USE OF PRIMARY CARE DATABASES IN EPIDEMIOLOGIC RESEARCH

Luis Alberto García Rodríguez

Spanish Centre for Pharmacoepidemiologic Research (CEIFE), Madrid, Spain

Index

The structure to successfully perform a pharmacoepidemiology study with automated primary care databases can be divided into the following items:

- **Hypothesis:** correct research question
- **Study design:** best suited design
- **Rich data source:** numbers and detail (quantity and quality)
- **Correct definition and classification** of the **outcome/exposure**
- **Analysis:** adequate analysis plan

Study Hypothesis

- Statement about the relationship between a drug and a disease that can be tested

- Hypothesis definition requires the researcher to have knowledge on:

  - **Biological mechanism:** detailed knowledge about the study drug and disease; mechanism of action (i.e. whether the effect is acute or chronic, local or systemic, modified by other factors…)

  - **Data sources**: detailed knowledge about the source of information (i.e. population included, type of health system, strengths and limitations…)

Study Hypothesis

The study design and methodological approaches will be subject to the study question

- **When there are prior data available:**
  - this must be considered when conducting a new study
  - especially when a new study aims to replicate what was already observed in a previous one
- **When no prior data are available:**
  - defining the hypothesis might be more challenging
  - the marginal contribution of a single study is the greatest

Study design

How to decide the most appropriate design under the observational approach:

- **Cohort:** e.g. follow-up study, long-term outcomes, survival
- **Case-control:** e.g. specific dose/duration effect
- **Case-crossover:** e.g. transient exposures/acute/prompt outcomes
- **Drug utilization:** e.g. treatment patterns, switching patterns, comparison with guidelines recommendations
- **Meta-analysis:** e.g. pooled estimates to reach conclusions on hypothesized association performed by different studies

Study design

Traditional “differences” between cohort and nested case-control studies

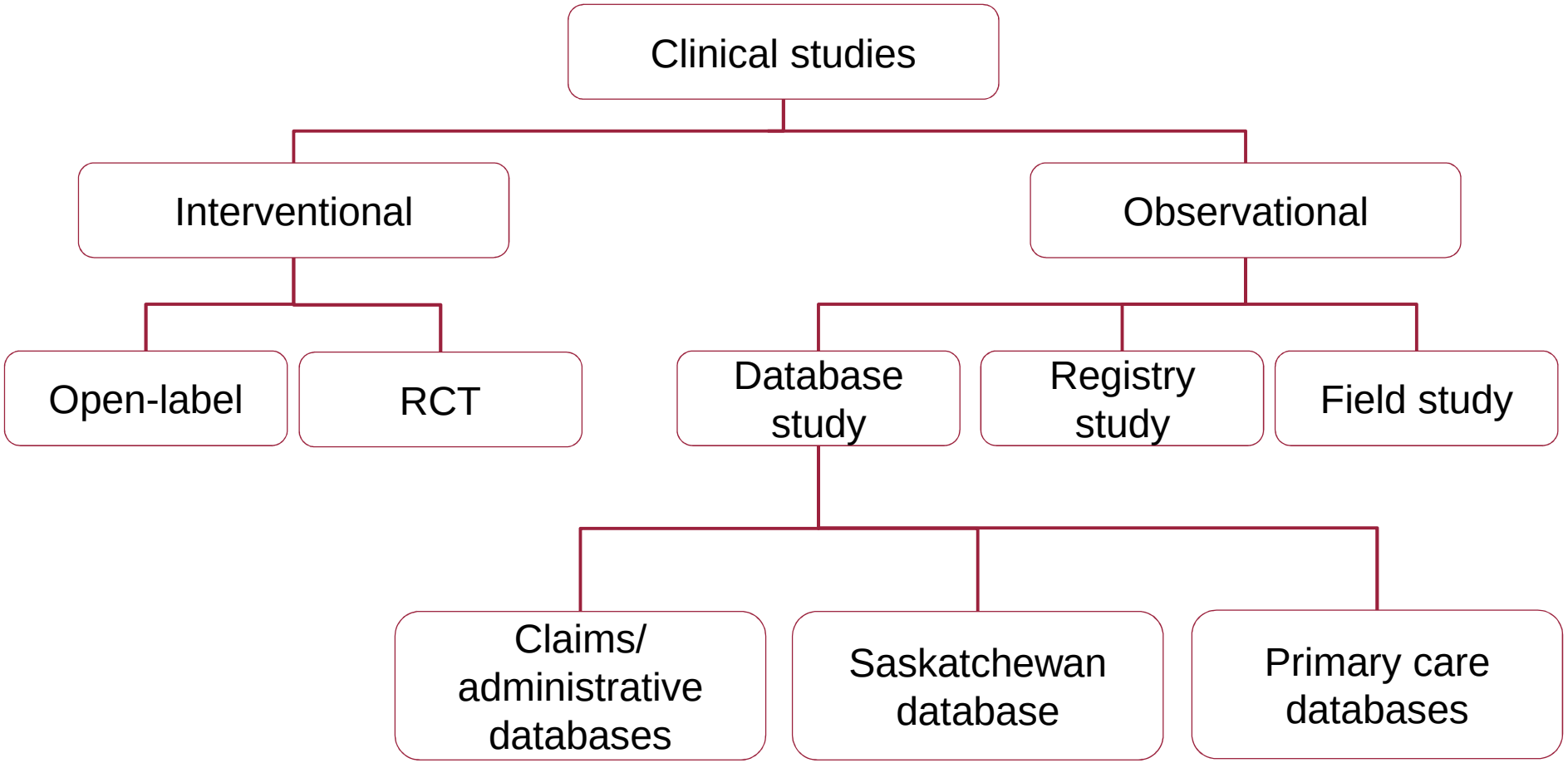
COHORT STUDIES

- Temporal sequence between exposure and outcome can be established
- Useful to study rare exposures, i.e. a specific chemical product
- Multiple outcomes associated with the exposure can be studied
- Information on confounding factors can be obtained
- Useful for estimating the risk of disease, the incidence rate and/or relative risks. Time-to-event analysis is possible as well
- Less prone to selection and information biases as compared to other designs

NESTED CASE-CONTROL STUDIES

- They may be less expensive and time consuming than cohort studies
- Rare diseases may be explored
- Diseases with long latency period can be studied
- It is possible to investigate multiple exposures
- When a risk set sampling is used to select the controls, the outcome estimate is simila to the risk ratio
- In "nested" case-control designs, informati on exposures have been collected before cases had been diagnosed, and may be le prone to bias.

Clinical studies: where do pharmacoepidemiology (PE) studies fit in?



Sources of information in pharmacoepidemiology

- **Field-based studies**: Researcher captures the information directly from the patient
  - Interview
  - Survey
  - Nurses

- **Registries**: Systematic collection of data about specific conditions. There can be population-based as well as hospitalized-based
  - Cancer
  - Pregnancy
  - Autoimmune diseases (Multiple Sclerosis)

- **Automated/computerized database:** Digital version of a paper chart that contains collections of clinical records with a defined structure and purpose. These databases include information at patient-level, on demographics, drug prescriptions, specialist referrals, hospital admissions, hospital discharge diagnoses, operations, ambulatory care

Automated databases in pharmacoepidemiology

- **Claims/administrative databases**: Captured claims data for individuals or multiple insurers, describing all transactions that results in a claim for reimbursement.
  - **Medicaid:** Federal state program for financing medical care to low income population.
  - **HMO database**: Consortium of health care delivery organizations with integrated research divisions (Kaiser Permanente, Harvard Pilgrim), group health cooperative. Collected for administrative purposes. Drug inventory, hospital discharges and demographic accounting
  - **Commercial insurance databases**: HealthCore, PharMetrics

- **Saskatchewan databases:** Provincial health system (government held database) , universal coverage for most residents, information accumulated on computer. Possibility of linking to other files (death records, cancer registry, mental health services)

- **Primary care databases**: Electronic medical records

Source of information

Revolution of primary care database started in the ‘80s

Some characteristics converged together to make the first EMR in the UK a possibility

- Universal Health Care system in the UK
- Replacement of medical charts with medical records (Alan Dean, VAMP, eventually EMIS, Vision, etc)
- Dictaphones to record diagnostic and treatment information
- Active counseling and training to primary care physicians willing to participate
- Validation exercises for recorded information to ensure quality of information (Boston Collaborative Drug Surveillance Program (BCDSP)

Source of information

Pioneers sources of information in pharmacoepidemiology…

- **CPRD and THIN (two primary care databases in the UK) are gold-standard sources of information to perform pharmaco-epidemiologic studies**
  - Strengths of these resources include:

    i) Highly granulated dictionaries (Read, drug Gemscript)

    ii) Drugs and devices prescriptions automatically recorded

    iii) Completeness of information: primary care practitioners (PCPs) are gatekeepers in the health care system

    iv) Feasibility of validation processes

Use of **large databases** with access to millions of patients enables the study of **rare diseases** or **outcomes difficult to capture**

The Health Improvement Network

- **Medical research database of systematically recorded anonymous patient records**
- **Validated for use in pharmacoepidemiology[1]**
- **Contains details on close to 4 million patients currently registered with primary care physicians (PCPs) in the UK**
- **Prospectively recorded information on patients' demographics, medical history (symptoms, diagnoses), prescriptions, additional health data (laboratory results), details of outpatient visits and hospitalisations**
- **Free text comments are available**
- **Subset of practices linked to HES files**

[1]Lewis JD *et al. Pharmacoepidemiol Drug Saf* 2007;16:393–401.

Source of information

Choosing a good data source alone does not guarantee the validity of the study results

- Transforming data into relevant information regarding the study hypothesis will depend on researcher's ability to**:**
  - **Extract**
  - **Process**
  - **View**
  - **Analyze**

➡ Importance of using specific tools to support efficient data collection and analysis

Outcome and exposure validation

How to start?

1) Outcome of interest

- Relevant operational definition: constructing specific diagnostic algorithms
- Validation of diagnostic algorithms

This validation process should include:

I) Manual review of patients' profiles with anonymized free text comments

II) Questionnaires sent to Primary Care Physicians requesting them to send anonymized copies of notes (hospital summaries, discharge/referral letters, reports of diagnostic procedures, etc.)

Outcome and exposure validation

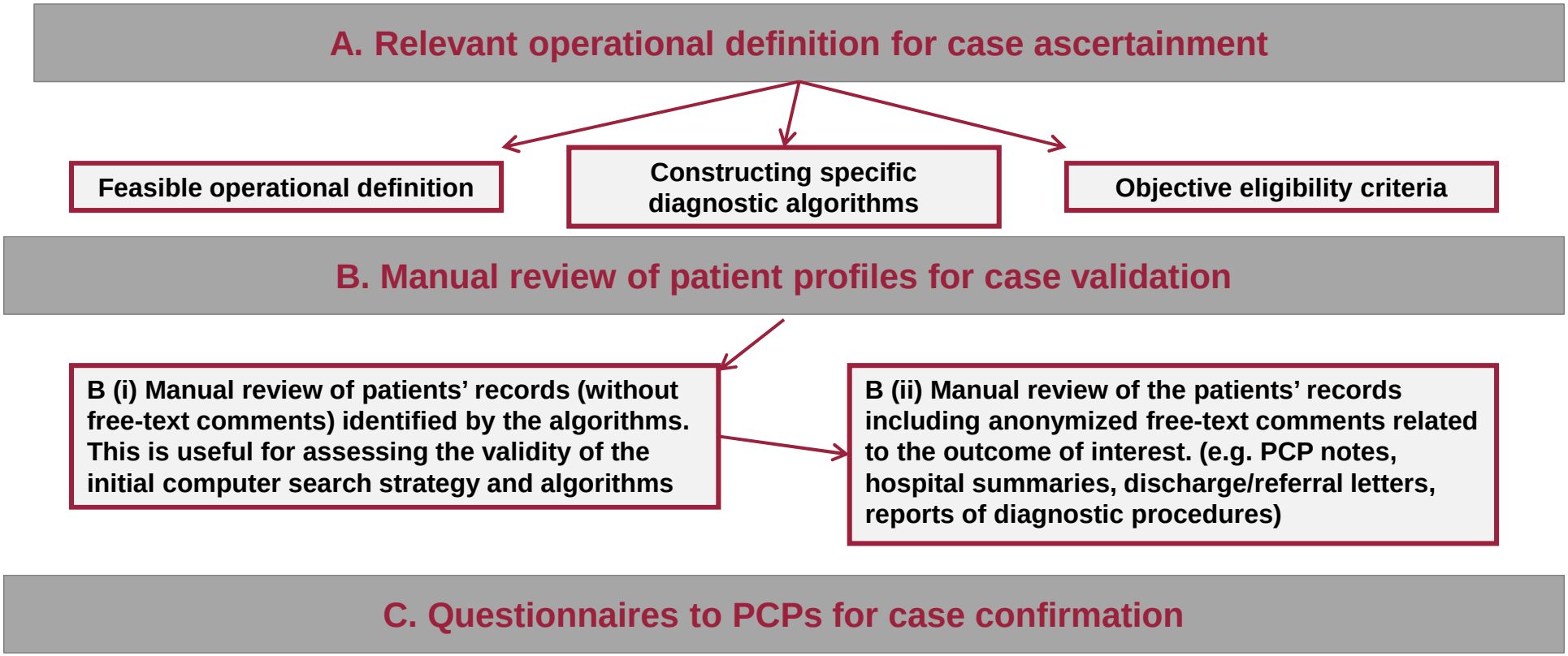
- The full validation process, including both steps, should always be performed when:
  - There is **no prior experience** with the particular disease in the database
  - **Concerns exist regarding our ability to validly identify** cases using only coded information

- When validation of the entire case set might not be feasible for logistic reasons:
  - **Random subset(s)** of all computer-detected cases **is required to undergo validation**

Validation in primary care databases

The task of transforming these data into relevant information addressing the study hypothesis depends on researchers' expertise in extracting, processing and analysing the data.

The process of case ascertainment and validation of the outcome of interest is summarised below



PCP, primary care practitioner

Outcome and exposure validation

2) Drug exposure

- Classification of individuals or person-time according to recency, dose, and duration of exposure

Extracting information from the database:

- Not always a straightforward process
- Often requires fairly **complex programming algorithms**, considering:
  - **time period elapsed** between repeat prescriptions, prescribed **daily dose** and **number of pills** in each prescription
  - should be able to **deal with partial missing data** and to assume some degree of non-compliance

Outcome and exposure validation

- Classification of individuals or person-time according to recency, dose, and duration of exposure:
  - might **vary between studies** depending on the strategy followed
  - these **differences of drug exposure definition** could have an impact on the results
- Invalid exposure ascertainment might prevent documenting dose or duration response when inferring causality
- Errors in exposure measurement will compromise interpretation of the results

Outcome and exposure validation

Our ability to correctly ascertain drug exposure might be limited by

quality of the information in the data source

e.g.

- drug exposures obtained from biennial self-reported questionnaires (e.g. Health Professionals or Nurses' Health Study), or
- drug exposure ascertained at initial discharge from hospital and assumed to remain constant over a long follow-up period

Primary care databases based on prospective recording avoid the likelihood of recall biases present in sources of information with either retrospective recording or recording based on memory of participants

Outcome validation: True Positive

```
#    7 Id:0007-004426  FamId:0002067 Pract. Start: 1/07/1992 Vision Start:28/12/1995 Reg:24/05/1990 RSt: DIED      Out:19/07/2001 DB:THIN1201
  MALE DOB: 1/07/1929  Age at Ix: 70 Marit.Stat: ?        Drawdown:01/12 Urb/Rural:5 Townsend Idx:2 Rxs:156  Evs:250  Fds:29   T.R.:37

47526  15/02/2000   -----      Discharged from hospital                    CVA RT HEMIPARESIS.      [O  ] <8HE..00>
                               CT SCAN SHOWS INFARCT LEFT PARIETAL LOBE WITH SMALL BLEED. OLD INFARCT RT FRONTAL
                               LOBE. CREAT 164 UREA 9.8 CHOLESTEROL 5.9 WARD [DELETED]
47529  18/02/2000   -----      Monitoring of patient                       NEVER GAVE HIM           [OF ] <8A...00>
                               ASPIRIN. DRAGGING TOES ARM CANT BE LIFTED. 4 HOME ASSESSMENT PHYSIO.

47529  18/02/2000 <98776997>     4908  ASPIRIN                                        TAKE 1 OR 2, DAILY D=0   N=56

47530  19/02/2000   ----------------------------------------------------------------------------------------------------
47534  23/02/2000   -----      Letter encounter                            4 OPD SPEECH THERAPY     [OF ] <9N33.11>

47557  17/03/2000 Blood pressure          DIASTOLIC      75        SYSTOLIC       125        <246..00> O/E - blood pressure reading

47557  17/03/2000   -----      Monitoring of patient                       OK WALKS WEAK R SIDE.    [OF ] <8A...00>
                               EXPRESSIVE DYSPHASIA SR BP OK.
47561  21/03/2000   -----      Monitoring of patient                       PAIN R KNEE ? OA         [OF ] <8A...00>
                               CO-CODAMOL 100
47574   3/04/2000   -----      Telephone encounter                                                  [OF ] <9N31.00>

47581  10/04/2000 <96950997>     6138  AMIODARONE                                     1 EVERY DAY D=0   N=60
47581  10/04/2000 <98776997>     4908  ASPIRIN                                 TAKE 1 OR 2, DAILY D=0   N=56
47581  10/04/2000 <93212998>     6794  FRUSEMIDE W AMILORIDE                TAKE ONE EACH MORNING D=0   N=56

47590  19/04/2000 Smoking                 Smoking status - Ex smoker                          <137..00> Tobacco consumption
47590  19/04/2000 Weight                  WEIGHT(KG)     85        BMI            27.7       <22A..00> O/E - weight

47590  19/04/2000   -----      Had a chat to patient                       NOT SLEEPING WHEN        [OF ] <8CB..00>
                               STOPS ZOPICLONE?ALSO BIT LOW,CONSTIPATION A PROB SINCE CVA AND NOT EATING MUCH,NO
                               DYSPHAGIA,CHAT++,TRY VARIOUS,CHECK 1M
47610   9/05/2000   -----      Ultrasound scan                             RENAL TRACT              [OF ] <58D..00>
                               PINDERS-,NORMAL KIDNEYS APPART FROM SMALL SCAR R,150ML RESIDUAL URINE,GALL STONE
                               NOTED

47618  17/05/2000 Weight                  WEIGHT(KG)     82        BMI            26.7       <22A..00> O/E - weight

47618  17/05/2000   -----      Had a chat to patient                       MORE MOBILE,ATTENDS      [OF ] <8CB..00>
                               HYDROTHERPY,SLEEP AND MOOD BETER,SOME WT LOSS BUT EATING BETER NOW,SKIN BIT
                               RED?AMIODARONE,KEEP ON DTP 9M,CHECK WT 2M

47620  19/05/2000   - -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- -
47620  19/05/2000   -----      Seen in hospital casualty                   CVA 18. [DELETED] A/E    [OF ] <9N19.00>
47620  19/05/2000    4360      Stroke and cerebrovascular accident unsp    CT-CEREBRAL              [DF ] <G66..00>
                               INFARC-WARFARINIZED,MED [DELETED]
47625  24/05/2000   -----      Telephone encounter                         [DELETED]-UPDATED RE     [OF ] <9N31.00>
                               U/S RENAL TRACT AT [DELETED] AS [DELETED] IN HOSPITAL

47641   9/06/2000 <93212998>     6794  FRUSEMIDE W AMILORIDE                TAKE ONE EACH MORNING D=0   N=56

47644  12/06/2000 International normalised     2.7       ratio                                <42QE.00> International normalised ratio

47644  12/06/2000   -----      International normalised ratio              2.7 CONTACTED LABS       [OF ] <42QE.00>
                               PREV RESULTS 3-2.9 CONT ON 1MG C1W
47644  12/06/2000   -----      Telephone encounter                         [DELETED],FELL OVER      [OF ] <9N31.00>
                               TODAY,OK NOW,?WAS SYNCOPAL,CHAT RE VBI,REVIEW IN NEXT FEW DAYS

47648  16/06/2000 <97119997>     5703  DIGOXIN                                        1 EVERY DAY D=0   N=60
47648  16/06/2000 <95617998>     1904  WARFARIN                          AS DIRECTED BY THE HOSPITAL D=0   N=100
```

Low dose ASA initiation
CVD indication

Follow up contribution

New Episode

Outcome validation: False Positive

```
#  117 Id:0181-009456  FamId:0006125 Pract. Start: 1/07/1992 Vision Start:26/06/1997 Reg:28/02/1984 RSt: DIED      Out: 7/10/2008
FEMALE DOB: 1/07/1936  Age at Ix: 70 Marit.Stat:MARRIED   Drawdown:01/12 Urb/Rural:4 Townsend Idx:1 Rxs:295  Evs:392  Fds:73   T.I

50088  20/02/2007   -----      Telephone encounter                          FROM HUSBAND, GETTING    [OF ] <9N31.00>
                               ON WELL , SPEECH IS COMING BACK, WANTS TO COME HOME ? FIT ENOUGH, HE WILL SEE
                               SPECIALIST ON THURSDAY

50091  23/02/2007 CAT scan                                                                          <567..13> Computerised tomograph
50091  23/02/2007 Echocardiogram                                                                    <5853.11> Echocardiogram

50091  23/02/2007   -----      Letter from specialist                       CLINICAL LETTER OTHER    [O  ] <9N36.00>
                               HOSPITAL GASTROENTEROLOGY
50091  23/02/2007   -----      Letter from specialist                       CLINICAL LETTER OTHER    [O  ] <9N36.00>
                               HOSPITAL GASTROENTEROLOGY
50092  24/02/2007   -----      Discharged from hospital                     DISCHARGE                [O  ] <8HE..00>
                               LETTER/SUMMARY OTHER HOSPITAL
50092  24/02/2007   -----      Discharged from hospital                     DISCHARGE                [O  ] <8HE..00>
                               LETTER/SUMMARY OTHER HOSPITAL

50097   1/03/2007 Urine test                                                                        <46...11> Urine tests
50097   1/03/2007 Urine test                                                                        <461..00> Urine exam. - general
50097   1/03/2007 Urinalysis - Protein                                                              <467..00> Urine protein test
50097   1/03/2007 Urine Biochemistry                                                                <46N..00> Urine protein
50097   1/03/2007 Other Bacteriology Tests                                                          <4J15.11> Culture sensitivity
50097   1/03/2007 Urine Dipstick for Nitrit                                                         <46X2.00> Urine dipstick for niti
50097   1/03/2007 Blood pressure           DIASTOLIC      70        SYSTOLIC       130             <246..00> O/E - blood pressure re

50097   1/03/2007   -----      Urinalysis - general                         POSITIVE , PROTEIN       [OF ] <461.
                               LEIUKS ETC , TO LAB
50097   1/03/2007    4360      Stroke and cerebrovascular accident unsp     REFER STROKE CLINIC ,    [OF ] <G66.  0>
                               AND DAY HOSPITAL

50097   1/03/2007 <93619997>     19103  SIMVASTIN                                1 EVERY NIGHT D=0   N=56
50097   1/03/2007 <97217998>      6716  BENDROFLUMETHIAZIDE                      1 EVERY MORNING D=0 N=56
50097   1/03/2007 <96950997>      6138  AMIODARONE                               1 EVERY MORNING D=0 N=56
50097   1/03/2007 <98776997>      4908  ASPIRIN                                  1 EVERY MORNING D=0 N=56
50097   1/03/2007 <89659997>      6188  BISOPROLOL FUMARATE                      1 EVERY MORNING D=0 N=56

50098   2/03/2007   ----------------------------------------------------------------------------------------------------
50102   6/03/2007   -----      Other rehabilitation                                                 [6 S] <8F...00>

50103   7/03/2007   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
50103   7/03/2007    4360      Stroke and cerebrovascular accident unsp                             [R S] <G66..00>
50104   8/03/2007   -----      Seen in speech and language clinic           CLINICAL LETTER          [O  ] <9N0Q.00>
                               [DELETED] PCT SPEECH AND LANGUAGE THEREPY
50105   9/03/2007   -----      Urine tests                                  ON CORRECT               [O  ] <46...11>
                               PRESCRIPTION/[DELETED]
50108  12/03/2007   -----      Encounter administration                     DR [DELETED]SEC RANG.    [O  ] <9N...11>
                               PATIENT HAS APPT WITH SPEECH THERAPY AT 9 AM AND APPT IN DR [DELETED] CLINIC AT
                               11 AM ON WEDNESDAY. [DELETED] SECRETARY, WANTS TO KNOW WHICH APPT SHOULD BE GIVEN
                               PRIORITY AS IT WON'T BE POSSIBLE FOR HER TO ATTEND BOTH. [DELETED] GIVEN MESSAGE
                               TO CALL [DELETED]

50109  13/03/2007 Blood pressure          DIASTOLIC      60        SYSTOLIC       110             <246..00> O/E - blood pressure re
```

Low dose ASA initiation
CVD indication

Refer to prior event

Questionnaire sent to PCPs

Over-the-counter low-dose aspirin use Questionnaire

Practice ID	Patient ID	Sex	Age on 1/1/2013

Please answer all the questions according to the information held in the patients written records by ticking the appropriate box or writing in the space provided.

1. Is the patient currently taking low-dose aspirin or in the last three months?

☐ Yes ☐ No ☐ Unknown

If YES:

- Are they taking ☐ prescription aspirin or ☐ over –the-counter aspirin?
- How long have they been taking aspirin? .....
- What is the indication for low-dose aspirin use?
  - ☐ Primary Prevention
  - ☐ Myocardial Infarction
  - ☐ Unstable Angina
  - ☐ Revascularization
  - ☐ Cerebrovascular disease (including stroke or transient ischaemic attack)
  - ☐ Other (please specify)_____

If NOT:

2. Did the patient ever take low-dose aspirin?

☐ Yes ☐ No ☐ Unknown

If yes:

- were they taking ☐ prescription aspirin or ☐ over –the-counter aspirin?
- When did they stop taking low-dose aspirin? .....
- What was the reason to stop taking low-dose aspirin?

☐ Adverse event: (list which_____) ☐ Change antiplatelet ☐ Other

- What was the indication for low-dose aspirin use?
  - ☐ Primary Prevention
  - ☐ Myocardial Infarction
  - ☐ Unstable Angina
  - ☐ Revascularization
  - ☐ Cerebrovascular disease (including stroke or transient ischaemic attack)
  - ☐ Other (please specify)_____

3. Is the patient currently or in the last three months taking *over-the-counter* one of the following drugs?

a. Proton Pump inhibitors ☐ Yes ☐ No ☐ Unknown

Please specify name of individual PPI: .....

b. H2 receptor antagonists ☐ Yes ☐ No ☐ Unknown

Please specify name of individual H2RA: .....

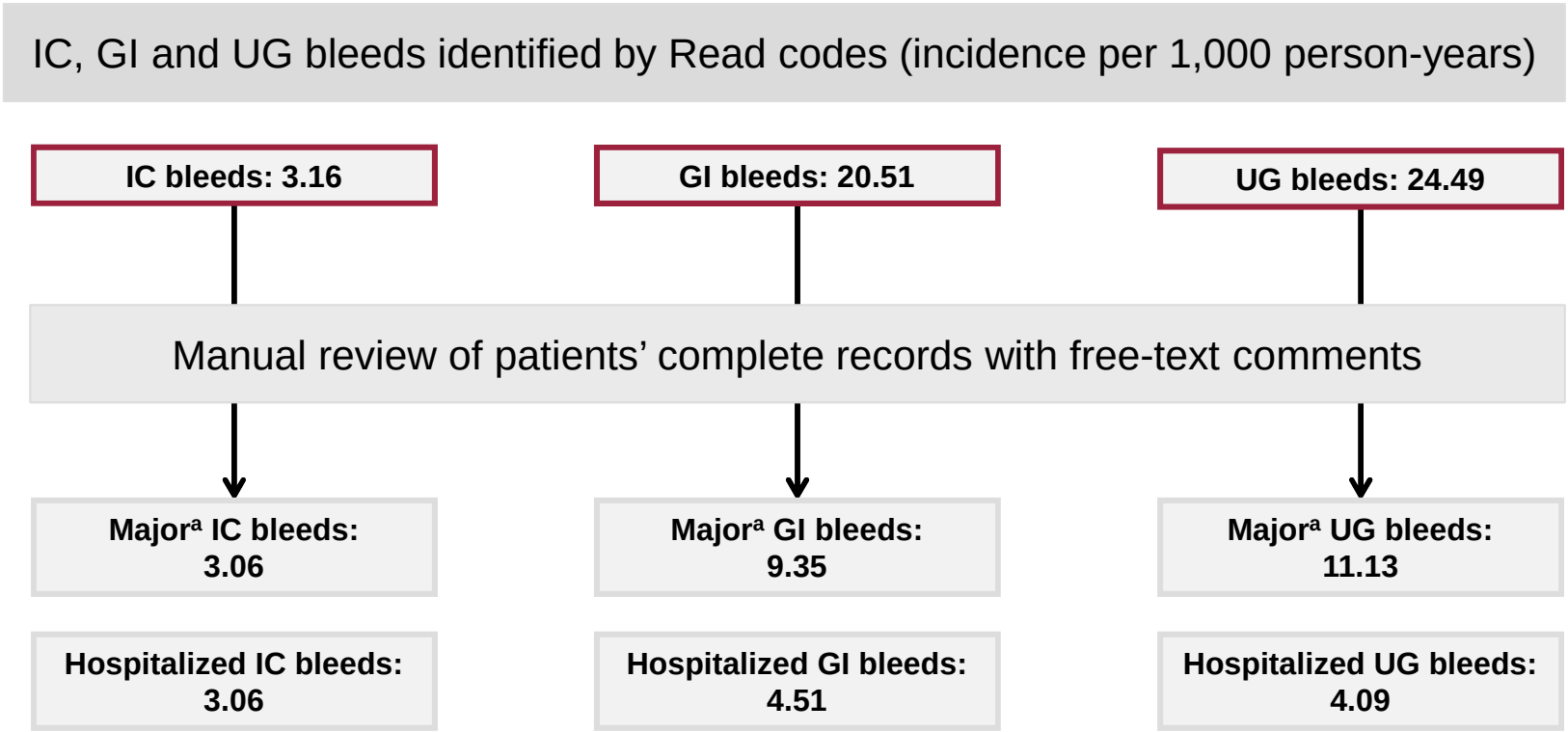
c. NSAIDs ☐ Yes ☐ No ☐ Unknown

Please specify name of individual NSAID: .....

Impact of case validation in THIN

Incidence of major bleeds among combined study cohorts

Including all bleeds identified by Read codes leads to misclassification of outcome and corresponding over-estimation of incidence of major bleeds



[a]Major bleed = requiring referral or hospitalization

GI, gastrointestinal; IC, intracranial; UG, urogenital

Outcome and exposure validation

Use of **large databases** with access to millions of patients enables the study of rare diseases or outcomes difficult to capture

Yet caution is needed:

- This phenomenal statistical power ensures precision on study results irrespective of their validity; this “amplifying” effect is particularly pernicious when erroneous data management or study design led to invalid estimates
- Requires careful planning and data analysis by experienced researchers
- Facilitate methodological mistakes translate into flawed conclusions

Analysis

Good hypothesis, good data source, and a good design might not be enough…we might spoil it in the analyses

- Issues to take into account:

  - **Discuss an analysis plan**
  - **Minimizing confounding**

    - Identify all potential confounders
    - Use the most adequate statistical test and other analyses such as stratification

  - **Sensitivity analysis**

In general, the use of complicated analytical methods is not recomended, unless they are well understood and add documented relevant contributions to conventional analyses

Summary

- **Observational studies that use electronic medical records provide valuable data and can be just as important as RCTs**
- **UK primary care data are of global significance**
- **The quality/experience of the researchers, as well as the quality of the data source, is key**