EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Identifying opportunities for 'Big Data' in medicines development and regulatory science.

Jim Slattery

Disclaimer

The views expressed in this talk are the personal views of the author and may not be understood or quoted as being made on behalf of or reflecting the position of the EMA or one of its committees or working parties.


Let's implement the latest catch-phrase to become more buzzword compliant.
yourecards
someecards.com

Attitudes to Big Data

- A huge advance that will allow computers to find the answers to everything
- A resource that appears promising across many areas of science and business
- A useful tool for hypothesis generation
- An asset we should try to use because it is there
- A ploy to attract funding for data scientists

Big Data

Data that arise as a product of the wide-spread use in human activities of digital equipment with storage capabilities

- Examples of human activities are communication, sales, professional consultations, monitoring, mapping etc
- Includes most data sources advertised as ‘Real World’.

Sometimes collected specifically for hypothesis generation but more often not

DATA/KNOWLEDGE

HYPOTHESIS

EXPERIMENT

DATA

ANALYSIS

NEW KNOWLEDGE

Often collected only to address the specific hypothesis

Potential use - 1

Hypothesis generation

- **Signalling safety issues**
- **Suggesting new directions for research**

Potential uses - 2

Elucidating the causal process underlying an association. In the pharmaceutical area these processes are those linking the products to alterations in the health of the recipient.

- **Appropriate targeting of medicines**

Potential uses - 3

Regulatory decision making

- **Supporting assumptions**
  - **Validation of surrogate outcomes**
  - **Validation of modelling and simulation**
  - **Extending clinical trial data**
    - **Longer term outcomes**
    - **Clinical pathways for cost-effectiveness analyses**
- **Outcome evaluation of regulatory interventions**
- **Evaluation of safety concerns**
- **Evaluation of efficacy**

A BIG QUESTION

How do we decide if the data are ‘fit for purpose’?

Fundamental requirement

- **Data whose form might be influenced by subject matter we are interested in investigating.**
- **Sometimes barn-door. E.g. Gene data.**
- **More often, for Big Data, the area of interest will only affect a small part of the data and hence preliminary analysis involves statistical screening to identify the relevant parts of the data. E.g. Drug effects from social media.**

Additional considerations

- Are the data available at the right time?
- Are they easily accessible?
- Are they sufficiently reliable for the current decision?
- Are they the only way of answering the question?
- If not, do they add value when used in parallel with the alternative methods?

An expert opinion on Big Data

> Things can go wrong, but that's where the power of big data comes in. If you're looking at ten tweets and you're getting a few wrong, you've got problems. If you're looking at ten billion tweets, basically it washes out as noise. The real patterns are the ones that survive the noise.
>
> Kalev Leetaru

- Stated objective may look like hype but is in fact quite modest - to find ‘real patterns’
- Real patterns are either truth or consistent untruths

Big data for hypothesis generation (1)

- Hypothesis generation is not free of prior conceptions
  - We know what kind of ideas interest us
  - Hence fundamental requirement can be applied for the class of interesting hypotheses
- Low standards of evidence may be acceptable
- **BUT** Every hypothesis carries a cost associated with further verification

Big data for hypothesis generation (2)

- Are the data we are considering the best way of generating this class of hypotheses?
- If not, do they add anything new to the other available sources of hypotheses?
- What is the (incremental) cost of generating hypotheses from these data?

Example 1

- Rapid learning for precision oncology
  - Statistical reverse engineering methods to hypothesize the putative driver networks for a given patient's tumour
  - Information from 'exceptional' responders
  - Information about cancer associated mutations
  - Feedback from personal treatment episodes

Shrager J, Tenenbaum J. Rapid learning for precision oncology. Nat. Rev. Clin. Oncol 11, 109-118 (2014)

Rapid learning for precision oncology

- Method of analysis unclear
- Vast assumptions. Responders or just resilient?
- Huge multiplicity issues and numerous sources of bias

- But, if no established treatment strategy, can a system like this help your doctor to choose a promising treatment?
- A valid use of big data but not a new paradigm for drug development

Example 2 Detection of unknown ADRs

- From spontaneous reports: large body of research characterising performance of methods and added value compared to other established pharmacovigilance techniques. Routine use at EMA.
- From EHR: An evolving area with mixed results
- From social media: early research shows:
  - That ADRs are mentioned in social media and can sometimes be mapped to formal coding dictionaries
  - As yet no clear characterisation of signal detection performance
  - No proof of added value compared to SR data

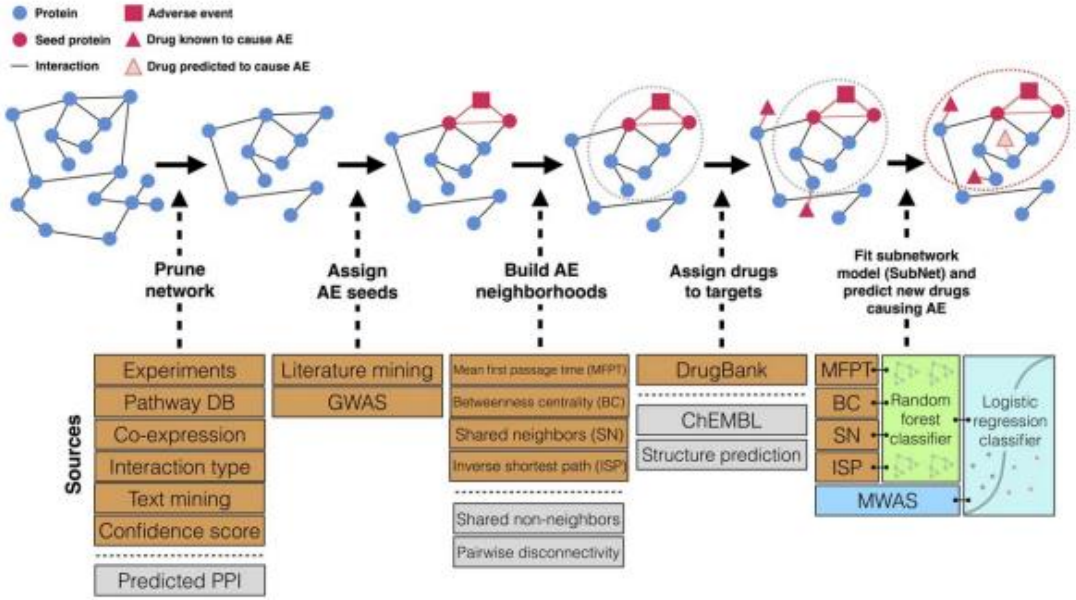
Freifeld C, et al. Digital drug safety surveillance: Monitoring pharmaceutical products in Twitter. Drug Saf 2014: 37: 343-350

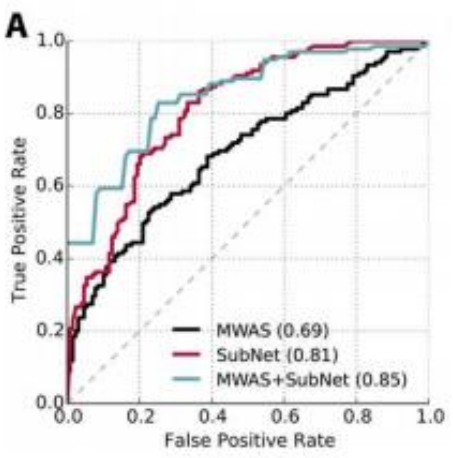
Big data for model building

- Are data available on variables that may modify the effects of drugs?
- Person/group-level data - Need these data to be linked to individuals or groups with known drug exposures and outcomes.
- Can the data be used to give prior probabilities for selecting among hypotheses?
- Growth area in research

Example – Systems pharmacology

- Using data on drug's target proteins and pathways to guide ADR detection





Lorberbaum T, Nasir M, Keiser M, Vilar S, Hripsak G, Tatonetti N. Systems pharmacology augments drug safety surveillance. Clin Pharm & Ther 2015: 97(2): 151-158

Messages

- Some quite non-specific information about chemical interactions seems to improve detection of adverse drug reactions
- Big data used to inform prior beliefs looks like a very exciting area for development

Regulatory decision: Safety

- Use of observational data is well-established in safety analyses.
- Often based on large detailed clinical datasets collected for routine clinical care
- Often the highest standard of evidence available
- Evaluation of such evidence is already a major component of the work of the Pharmacovigilance Risk Assessment Committee.

Example: Fluoroquinolones + Retinal Detachment

Study name	Rate ratio	Lower limit	Upper limit	p-Value
Etminan et al, 2012 (case-control)	4.50	3.56	5.69	0.00
Fife et al, 2014 (case-control_CCAE)	1.39	1.01	1.91	0.04
Fife et al, 2014 (case-control_Optum)	1.19	0.68	2.07	0.54
Kapoor et al, 2014 (cohort)	1.82	0.26	12.89	0.55
Eftekhari et al, 2014 (cohort)	0.56	0.08	4.03	0.56
Pasternak et al, 2013 (cohort)	1.29	0.53	3.13	0.57
Kuo et al, 2014 (cohort)	2.07	1.45	2.96	0.00
Chui et al, 2014 (case series)	1.26	0.65	2.46	0.50
Fife et al, 2014 (case series_CCAE)	1.13	0.99	1.29	0.07
Fife et al, 2014 (case series_Optum)	0.85	0.66	1.09	0.20
	1.47	0.95	2.27	0.09

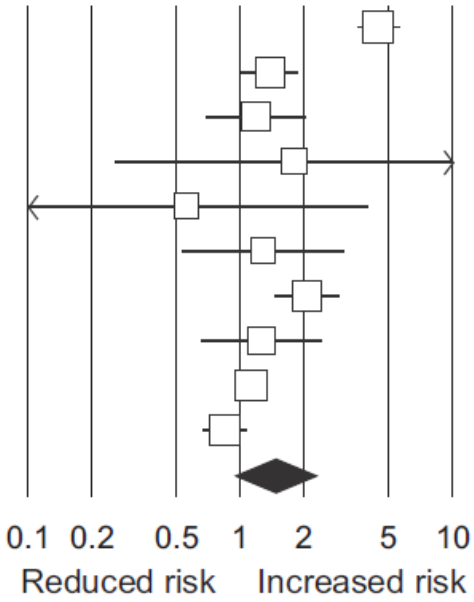


Fig. 2. Pooled rate ratio and 95% CI of retinal detachment associated with fluoroquinolones.

Alves C, Penedones A, Mendes D, Marques F. A systematic review and meta-analysis of the association between system fluoroquinolones and retinal detachment. Acta Ophthalmol. 2016: 19: e251-e259

Context

- These are all carefully designed studies by researchers who understood the data and adjusted estimates appropriately.
- $I^{2,}$ a statistic that describes the proportion of total variation in study estimates that is due to heterogeneity, was 92.8%.
- For PRAC to reach agreement on results of this nature requires extended deliberation.
- Use of RWE is essential – but not a quick or easy option!

Regulatory decision: Verifying Efficacy

- Extend time horizon for efficacy
- Validating assumptions of modelling and simulation
- Linking surrogate outcomes to clinically important effects
  - E.g. Is BMD associated with fractures?
    - In untreated subjects?
    - In treated patients?
      - In patients with this specific treatment?

Context

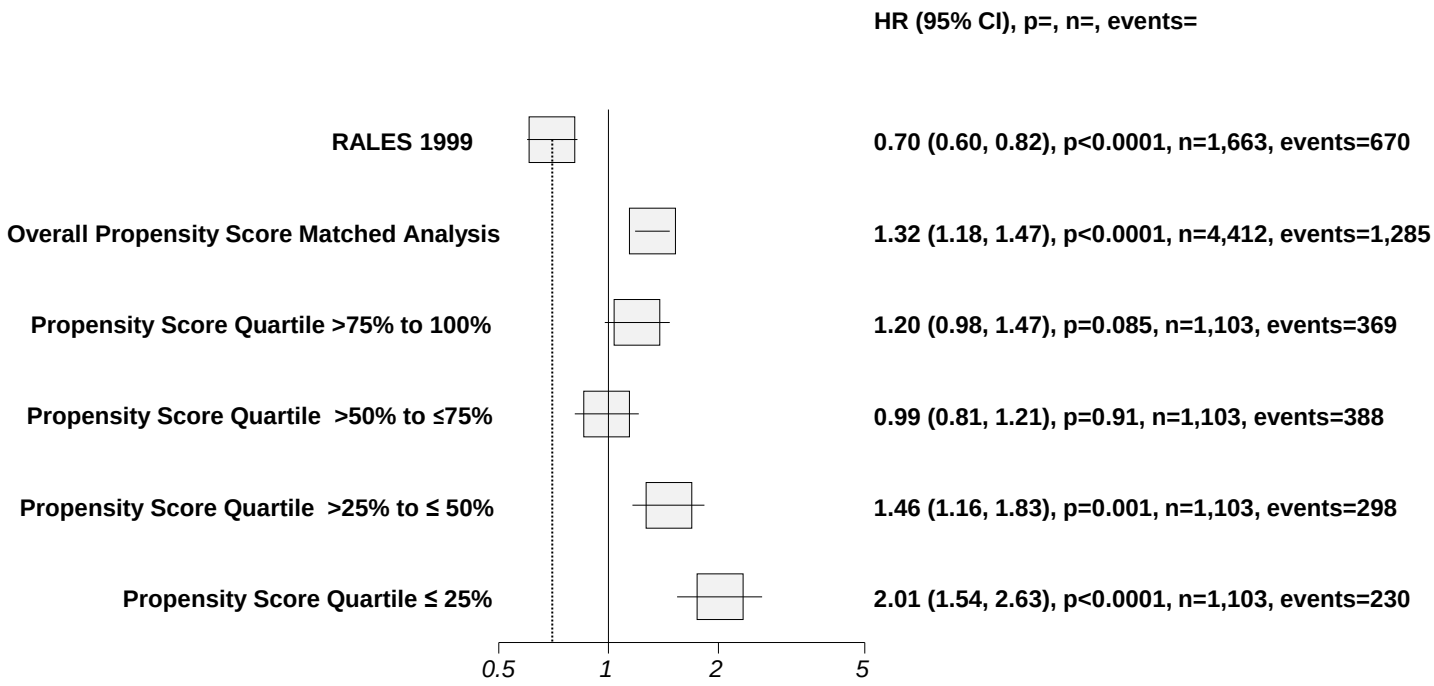
- Attempts using RWE/Big data to extend limited but reliable observations within the same patient group can start by matching the initial results – A reality check.
- Even if the absolute results do not match it may be reasonable to estimate relative measures from the observational data.
- RWE/Big data often supply insights not otherwise obtainable but need careful interpretation

Level of evidence for efficacy?

- Spironolactone in heart failure
- Case study attempting to replicate results of randomised clinical trial

Freemantle N, Marston L, Walters K, et al. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ 2013;347:f6409.

Propensity scores, confounding by indication and other perils for the unwary in observational research



Take home messages for propensity score based analyses of treatment effects

- Start from somewhere you know
  - Replicating an existing trial
- Examine the behaviour of the propensity score across its range
  - Test for interaction with exposure
- Estimates of efficacy for treatments given according to well-defined rules based on known variables are less likely to be biased.
- Doubts about standard of evidence may limit regulatory use.

Regulatory decision: Efficacy in new patient groups

- Pre-authorisation RCTs enrol patients for whom no licence yet exists
- Exposure of patients beyond the current label demands extreme caution
- Where would observational data on a well-characterised group of off-label patients arise?
- Main limitation of observational approach.

Efficacy: An aside

RCT: The right answer to (maybe) the wrong question

OR

Observational study: The (maybe) wrong answer to the right question

If we chose to put one of these right, which would it be?

Conclusions

1. **Optimism:** Big data/RWE hold promise. They are already used productively in safety evaluation and can provide an evidence-base in confirmation of efficacy and support of modelling assumptions.
2. **Realism:** Big Data/RWE have multiple uses in regulation but no data analytics approach can free the analyst from having to understand the data
   - Big Data are an accumulation of diverse datasets with heterogeneous properties
   - Regulatory data uses are also diverse with heterogeneous requirements
   - Every data source must be evaluated relative to its potential use
   - Clear and rational criteria for use of each dataset in each purpose are required.

Conclusions

3. **Caution:** Some regulatory decisions require very robust and reliable standards of evidence and this limits use of observational data.
4. **Future:** EMA considers that RWE might have a role in medicines regulation as a source of insights and it will explore this in the future by working with stakeholders to evaluate the reliability and utility of available sources of big data.

Real World Data Workshops at EMA

EUROPEAN MEDICINES AGENCY