PROTECT

imi efpia

Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium

The PROTECT symposium

Training: Testing a statistical signal detection system

February 2015

PROTECT is receiving support from the Innovative Medicine Initiative Joint Undertaking (www.imi.europa.eu), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

PROTECT Goal

To strengthen the monitoring of benefit-risk of medicines in Europe by developing innovative methods

to enhance early detection and assessment of adverse drug reactions from different data sources (clinical trials, spontaneous reporting and observational studies)

to enable the integration and presentation of data on benefits and risks

These methods will be tested in real-life situations.

Work Package 3: Signal Detection

Objective:

To improve early and proactive signal detection from spontaneous reports, electronic health records, and clinical trials.

Key learning objectives

- Why do we need with empirical evaluation of statistical signal detection
  - PROTECT Recommendations
  - Advantages and disadvantages of retrospective and prospective evaluation
- Describe a Reference Dataset
  - What kind of controls to use?
  - Access to the PROTECT dataset. How to download it, how to adapt it, how to construct a new dataset if none of your products are present
- Metrics for quantitative performance
  - How we choose them, what they measure
  - Sensitivity, precision (PPV), time to detection, specificity, NPV
  - Why specificity was not used
- Thresholds
  - Why they are important
  - How to choose them
- SAS macros
  - Where they are, general idea of how they work
  - Output: how to use it

Statistical Signal Detection

- Start with a set of drugs of interest and a classification of adverse events – for this discussion we assume that these decisions have been made.

- Calculation of summary statistics for each drug-AE pair in a database designed to draw attention to those pairs that might need further investigation.

Desirable properties for the SD method

- It should detect ADRs in a reasonable time frame.
  - Consider relative to alternatives
- It should not highlight too many false positives
- Its performance in these respects should be stable and predictable.

Choice of SD method

- There are a large number of different methods available
- An important question is whether the method currently in use is as good as it can be
- Where do we start in answering a question like this?

WP3 Recommendation B1.3

For moderate to large spontaneous report databases, the relative performance of a quantitative signal detection algorithm in one database can be predicted from research in other databases.

Rationale: In the PROTECT study, signal detection algorithms with good signalling properties compared to other signal detection algorithms in one spontaneous report database also had relatively good signalling properties in other spontaneous report databases. The databases were both regulatory and company based and ranged in size from about 500,000 to 5,000,000 reports. Hence relative performance in moderately large databases can be reliably inferred from evaluations in other settings.

WP3 Recommendation B1.4

Absolute performance of the selected quantitative signal detection algorithm must be validated in the target spontaneous report database.

Rationale: Although the relative performance of signal detection algorithms is similar in different spontaneous report databases the absolute performance characteristics may vary substantially. Hence it is advisable to test the chosen disproportionality statistic with a range of signal detection algorithms within the target database.

Source: Candore G, Juhlin K, Manlik K, Thakrar B, Quarcoo N, Seabroke S, Wisniewski A, Slattery J. Comparison of statistical signal detection methods within and across databases. *In Press.*

In simple terms

- At least one evaluation of signal detection performance is needed within the specific reporting database to give a baseline against which to predict how changes of SD method will affect the ability of the system to highlight ADRs and the changes in workload for signal detection assessors

PROTECT

Options

- Prospective
- Retrospective

Prospective evaluation

- Record as they arise:
  - drug-AE pairs highlighted by method.
  - nature and outcome of clinical evaluation and regulatory processes
  - ADR detection arising from other SD processes but not highlighted by statistical method

Good and bad of prospective

- Strength
  - Should be doing this anyway as a QA measure and to check for any changes in process
  - Post SDR decision process is 'real life'

- Weakness
  - Takes a long time to accumulate data for precise estimates
  - Can only test method that was in use – which may have evolved over time

Retrospective evaluation

- Reconstruct SD calculations in accumulated data using any chosen SD method.
- Compare against a reference database of ADRs

Good and bad of retrospective

- Strength
  - Reproducible process
  - Give estimates comparable to studies like PROTECT
  - Once set up can quickly test effect of changes to system – I.e Multiple algorithms can be tested

- Weakness
  - Requires preparation of reference dataset
  - Requires some technology (explained later)
  - Assumes that statistical findings would have stimulated further evaluation of drug-AE pair

Focus for this session

Retrospective evaluation

- **Reference dataset**
- **How to access and use routines used in PROTECT to evaluate performance**

PROTECT

Reference Standard

- We assume the best current information about ADRs, often the final resolution of the regulatory process, is true.

- Thus we can categorise the SDRs as either related to an established ADR or not known and, occasionally, not ADRs. Experience shows that most SDRs are false positives and we assume for analysis that these latter are not ADRs

Reference database

- Records current knowledge about ADRs
- Essential fields:
  - Product. Formally coded
  - ADR. Formally coded
- Very useful:
  - Date of signal for ADR

Reference dataset used in PROTECT

- Based largely on PROTECT WP3.3 – Structured database of SPC Section 4.8 for CAPS
- Augmented with some additional products from companies based on core safety databases.
  - Note that these do not accord perfectly with SPC
- Date of first signalling of ADRs was not available – only pre- versus post-authorisation for CAPS
  - Field not yet available in public access version

Address for PROTECT reference database

- http://www.imi-protect.eu/methodsRep.shtml

Note: The ADR database for CAPs is found at the same address

An aside about the WP3.3 ADR database

- Designed to integrate ADR data into automated systems

- Multiple uses

Structured database of SPC 4.8

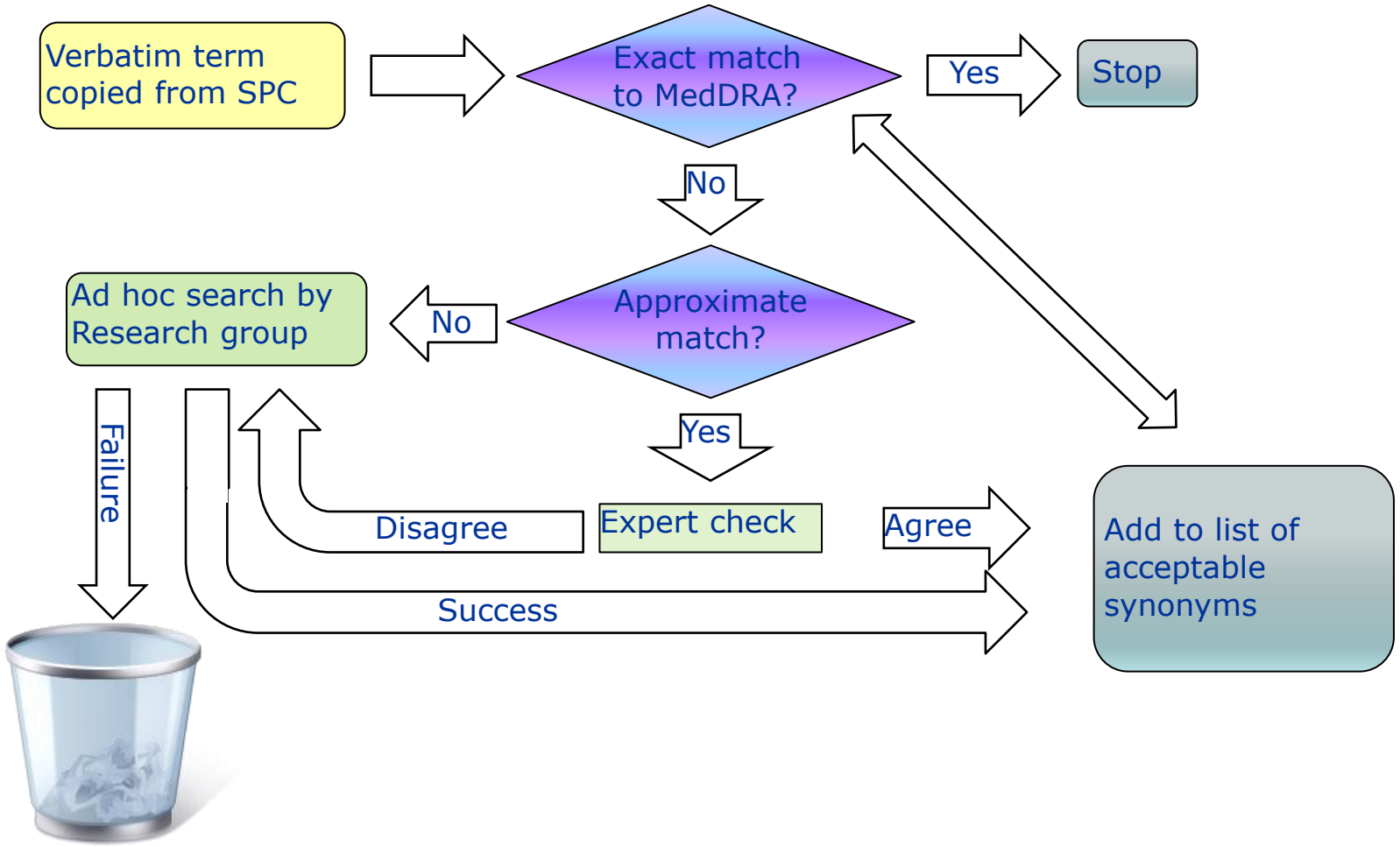
- Objective

  Making available, in a **structured** format, already known ADRs to allow for:

  - Automated classification of SDRs as known ADRs
  - Automatic reduction of masking effects

- Current status
  - Database for centrally authorised products (CAP) fully implemented
  - Contributed to gold standard for 3.01
  - Maintenance procedure agreed
  - Published on PROTECT website
  - Extension to non-CAP products being tested

How was it constructed?



Structured database of SPC 4.8

- Fuzzy text matching (automatic algorithm) to match MedDRA terms from manual extracted ADRs from the SPCs
  - Stemming, Stop words, Permutations, Synonyms and Spelling variations
  - ➲ **Sensitivity of verbatim matching increased from 72% → 98%**

Drug	SPC Term	Verbatim match	Fuzzy matching algorithm
Aclasta	FLU-LIKE SYMPTOMS		Flu symptoms
Advagraf	OTHER ELECTROLYTE ABNORMALITIES	-	Electrolyte abnormality
Advagraf	PAIN AND DISCOMFORT	-	Pain and discomfort NEC
Advagraf	PRIMARY GRAFT DYSFUNCTION	-	Primary graft dysfunction*
Advagraf	PRURITUS	PRURITUS	Pruritus*
Advagraf	PSYCHOTIC DISORDER	PSYCHOTIC DISORDER	Psychotic disorder*
Advagraf	PULSE INVESTIGATIONS ABNORMAL	-	***Investigation abnormal***
Advagraf	RASH	RASH	Rash*
Advagraf	RED BLOOD CELL ANALYSES ABNORMAL	-	***Red blood cell analyses****
Advagraf	RENAL FAILURE	RENAL FAILURE	Renal failure*
Advagraf	RENAL FAILURE ACUTE	RENAL FAILURE ACUTE	Acute renal failure, Renal failure acute*
Advagraf	RENAL IMPAIRMENT	RENAL IMPAIRMENT	Renal impairment*
Advagraf	RENAL TUBULAR NECROSIS	RENAL TUBULAR NECROSIS	Renal tubular necrosis*
Advagraf	RESPIRATORY FAILURES	-	Respiratory failure, Failure respiratory
Advagraf	RESPIRATORY TRACT DISORDERS	-	Respiratory tract disorders NEC
Advagraf	SEIZURES	-	Seizure, Seizures*
Advagraf	SHOCK	SHOCK	Shock*

Better option:
Red blood cell abnormal

Structured database of SPC 4.8 – published

http://www.imi-protect.eu

PROTECT

imi efpia

Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium

Home | Contact Us | Search

PROJECT
- About PROTECT
- Objectives
- Governance structure
- Partners
- Work programme

News

Results

General Presentations

eRoom - partners only

Links
- General Links
- Collaborations
- Training Opportunities
- Pregnancy Study
- Adverse Drug Reactions Database NEW
- Drug Consumption Databases in Europe NEW

Welcome to the PROTECT website!

The Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium is a collaborative European project that comprises a programme to address limitations of current methods in the field of pharmacoepidemiology and pharmacovigilance. The European Medicines Agency (EMA) is the coordinator of PROTECT and GSK is the deputy co-ordinator of PROTECT. They manage a multi-national consortium of 33 partners including academics, regulators, SMEs and EFPIA companies

Latest

February 2013

Paste | Format Painter | Clipboard | Font | Alignment | Merge & Center | Number | Conditional Formatting | Format as Table | Cell Styles | Styles | Insert | Delete | Format | Cells | Clear | Sort & Filter | Find & Select | Editing

J30

Bold (Ctrl+B)

Make the selected text bold.

	A	B	C	D	E	J
1	PRODUCT	SCIENTIFICGROUP_ID	SUBSTANCE	DATE OF THE SPC	ADR AS IT APPEARS IN THE SPC	MEDDRA PT
2	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ABDOMINAL DISCOMFORT	ABDOMINAL DISCOMFORT
3	Abilify	17061	ARIPIPRAZOLE	2011-01-21	AGITATION	AGITATION
4	Abilify	17061	ARIPIPRAZOLE	2011-01-21	AKATHISIA	AKATHISIA
5	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ALLERGIC REACTION	HYPERSENSITIVITY
6	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ALOPECIA	ALOPECIA
7	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ANAPHYLACTIC REACTION	ANAPHYLACTIC REACTION
8	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ANGIOEDEMA	ANGIOEDEMA
9	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ANOREXIA	DECREASED APPETITE
10	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ANXIETY	ANXIETY
11	Abilify	17061	ARIPIPRAZOLE	2011-01-21	ASPIRATION PNEUMONIA	PNEUMONIA ASPIRATION
12	Abilify	17061	ARIPIPRAZOLE	2011-01-21	BLOOD GLUCOSE FLUCTUATION	BLOOD GLUCOSE FLUCTUATION
13	Abilify	17061	ARIPIPRAZOLE	2011-01-21	BLOOD GLUCOSE INCREASED	BLOOD GLUCOSE INCREASED
14	Abilify	17061	ARIPIPRAZOLE	2011-01-21	BLURRED VISION	VISION BLURRED
15	Abilify	17061	ARIPIPRAZOLE	2011-01-21	BRADYCARDIA	BRADYCARDIA
16	Abilify	17061	ARIPIPRAZOLE	2011-01-21	CARDIAC ARREST	CARDIAC ARREST
17	Abilify	17061	ARIPIPRAZOLE	2011-01-21	CEREBROVASCULAR ADVERSE REACTIONS	CEREBROVASCULAR ACCIDENT
18	Abilify	17061	ARIPIPRAZOLE	2011-01-21	CHEST PAIN	CHEST PAIN
19	Abilify	17061	ARIPIPRAZOLE	2011-01-21	COMPLETED SUICIDE	COMPLETED SUICIDE
20	Abilify	17061	ARIPIPRAZOLE	2011-01-21	CONSTIPATION	CONSTIPATION
21	Abilify	17061	ARIPIPRAZOLE	2011-01-21	DEEP VEIN THROMBOSIS	DEEP VEIN THROMBOSIS
22	Abilify	17061	ARIPIPRAZOLE	2011-01-21	DEPRESSION	DEPRESSION
23	Abilify	17061	ARIPIPRAZOLE	2011-01-21	DIABETES MELLITUS	DIABETES MELLITUS
24	Abilify	17061	ARIPIPRAZOLE	2011-01-21	DIABETIC HYPEROSMOLAR COMA	DIABETIC HYPEROSMOLAR COMA

How to construct your own dataset

- Search the protect reference dataset for any products of interest to you
- Address for dataset: http://www.imi-protect.eu
- Check for recent additions to SPC section 4.8 for these products and add manually
- Decide which other products you would like to include
- NOTE: It is not necessary to include ALL your products but a reasonably sized representative sample would be good.
- Coding process as already described.

Coding

- Core safety data – if already MedDRA based – may be best starting point.
- If not, copy and paste verbatim terms from SPC to a spreadsheet or database.
- Note that the PROTECT ADR database includes verbatim terms from many products and the selected mapping in MedDRA. This may help if no exact match.
- If large numbers of still unmatched terms, consider approximate matching process.
- Consider making your coding public!

QUANTITATIVE PERFORMANCE METRICS

The Screening Paradigm

- Defined population who are either cases or not
- Uniform test applied independently to each member of population
- Gold standard information
- Usual measures - Sensitivity and Specificity

- Is this applicable to analysis of spontaneous reports?

Screening paradigm

	Positive	Negative
Case	True Positive	False negative
Not case	False positive	True negative

Sensitivity = TP/(TP+FN)

Specificity = TN/(TN+FP)

A note on terminology

True Positive => Case

Case ≠> True Positive

Corollary: In SD context avoid terms 'positive control' and 'negative control'

PROTECT

Sensitivity in Signal Detection?

- Requires a well defined set of cases
- We have a fairly large number of well established ADRs
- But we suspect there are many we don't know
- Can construct a 'reference' if not a 'gold' standard
- Sensitivity will be underestimated

PROTECT

Specificity in Signal Detection?

- Requires a well defined set of non-cases
- MedDRA PTs are heterogeneous set of events
  - Wide range of clinical importance
  - Some may occasionally occur as ADRs. Others never will.
  - Some are more likely than others with specific products
  - Thus not all drug-AE pairs are potential ADRs
- No credible standard – but several have been tried

Message: Should be cautious about beating problem into screening framework

So if not specificity – what?

PROTECT

What matters in signal detection?

Detecting ADRs	**Sensitivity**
Avoiding too many clinical assessments	**Positive predictive value a.k.a Precision**
Finding the ADRs quickly	**Time to SDR**

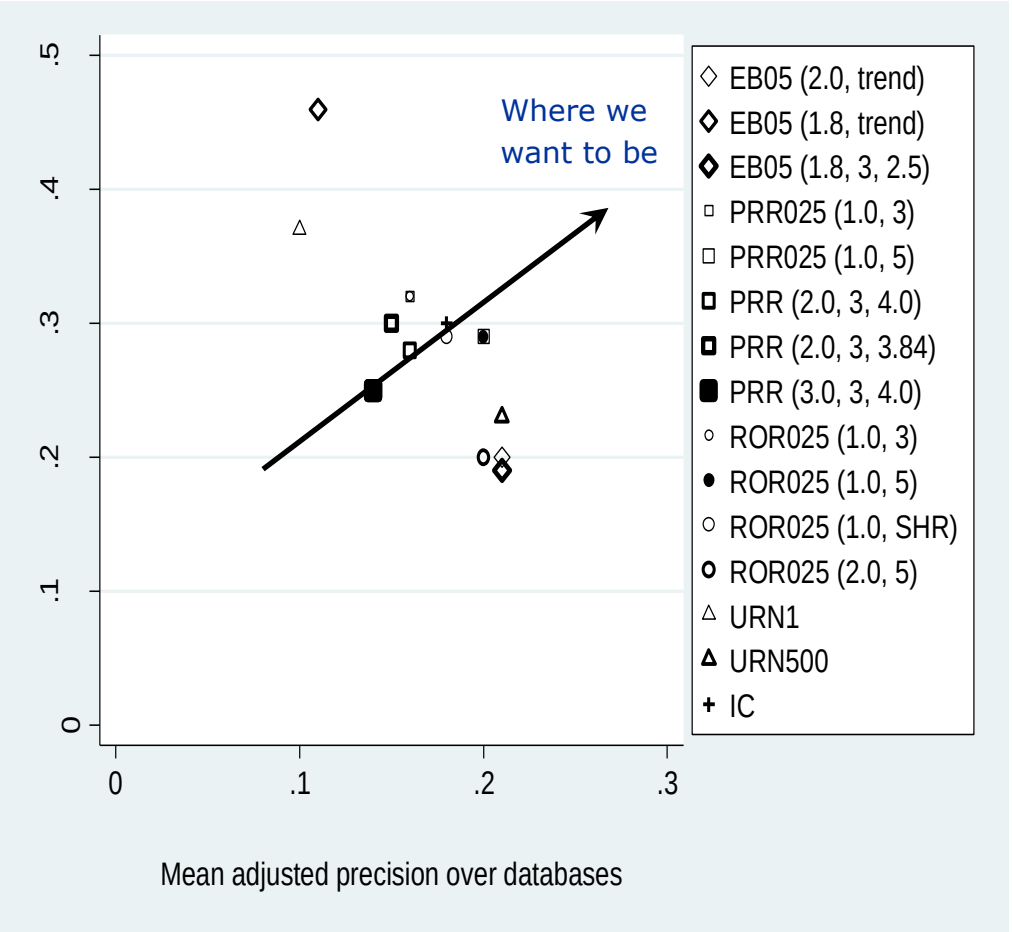
Screening for ADRs

	SDR	Not SDR
ADR	True Positive	False negative
Not ADR	False positive	TRUE NEGATIVE

Sensitivity = TP/(TP+FN)

PPV= TP/(TP+FP)

Graphical presentation of results



Thresholds

- ## Why needed

  To define when a signal of disproportionate reporting (SDR) has been observed

  Usually trigger the decision to investigate further a potential adverse reaction

- ## What they are

  A set of rules

  Usually based on:

  - the observed value of the disproportionality measure (either absolute or lower limit of confidence interval)
  - absolute number of reports for a particular drug/adverse event combination

  Other rules can be applied (trend flags,…)

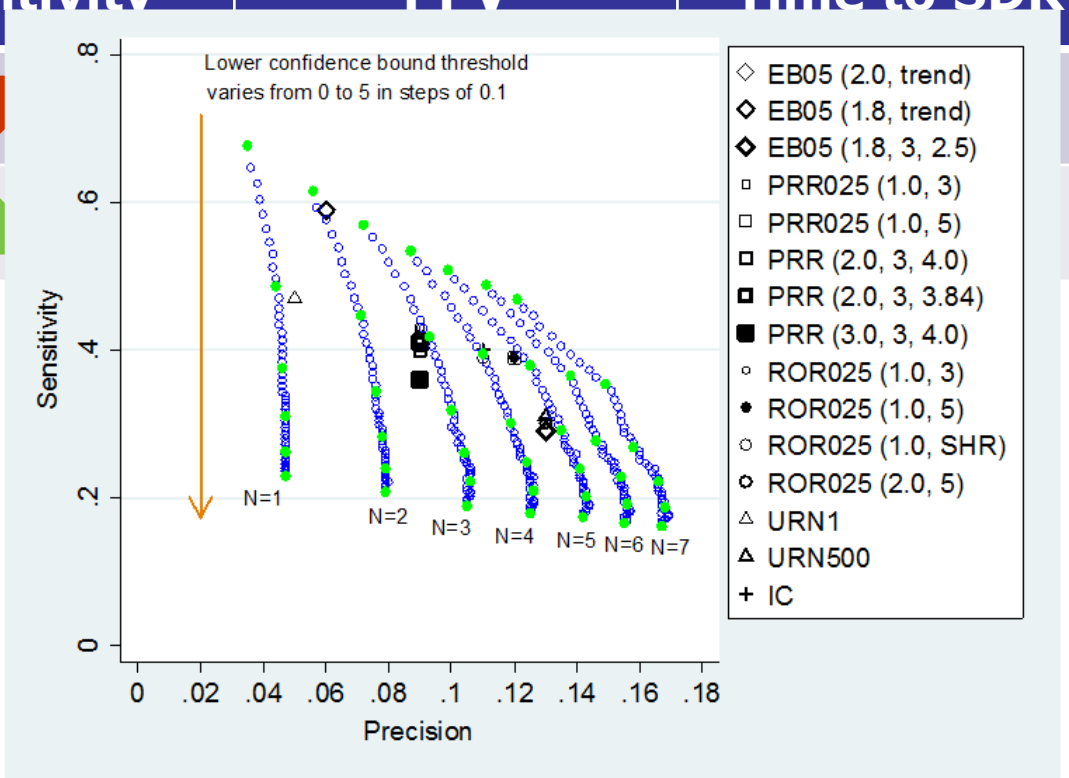
Thresholds: example in WP 3.1

Statistics	Partner / Current Use	Implementations
Proportional Reporting Ratio PRR	EMA	PRR lower bound 95% c.i. ≥ 1 & n ≥ 3
	EMA	PRR lower bound 95% c.i. ≥ 1 & n ≥ 5
	MHRA (No)	PRR ≥ 3 & $\chi 2$ ≥ 4 & n ≥ 3
	Bayer	PRR ≥ 2 & $\chi 2$ ≥ 4 & n ≥ 3
	Roche	PRR ≥ 2 & p ($\chi 2$) <= 0.05 & n ≥ 3
Reporting Odds Ratio ROR	UMC	ROR with shrinkage, lower bound 95% c.i. > 1
	MEB	ROR lower bound 95% c.i. > 2 & n ≥ 5
	None	ROR lower bound 95% c.i. ≥ 1 & n ≥ 3
	None	ROR lower bound 95% c.i. ≥ 1 & n ≥ 5
Information Component IC	UMC	IC lower bound 95% confidence interval (c.i.) > 0
Empirical Bayes Geometric Mean EBGM	MHRA	EB05 ≥ 1.8 & n ≥ 3 & EBGM ≥ 2.5
	AZ	EB05 ≥ 1.8 or positive trend flag
	GSK	EB05 > 2.0 or positive trend flag
URN model	None	Reporting Ratio > 1 & unexpectedeness > 1 / 0.05
	None	Reporting Ratio > 1 & unexpectedeness > 500 / 0.05

Thresholds

- Why important
  - Higher or lower thresholds significantly impact the outcome measures
  - Impact on the effectiveness of the detection process

Threshold	Sensitivity	PPV	Time to SDR
Higher	↓		
Lower	↑		



Thresholds

- How to choose them

  Empirical evaluation

  Resource availability

  - Available resources: any reduction in sensitivity may not be acceptable
  - Limited resources: need to prioritise

Number of SDR	PPV	ADR	Capacity (investigate SDR)	ADR detected
200	10%	20	100	10
100	15%	15	100	15

Thresholds might be applied differently to selected products

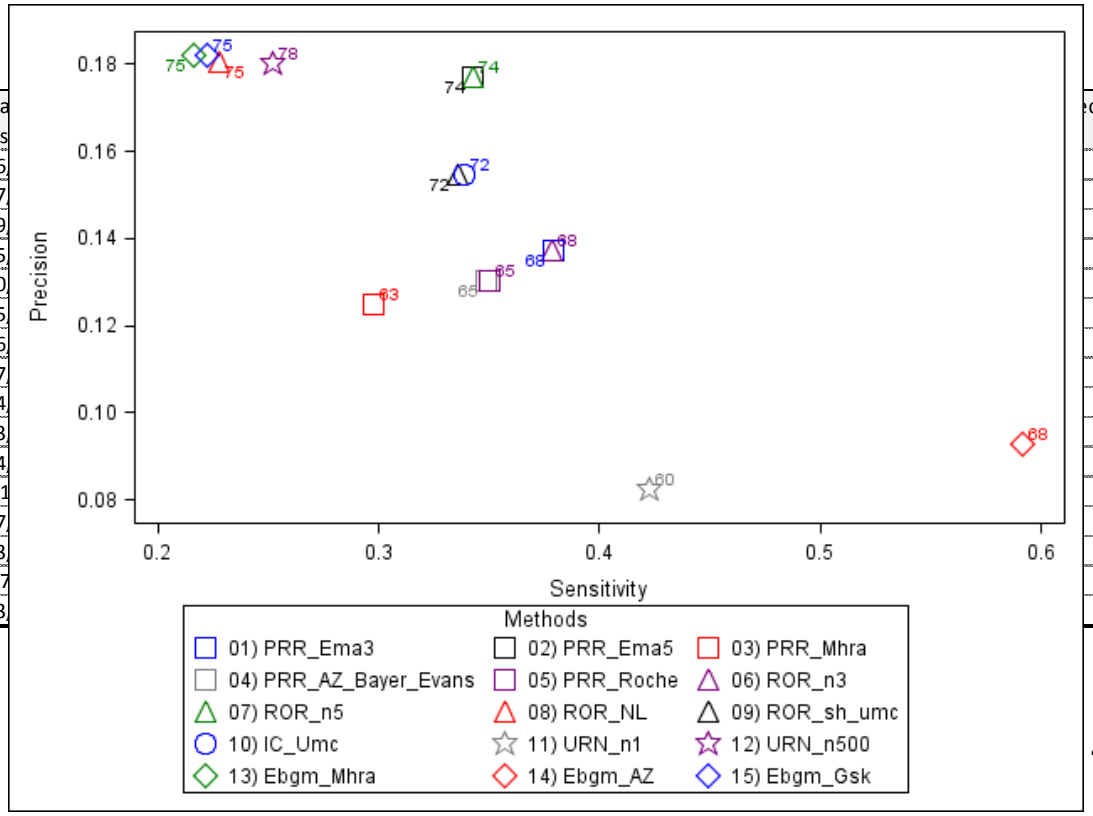
SAS Macros

- Objective

  Estimating the performance of the disproportionality methods tested in PROTECT WP 3.1 on your database

  - Estimating sensitivity, positive predictive value (precision) and time to SDR
  - Creating graphs

Methods	Total Labelled	Total Labelled with data	True Positive	Fa Pos	edian Time Drug
01) PRR_Ema3	23,742	17,054	9,000	56	57
02) PRR_Ema5	23,742	17,054	8,135	37	65
03) PRR_Mhra	23,742	17,054	7,057	49	52
04) PRR_AZ_Evans	23,742	17,054	8,272	55	55
05) PRR_Bayer	23,742	17,054	7,911	50	55
06) PRR_Roche	23,742	17,054	8,306	55	54
07) ROR_n3	23,742	17,054	8,988	56	57
08) ROR_n5	23,742	17,054	8,127	37	65
09) ROR_NL	23,742	17,054	5,394	24	66
10) ROR_sh_umc	23,742	17,054	7,968	43	62
11) IC_Umc	23,742	17,054	8,048	44	63
12) URN_n1	23,742	17,054	10,023	111	48
13) URN_n500	23,742	17,054	5,982	27	70
14) Ebgm_Mhra	23,742	17,054	5,131	23	66
15) Ebgm_AZ	23,742	17,054	14,052	137	59
16) Ebgm_Gsk	23,742	17,054	5,272	23	66



SAS Macros

- How
  - SAS macros available on PROTECT website:

    http://www.imi-protect.eu/methodsRep.shtml
  - More detailed instructions / set of slides might be available soon

- Disclaimer
  - The code makes most of the operations automatic, but manual input is required (e.g. input dataset, reference dataset,…)
  - Basic knowledge of SAS is strongly recommended
    - Different SAS versions, different system configurations, mistakes in preparing the inputs
  - If you encounter problems in its use we will do what we can to help
  - Acknowledge support of Ivan Zorych
  - The code assesses the disproportionality methods investigated in WP 3.1

Methodology

SAS Macros	Create Input	Calculate Statistics	Analysis SDR	Summary Final
Aim	Reconstruct dataset in the past Create datasets with monthly cumulative counts	Calculate statistics from disproportionality methods	Add reference DS information Remove all the records not related to the products in the analysis Flag SDR	Create summary tables and graphs

Create Input: introduction

- Objective

  Create the historical datasets with monthly cumulative counts

- Input

  Your raw dataset

  As if computing signal detection algorithm in your company

  - Raw dataset needs to **include all the reports** in the database for the report types (spontaneous, literature,…), drug characterisation (suspect, interacting,…), seriousness and/or non serious and medically confirmed and/or non medically confirmed **as used in the routine SD**
  - Raw dataset needs to have the same cleaning / grouping for the drugs / substances as used in the routine SD
  - Raw dataset needs to **include all the drugs** (not just the ones analysed in the reference dataset) as used in the routine SD

Create Input: example

- Example of raw dataset 'EMA'

Drug / Substance identifier

Reaction identifier

Case identifier

Date identifier for reconstructing the dataset in the past

Pk_SafetyReport	ScientificGroup_Id	Pt_Code	Receive Date
26043	20823	10041660	06/03/1992
26054	20823	10019668	06/03/1992
26055	20823	10019717	06/03/1992
26052	20823	10008635	01/01/1995
26049	20823	10003445	17/01/1995
26049	20823	10028813	17/01/1995

- Please note that the Drug / Substance identifier is the way the drug / substance is identified and reported **in your dataset**
  - It can be a numerical or a character value
- Drug identifier and reaction identifier need to be **harmonised** with the reference dataset

Create Input: output

- Output:

  One dataset per month

  – If the time window considered is from Jan 1995 to Dec 2011, 204 datasets are created (17 years * 12 months)

  Each dataset contains the count for all the reports **up to that month** for each drug / substance – event combination

- Example of output dataset 'Basic_Table_31Jan1995'

Date	ScientificGroup_Id	Pt_Code	a	b	c	d
31-Jan-95	20823	10000059	3	431	12	3,715
31-Jan-95	20823	10000081	5	429	89	3,638
31-Jan-95	20823	10000087	3	431	22	3,705
31-Jan-95	20823	10000097	1	433	3	3,724
31-Jan-95	20823	10000210	1	433	8	3,719
31-Jan-95	20823	10000378	1	433	27	3,700
31-Jan-95	20823	10000381	1	433	36	3,691
31-Jan-95	20823	10000389	1	433	1	3,726
31-Jan-95	20823	10000486	1	433	9	3,718

Calculate Statistics: introduction

- Objective

  Calculate the statistics from disproportionality methods

- Input

  Monthly datasets with cumulative counts for 'a', 'b', 'c' and 'd' for each drug / substance – event combination

- Example of input dataset 'Basic_Table_31Jan1995'

Date	ScientificGroup_Id	Pt_Code	a	b	c	d
31-Jan-95	20823	10000059	3	431	12	3,715
31-Jan-95	20823	10000081	5	429	89	3,638
31-Jan-95	20823	10000087	3	431	22	3,705
31-Jan-95	20823	10000097	1	433	3	3,724
31-Jan-95	20823	10000210	1	433	8	3,719
31-Jan-95	20823	10000378	1	433	27	3,700
31-Jan-95	20823	10000381	1	433	36	3,691
31-Jan-95	20823	10000389	1	433	1	3,726
31-Jan-95	20823	10000486	1	433	9	3,718

Calculate Statistics: output

- Output

  One dataset per month

  Each dataset contains for each drug / substance – event combination:

  - Count for all the reports up to that month
  - Statistics from signal detection algorithms

- Example of output dataset 'Sdr_31Jan1995'

Date	ScientificGroup_Id	Pt_Code	a	b	c	d	prr	prr_left_95	chi_sq	chi_yate	ic_025	unexp	Ebgm	Ebgm05	Ebgm01
31-Jan-95	20823	10000059	3	431	12	3,715	2.1469	0.6082	1.4758	0.6267	-1.2889	0.2006	1.3786	0.4925	0.2803
31-Jan-95	20823	10000081	5	429	89	3,638	0.4824	0.1971	2.6892	2.1585	-2.4332	0.9743	0.4780	0.2177	0.1450
31-Jan-95	20823	10000087	3	431	22	3,705	1.1710	0.3519	0.0663	0.0050	-1.8789	0.4922	0.9296	0.3320	0.1890
31-Jan-95	20823	10000097	1	433	3	3,724	2.8625	0.2984	0.9098	0.0184	-3.0876	0.3565	0.9894	0.1626	0.0517
31-Jan-95	20823	10000210	1	433	8	3,719	1.0734	0.1346	0.0045	0.2294	-3.7370	0.6293	0.6529	0.1073	0.0341
31-Jan-95	20823	10000378	1	433	27	3,700	0.3181	0.0433	1.4195	0.7766	-4.9865	0.9547	0.2848	0.0468	0.0149

Analysis SDR: introduction

- Objective

  Add the ADRs in the reference dataset for the products of the analysis (first part of the macro)

  Flag a SDR according to the different definitions used by the partners in WP 3.1 (second part of the macro)

- Input

  1) Monthly datasets containing the count up to that month and statistics from signal detection algorithms
     - Example of raw dataset 'Sdr_31Jan1995'

Date	ScientificGroup_Id	Pt_Code	a	b	c	d	prr	prr_left_95	chi_sq	chi_yate	ic_025	unexp	Ebgm	Ebgm05	Ebgm01
31-Jan-95	20823	10000059	3	431	12	3,715	2.1469	0.6082	1.4758	0.6267	-1.2889	0.2006	1.3786	0.4925	0.2803
31-Jan-95	20823	10000081	5	429	89	3,638	0.4824	0.1971	2.6892	2.1585	-2.4332	0.9743	0.4780	0.2177	0.1450
31-Jan-95	20823	10000087	3	431	22	3,705	1.1710	0.3519	0.0663	0.0050	-1.8789	0.4922	0.9296	0.3320	0.1890
31-Jan-95	20823	10000097	1	433	3	3,724	2.8625	0.2984	0.9098	0.0184	-3.0876	0.3565	0.9894	0.1626	0.0517
31-Jan-95	20823	10000210	1	433	8	3,719	1.0734	0.1346	0.0045	0.2294	-3.7370	0.6293	0.6529	0.1073	0.0341
31-Jan-95	20823	10000378	1	433	27	3,700	0.3181	0.0433	1.4195	0.7766	-4.9865	0.9547	0.2848	0.0468	0.0149

  2) Dataset containing the circulated list of products and ADR listed in the SPC
     - Example of raw dataset 'product_adr' for EMA

Scientificgroup_id	MedDRA Pt	Pt_code
Adefovir	Abdominal Pain	10000081
Adefovir	Renal Impairment	10062237
Adefovir	Asthenia	10003549
Adefovir	Carnitine Decreased	10007668

Analysis SDR: output

- Output

  One dataset

  The dataset contains only the products that are in the reference dataset used

- Example of output dataset 'Flagged_Sdr'

Count of all the reports up to that month (from 'Create Input')

Statistics from signal detection algorithm (from 'Calculate Sdr')

Flagged SDRs (from 'Analysis Sdr')

Whether the ADR is listed

Date	ScientificGroup_Id	Pt_Code	SPC	a	...	prr	...	prr_ema	prr_mhra	prr_az_bayer_evans	prr_roche	ic_umc	ebgm_mhra	ebgm_az	ebgm_gsk
31-Dec-04	23689	10012239	0	1		8.6		0	0	0	0	0	0	0	0
31-Dec-04	23689	10012373	0	1		3.8		0	0	0	0	0	0	0	0
31-Dec-04	23689	10012378	1	4		5.3		1	1	1	0	1	0	1	0
31-Dec-04	23689	10012503	0	1		60.0		0	0	0	0	0	0	0	1
31-Dec-04	23689	10012601	0	1		10.0		0	0	0	0	0	0	0	0
31-Dec-04	23689	10012735	1												

Reaction 10012735 is listed in the SPC but no reports have been received in the database up to 31-Dec-2004, therefore it has been added in the output dataset

Summary final: output

- Output

Dataset with summary performance

– Example of dataset 'A1_Summary'

Methods	Total Labelled	Total Labelled with data	True Positive	False Positive	Sensitivity	Sensitivity (data av)	Precision	Ratio FP / TP	Average Time Dec	Median Time Dec	Average Time Drug	Median Time Drug
01) PRR_Ema3	23,742	17,054	9,000	56,603	0.38	0.53	0.14	6.29	29	16	68	57
02) PRR_Ema5	23,742	17,054	8,135	37,833	0.34	0.48	0.18	4.65	37	25	74	65
03) PRR_Mhra	23,742	17,054	7,057	49,384	0.30	0.41	0.13	7.00	23	12	63	52
04) PRR_AZ_Evans	23,742	17,054	8,272	55,145	0.35	0.49	0.13	6.67	25	14	65	55
05) PRR_Bayer	23,742	17,054	7,911	50,742	0.33	0.46	0.14	6.41	27	15	66	55
06) PRR_Roche	23,742	17,054	8,306	55,517	0.35	0.49	0.13	6.68	25	14	65	54

Dataset with performance from time to authorisation

– Example of dataset 'A2_origin_sem'

Methods	Total Labelled	Total Labelled with data	True Positive	False Positive	Sensitivity	Sensitivity (data av)	Precision	Ratio FP / TP	Sem	Tot Lab with data Six Month incr	True Pos Six Month incr	False Pos Six Month incr	Sens (data av) Six Month incr	Precision Six Month incr	Ratio FP / TP Six Month incr
01) PRR_Ema3	23,742	704	57	32	0.00	0.08	0.640	0.561	1	704	57	32	0.081	0.640	0.561
01) PRR_Ema3	23,742	2,368	522	566	0.02	0.22	0.480	1.084	2	1,664	465	534	0.279	0.465	1.148
01) PRR_Ema3	23,742	3,728	1,003	1,443	0.04	0.27	0.410	1.439	3	1,360	481	877	0.354	0.354	1.823
01) PRR_Ema3	23,742	4,979	1,501	2,526	0.06	0.30	0.373	1.683	4	1,251	498	1,083	0.398	0.315	2.175

More datasets are created and more complex versions of the macros exit: please do contact us directly if interested

Computing Time

- Bayer Dataset (thanks to Katrin)
  - Bayer raw data: 1,522,493 observations, 1.93 GB
  - Time window: January 1995 – December 2011
- 'Create Input'
  - Computing time: 5 hours 39 minutes
  - 204 datasets, 11.5 GB
- 'Calculate Statistics'
  - Computing time: 9 hours 53 minutes
  - 204 datasets, 12.4 GB
- 'Analysis SDR'
  - Computing time: 3 minutes

More information?

Website: www.imi-protect.eu

Email: Protect_Support@ema.europa.eu