

# Data Quality in Electronic Patient Records: Why its important to assess and address

*Dr Annette Gilmore PhD, MSc (Econ) BSc, RGN*

## What this presentation covers

Why GP EPRs are important?

Uses of GP EPRs

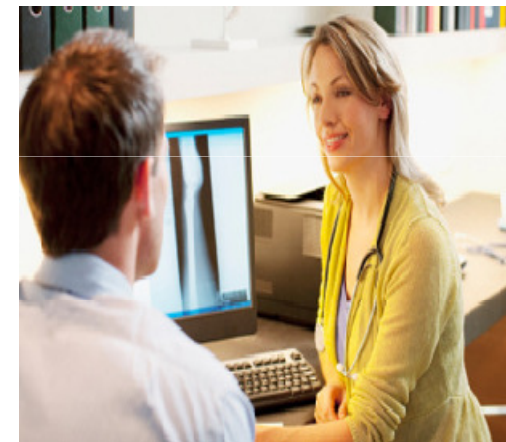
Dimensions of data quality

Assessment of EPR quality

Effects of quality on data uses

How to improve quality

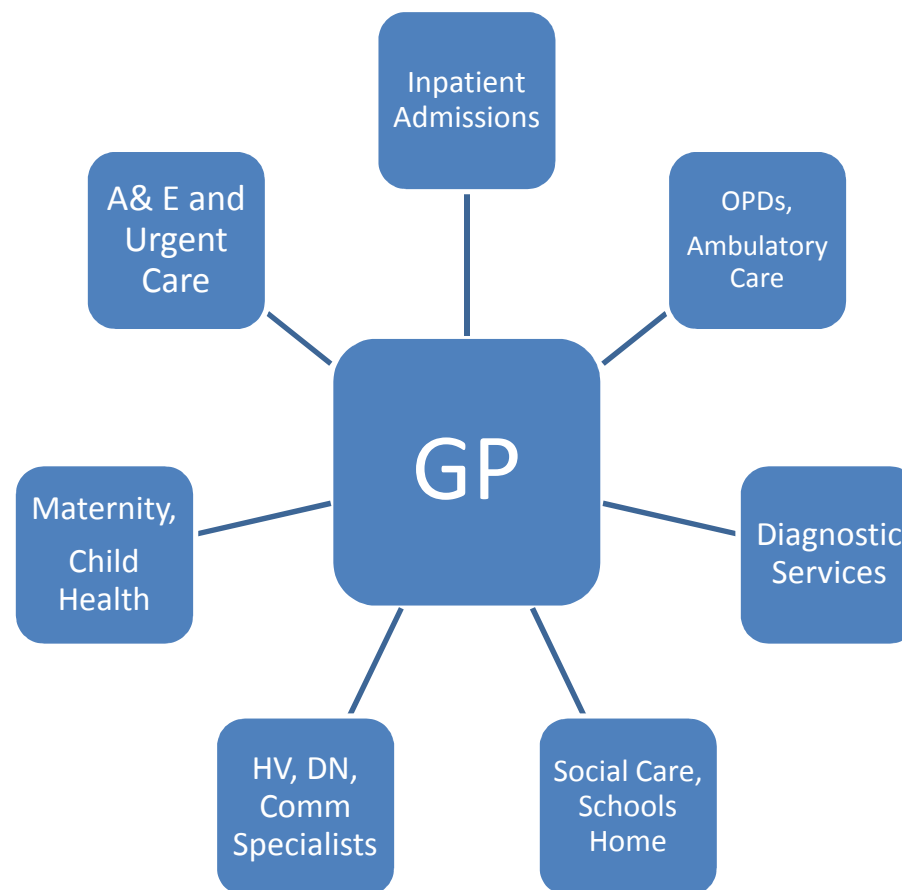
Where to get help



# Why GP EPRs are important?

- National policies and legislation to:
  - promote care in the community and peoples home
  - evaluate patient outcomes
  - ↑ linkage of health records to drive improvements in care
- Best placed to co-ordinate patients' care
- Repository for pts' complete healthcare records
- Covers total population (geographically, case identification)
- Linkage for validation with other sources- disease registries, HES, RCTs

## GP Patient information sources



From multiple healthcare settings in private and public sectors

# EPR aspirations

## Primary information uses:

- Access to complete, accurate information at point of care
- ↓ Duplicate data entry (single entry shared across eprs)
- Safer care and supporting care (i.e. decision support)

## Data reuse:

Research (epidemiology, rcts), monitoring health outcomes, comparative audit, surveillance

## Supporting:

Initiatives to improve patient safety, health outcomes, patient empowerment, health policy, commissioning etc.

Juran JM, 1988, defined quality through  
***‘fitness for use’***

In the context of EPRs interpret as data are  
of sufficient quality when they serve the  
needs of a given user pursuing specific goals  
(Gray Weiskopf and Weng, JAMIA, 2013)

***However you need to know the quality of the  
data to make this interpretation***

---

# Main Quality Features

- **Completeness**
- **Accuracy**

Assessment required to:

- Determine how the information can be used
- What adjustment to data are required
- What assumptions can be made from outputs
- Confidence in uses

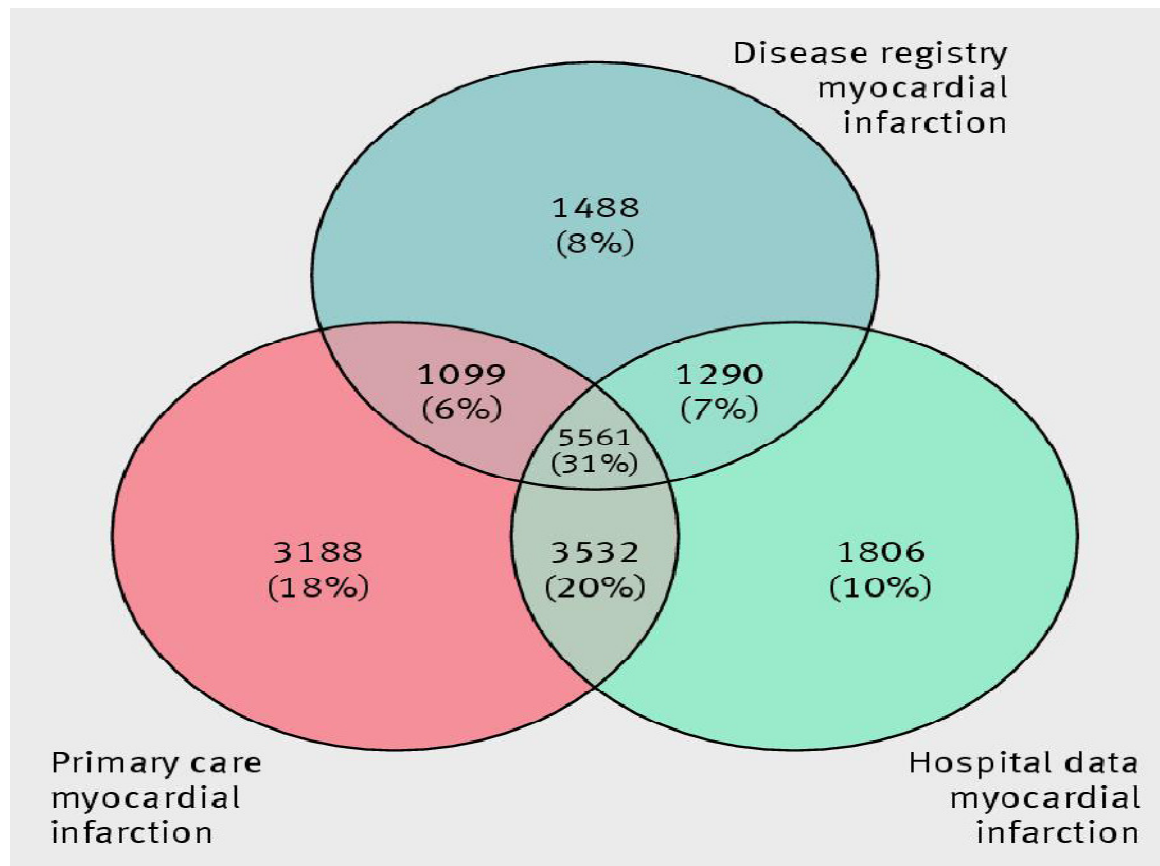
---

## Dimensions of Coverage/ Completeness:

- Representative of population (geographical)
- % Recruitment/ Case ascertainment
- Extent of dataset (variables included)
- Data collection (% variables  $\geq$  95% complete)

*Information about all 4 aspects required for  
intelligent use of data (Black and Payne, QSHC,2003)*





**Example of completeness of case identification in 3 data sources: GP, HES and Disease register**

**Fig 3** Number and percentage of records recorded in primary care (Clinical Practice Research Datalink), hospital care (Hospital Episode Statistics), and disease registry (Myocardial Ischaemia National Audit Project) for non-fatal myocardial infarction across the three sources (n=17 964 patients),

***Reproduced from Herrett et al, BMJ 2013: CALIBER study based at UCL, London***

---

## Incomplete case identification: potential issues

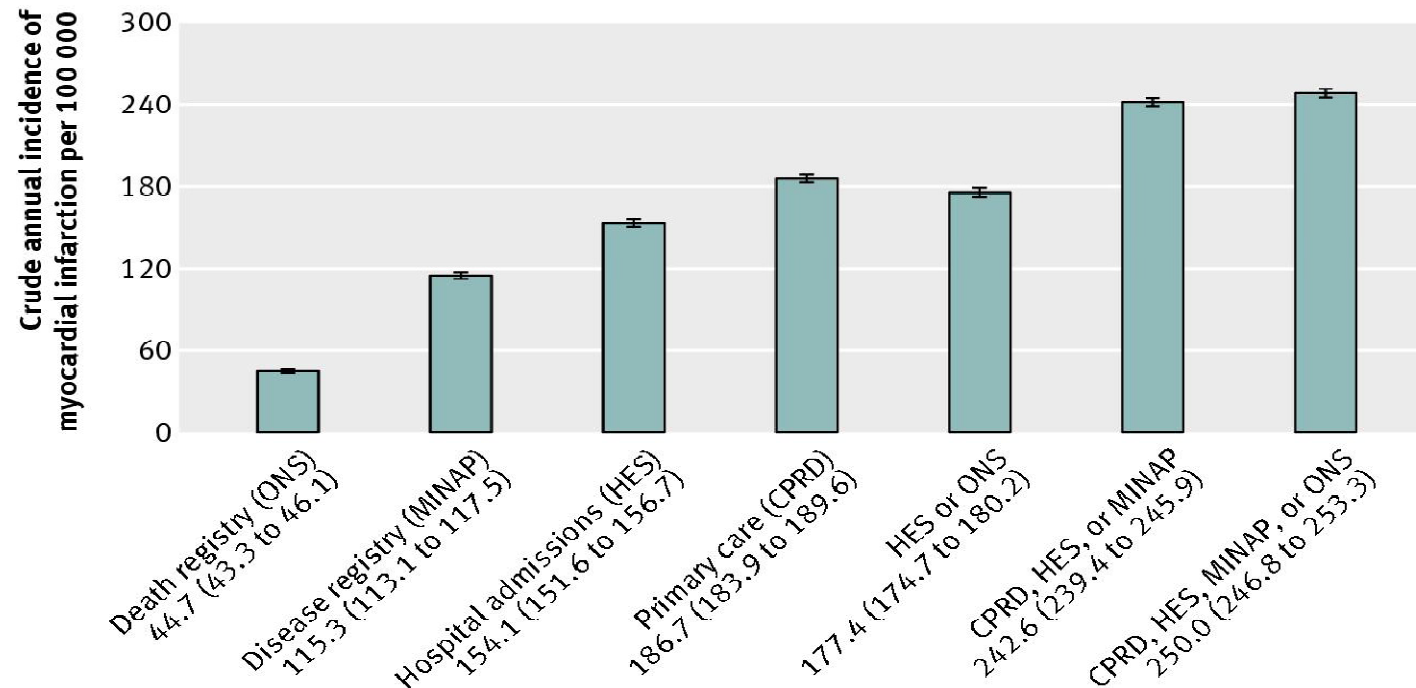
- Are missing cases random or non random?
- May cause selection bias if non random
- Usually miss mild and very severe cases
- ↑ DCOs cases (death certificate only) –red flag
- ↑ DCOs – don't know direction of bias and fixing causes selection bias

---

# Effect of missing cases

- Inflate/deflate incidence and prevalence rates
- Incorrect outcome estimates e.g. survival
- Problematic if small numbers involved
- Improving completeness alters trend analysis

## Effect of completeness variations on crude incidence rates from single and linked data sources



**Fig 1** Crude incidence of acute fatal and non-fatal myocardial infarction estimated using different combinations of data from primary care (Clinical Practice Research Datalink), hospital admissions (Hospital Episode Statistics), disease registry (MINAP, Myocardial Ischaemia National Audit Project), and death registry (Office for National Statistics). Incidence derived using denominator of all adults in the CALIBER primary care population

***Reproduced from Herrett et al., BMJ 2013. Research from CALIBER study, based at UCL, London***

---

## Areas prone to incomplete and inaccurate information

- In complex diagnoses, missing and inconclusive information and lack of objective diagnostic tests
- ambiguous definitions and non adherence to reporting rules
- time lags and availability of supporting data
- Complicated pt. pathway
- Rules and variables interpreted differently by different staff groups or centres
- Single sources of information

---

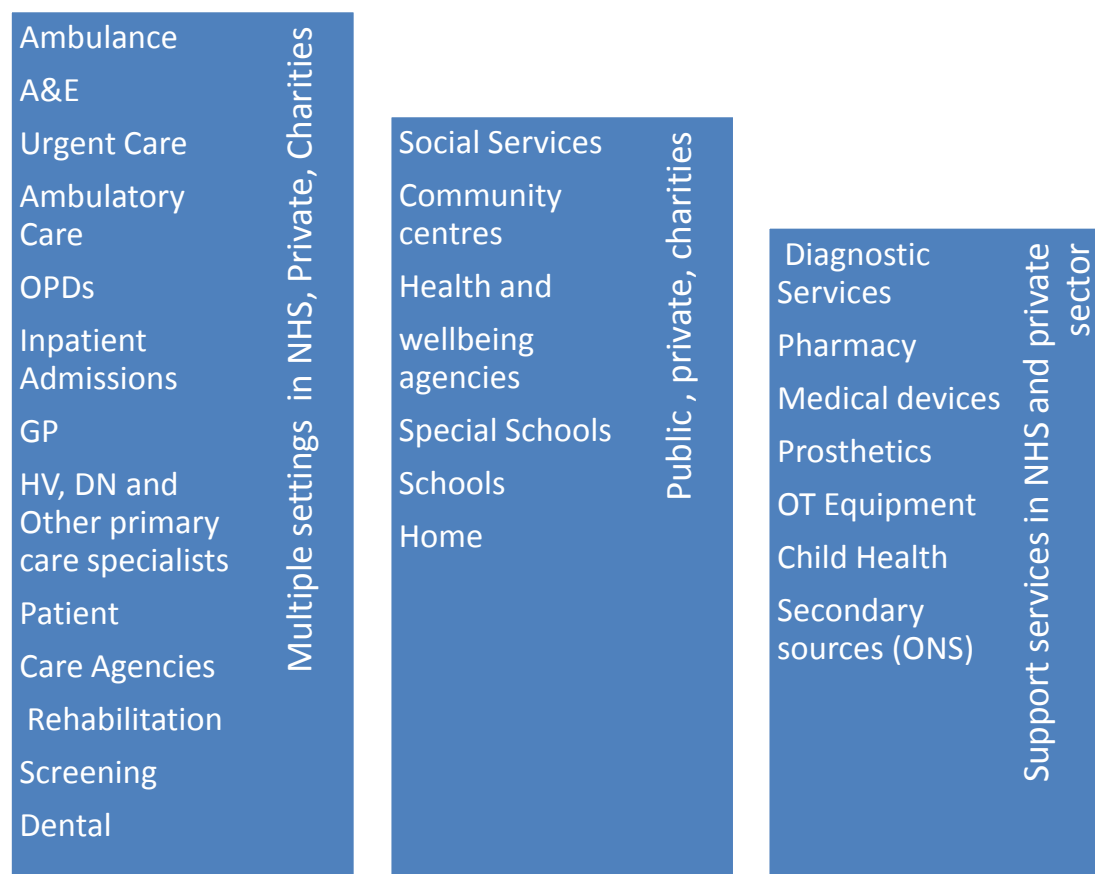
# Implications of validity and reliability problems

- False positives (inflate incidence rates)
- Underestimates benefits of interventions (screening)
- Limits comparative audit
- Outcomes analysis –requires detailed risk factors for case mix adjustment
- May cause bias in survival reporting

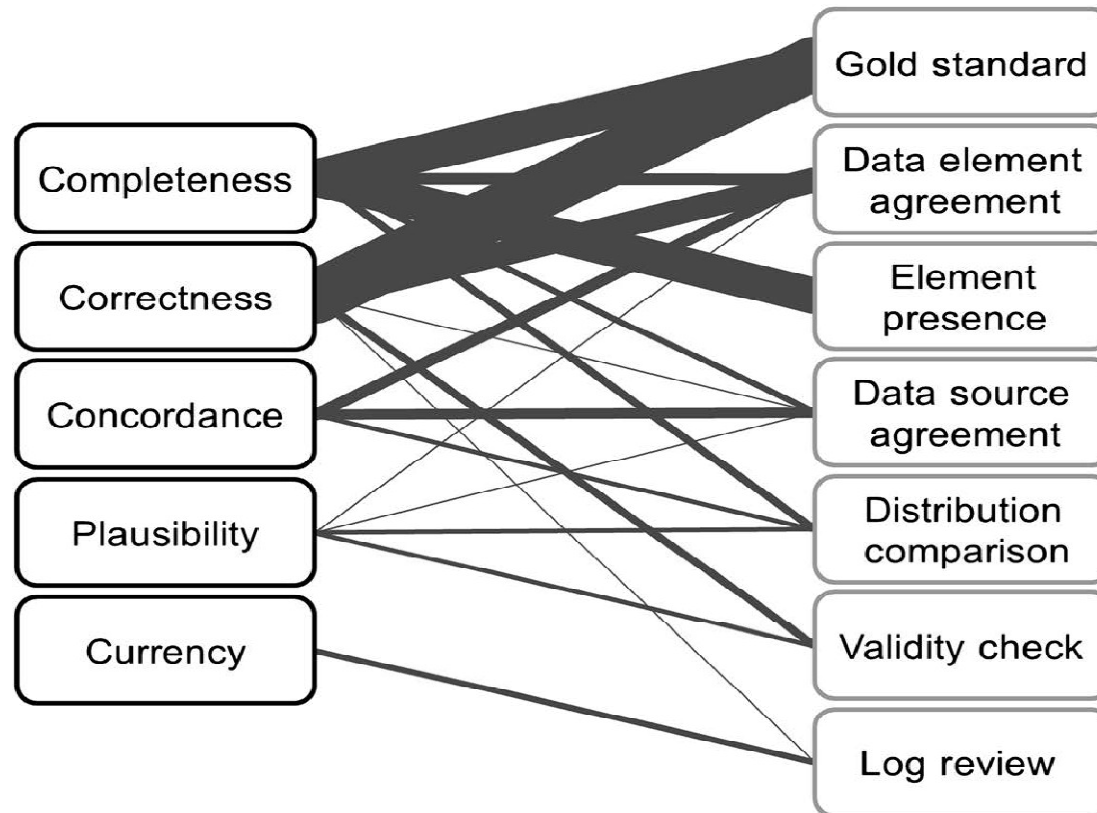
# Determinants of accuracy

- Explicit definitions of variables
- Explicit rules for data collection
- Reliability (reproducibility) of data entered/coding
- Identify requirements-quality vs quantity
- Extent to which data is validated
- Extent of collection of raw data

Example of a person's possible health and social services seeking behaviour and potential support services accessed. All these services may hold personal, health and social data on individuals.







**Figure 1** Mapping between dimensions of data quality and data quality assessment methods. Dimensions are listed on the left and methods of assessment on the right, both in decreasing order of frequency from top to bottom. The weight of the edge connecting a dimension and method indicates the relative frequency of that combination. *Reproduced from Gray Weiskopf and Weng, JAMIA, 2011*

# EPR quality assurance principles

- Clinical engagement and staff training
- Engage experts from related fields
- Linkage to multiple data sources
- Data quality expectations and incentives
- Common Standards
  - SNOMED CT (READ, ICD 10), PRSB Assurance
- Quality assurance process
  - Automated validations, range and consistency checks, statistical techniques to reduce bias, (CHART, PRIMIS)
  - Ref. to gold standard, algorithm of pt pathway, information sources and external validity and reliability studies



*A Community Interest Company owned by UK health and social care professional bodies and patient organisations (Company No. 8540834)*

- Forum for effective engagement of patient and care professional organisations and the vender community in the 4 UK nations to influence and direct care record standards development and implementation.
- Purpose is to ensure that the requirements of those who provide and receive care can be fully expressed in health and social care records.
- Is the **first point of call** for care professionals, service providers, commissioners, policy makers, professional bodies and system suppliers for expertise and all matters relating to care records.
- Visit <http://www.theprsb.org.uk/> and follow on Facebook and Twitter



***Information entered  
into EPRs are primary  
for the purpose of  
patient care  
not secondary reuse***

*Achieving interoperability  
and ensuring adherence to  
common standards is vital*

*Engaging with all the communities involved and harnessing  
the knowledge and skills already amassed is key  
to effective implementation and usage.*

Thank you

**Dr Annette Gilmore**

Researcher, IT project manager  
and Nurse Specialist  
Chair of PRSB Assurance Committee

Department of Haematology  
London North West NHS Trust

[Annette.Gilmore@nhs.net](mailto:Annette.Gilmore@nhs.net)