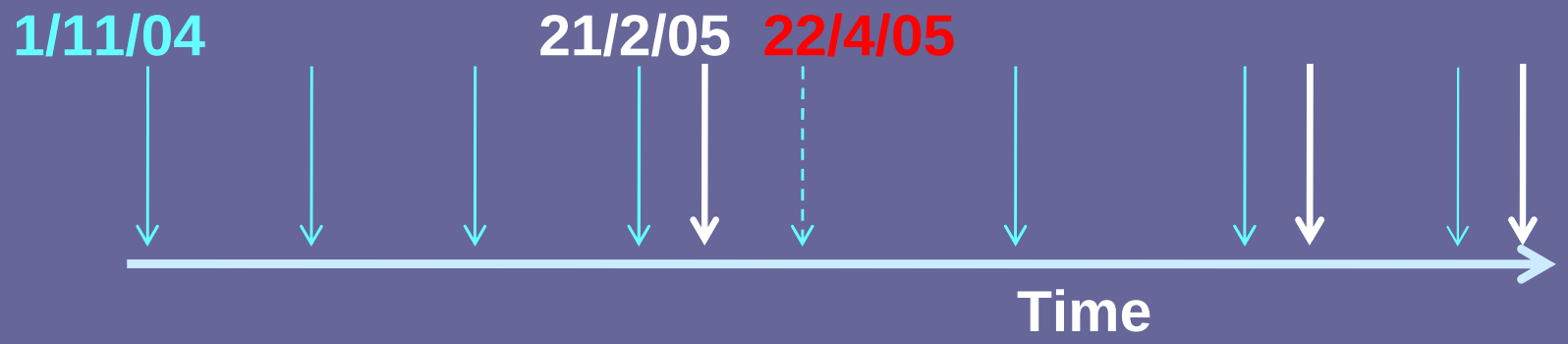
Assumption in diabetes therapy

- Dual therapy = Rx of 2 drugs on same day
- Single exposure = Rx 1 drug on same day

5. Quality and validation

Metformin: started 1/11/04

Gliclazide: added 21/2/05



5. Quality and validation

Vaccine study:

- designing a study of one vaccine brand
- previously studied 1 vaccine class - assumed vaccine record was correct
- preliminary analysis of brand – batch number recording
- Brand identified for 94% of influenza vaccines

5. Quality and validation

Checks on the process:

- code lists
- extraction
- data merging
- programming code

5. Quality and validation

Testing techniques

- External, logical, internal
- Cross-sectional, longitudinal and up to date
- Dummy datasets

5. Quality and validation

Sustainability

- Code lists
- Healthcare policy
- Reimbursement policy
- Software changes

6. Documentation

- Documentation of every stage

PASS of childhood vaccine (age 13) – Bell's palsy + anaphylaxis

Database requirements:

- a general paediatric population
- large sample size as rare outcomes
- data likely to be collected over coming years
- data available promptly

Childhood vaccine at age 13 years – Bell's palsy + anaphylaxis

Exposure

- school vaccination details
- need the date of vaccination not of report
- 1 brand studied; need brand name or batch

Childhood vaccine – Bell's palsy + anaphylaxis

Outcome

- primary care diagnosis (Bell's palsy)
- secondary care diagnosis (anaphylaxis)
- can the outcome be identified?
  - a specific code
  - cause of death

Childhood vaccine – Bell's palsy + anaphylaxis

Checks

- Assumptions
  - School vaccination included on database
  - Date of vaccination recorded
  - Batch number can be linked to vaccine
- Have the outcomes been validated?

Other guidance

- A number of complementary guides
  - Good pharmacoepidemiology practices
  - STROBE statement
  - ISPOR guidance
  - ENCePP
  - FDA