IMI1 Final Project Report
Public Summary

Project Acronym: PROTECT

Project Title:
Pharmacoepidemiolocal Research on Outcomes of Therapeutics by a European ConsorTium

Grant Agreement: 115004
Project Duration: 01/09/2009 - 30/04/2015

Innovative Medicines Initiative

TEMPLATE

IMI PROJECT FINAL REPORT
INCLUDING THE PERIODIC REPORT FOR THE LAST PERIOD[1]

Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium

PROTECT

Grant Agreement No 115004

[Name of the scientific representative of the Coordinator][2]

Xavier Kurz, MD, DTM&H, MSc, PhD
European Medicines Agency
+44 20 3660 7269
xavier.kurz@ema.europa.eu

Last Period: **1st October 2013 – 30 April 2015**

Reporting Period **5**

Duration of the project: 1 September 2009 – 30 April 2015

Description of work – Version 1.0; 21/04/2015

April 2017

[1] See Articles II. 4.1 and II 4.2 of the IMI model Grant Agreement.
[2] Usually the person mentioned in the coordinator A2.4 form (in SOFIA IT tool).

Declaration of the coordinator

I, the coordinator of this project, declare that,

The final report submitted is in line with the obligations as stated in Article II.2.3 of the Grant Agreement:

The attached report represents an accurate description of the work carried out in this project for the last reporting period as well as for the whole duration of the project;

For the last period, the project *(tick as appropriate):*

- ☐ has fully achieved its objectives and technical goals; has achieved most of its objectives and technical goals for the period with relatively minor deviations[3];
- ☐ has failed to achieve critical objectives and/or is not at all on schedule[3].

For the whole duration of the project, the project *(tick as appropriate):*

- ☐ has fully achieved its objectives and technical goals;
- ☐ has achieved most of its objectives and technical goals with relatively minor deviations[3];
- ☐ has failed to achieve critical objectives and/or is not at all on schedule[3].

The public project website http://www.imi-protect.eu [4] is up to date.

To my best knowledge, the financial statements which are being submitted as part of this final report are in line with the actual work carried out and are consistent with the report on the resources used for the project (section 7) and if applicable with the certificate on financial statement.

All participants, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes or deviations have been reported under section 6 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of the Coordinator: Xavier Kurz……

Date: ...../ ...../ .....

Signature of the Coordinator: ……..

[3] If either of these boxes is ticked, the report should reflect these and any remedial actions taken.

[4] Please add the address of the public project website. The home page of the website should contain the generic IMI logo which is available in electronic format at the IMI website. The area of activity of the project should also be mentioned.

1. Executive summary

1.1. Project rationale and overall objectives of the project

PROTECT is a research programme that was designed based on an in-depth analysis of knowledge gaps and data needs that represent barriers to product development and continuous benefit- risk monitoring of medicinal products during their life cycle. PROTECT aims to address limitations of current methods used in pharmacovigilance and pharmacoepidemiology and to significantly strengthen the monitoring of benefit-risk (B-R) of medicines marketed in Europe.

The objectives of PROTECT are: to enhance data collection directly from consumers in their native language in several countries using modern tools of communication; to improve early signal detection from spontaneous reports, electronic health records and clinical trials; to develop and disseminate methodological standards for the design, conduct and analysis of pharmacoepidemiological studies applicable to different safety issues and different data sources; to develop methods for continuous benefit-risk monitoring of medicines, by integrating and presenting data on benefits and risks from clinical trials, observational studies and spontaneous reports; and to validate various methods developed in PROTECT using different data sources in order to identify and help resolve operational difficulties linked to multi-site investigations.

1.2. Overall deliverables of the project

PROTECT has been designed as a comprehensive and integrated project aiming to develop and validate a set of innovative tools and methods. The objectives of PROTECT and the main deliverables generated for each of them are as follows.

Objective 1. To enhance data collection directly from consumers of medicines in their natural language in several EU countries, using modern tools of communication.

Overall deliverables:

- Report on length of time and consistency with which pregnant women recruited via internet may provide data on drug utilisation and pregnancy outcomes.
- Report on usage, accuracy and completeness of self-reported prescription drug use by pregnant women, including on over-the-counter products, homeopathic and herbal medicines.
- Report on validity, completeness and usefulness for pharmacovigilance of data collected on drug utilisation and self-reported pregnancy outcomes (incl. representativeness of women and stage of pregnancy status at recruitment).
- Report on evaluation and comparison of methods for collecting data directly from pregnant women, and comparison with data from electronic health records.

Objective 2. To improve early and proactive signal detection (SD) from spontaneous reports, electronic health records and clinical trials.

Overall deliverables:

- Report on use of standard MedDRA groupings to expedite the detection of disproportional reporting for historical safety signals, showing no overall benefit in

conducting signal detection using MedDRA HLTs or SMQs compared with using individual PTs.

- Creation of OntoADR, an ontology for MedDRA preferred terms to facilitate dynamic definition of groups based on the relevant dimensions for a specific topic of interest.
- Structured database of the ADR information in section 4.8 of the Summary of Product Characteristics (SPC) for all European centrally authorised products, in MedDRA.
- Report on the assessment of methods to address challenges with single drug- single event associations such as masking, effect modification and confounding and to explore more complex patterns such as suspected drug-drug interactions.
- Report on the evaluation of the performance of common statistical signal detection algorithms based on a study across spontaneous report databases from pharmaceutical companies, national and international pharmacovigilance organisations.
- Report of a study exploring the use of stratification and subgroup analysis in disproportionality analysis for spontaneous report databases.
- Report on the impact of masking on statistical signal detection in pharmacovigilance.
- Report showing that statistical interaction measures with additive baseline models outperformed those with multiplicative baseline models typically available in standard statistical software.
- Report showing that probabilistic record matching for duplicate case detection performed better than rule based screening and should be considered as an alternative to such methods.
- Report on the comparison between estimates of association from formal epidemiological studies and proportional reporting ratios (PRRs) in spontaneous reporting data for a set of known ADRs, with the finding of a correlation, at a point in time before the ADR was first publicly recognised.
- Process for structured clinical and epidemiological assessment of temporally associated prescriptions and events in electronic health records with application in The Health Improvement Network (THIN) database of longitudinal electronic health records from general practices in the UK.
- Report on signal detection performance in clinical trials based on extreme value modeling as a basis to predict drug toxicity.

Objective 3. To develop, test and disseminate methodological standards for the design, conduct and analysis of pharmacoepidemiological (PE) studies applicable to different safety issues and using different data sources.

This work was organised in three working groups:

Databases, with the objective to explain differences in drug-adverse event associations due to choices in methodology and databases by comparing results from pharmacoepidemiological studies;

Confounding, with the objective was to evaluate and improve methods to control confounding by testing them in simulation studies and applying them in real-life data sets.

Drug utilisation (DU), with the objective to provide guidelines on how to identify, assess validity and use national drug utilisation data to estimate public health impact of adverse events.

Overall deliverables

- Report/publications on key adverse events and feasibility to study these events in electronic health/epidemiological databases.
- Study protocols for each of the 5 key selected adverse events and related drugs for each of the databases in the Consortium.
- Report/publication with an inventory of data sources (including three yearly updates) on the consumption of the medicines of interest in the EU (including validity of data).
- Publications on evaluation and application of methods to control for confounding.
- Report/publication with examples of estimation of the exposed population related to the drug-AE pairs of interest, and patterns of use of the products of interest.
- Report/publications on analysis of discrepancies between studies using common protocols and definitions in several data sources.
- Report/publications on statistical methods to analyse multi-database studies.
- Guidelines and methodological standards to scientific community and industry for conceptualization of PE studies.

Objective 4. To develop methods for continuous B-R monitoring of medicines, by integrating data on benefits and risks from clinical trials, observational studies and spontaneous reports, including both the underpinning modelling and the presentation of the results, with a particular emphasis on graphical methods.

Overall deliverables:

- Report with extensive review and evaluation of the methods used in benefit risk assessment.
- Reports on results of the test of key methods via a case study approach (wave 1 case studies).
- Report with extensive review and evaluation of the graphical/visual representations that could be used in presenting benefit-risk information.
- Reports on complex case studies to further develop B-R methodologies and explore visual representation/recommendations (wave 2 case studies).
- Report on benefit-risk assessment according to different perspectives, including those of regulators, prescribers, public and patients.
- Report on criteria to be used for an extended case study.
- Synthesis of findings into final umbrella recommendations.
- Report on the Patient and Public Involvement study.
- Training materials on benefit-risk evaluation.

- Website for the dissemination of project findings to the public (http://www.PROTECTBenefitRisk.eu).

Objective 5. To test and validate various methods developed in PROTECT using different methods and different methods. This work has been performed in relation to the objectives 3 and 4 above.

Overall deliverables

In relation to objective 3 (methodological standards for pharmacoepidemiological studies), reports on:

- Replication of study in a same database (Is a study replicable when conducted independantly ?)
- Replication of a study in different databases (Do the results have external validity ?)
- Negative control study (Does a study provide absence of evidence of an association where the exposure is such that the expected result is one of no association ?)
- Use of alternative outcome definition (What is the impact of different levels of certainty of the outcome on the effect estimate ?)
- Validation of outcome (What is the impact of validation, e.g. through clinicl record review, on the effect estimate ?)
- Assessment of confounders (How does better control for confounding impact on the effect estimate ?).

In relation to objective 4 (methods for benefit-risk evaluation), reports on:

- Results of case studies testing and extending recommendations on B/R assessment methods in real-life setting.
- Results of a study performed in patients, health care professionals and regulatory assessors to test the applicability and acceptability of different visual methods on B/R evaluation for patients diagnosed with atrial fibrillation, diabetes and breast cancer.
- Testing and comparison of methods to collect data on patients' preference.

Objective 6: To identify tools for disseminating Project results, including training programme uptake of methodologies by interested stakeholders and identify training opportunities that may be offered to Consortium Partners and to students of the IMI pharmacovigilance Training Consortium.

Overall deliverables

- Training modules on PROTECT results in collaboration with the IMI Call No 18 Consortium;
- Platform of training opportunities listing and promoting training opportunities;
- List of conferences, workshops, symposia and other forums appropriate for a presentation of results achieved by the PROTECT Consortium;
- Communication plan for the dissemination of PROTECT results.

1.3. Summary of progress versus plan since last period

Objective 1 (Data collection, Work package 4)

In the original plan, the intention was to stop recruitment in August 2013; however, as detailed in Report #4, due to the delays encountered during the set-up of the project (see tasks 4.1.3, 4.1.4), the enrolment was extended until the end of January 2014 (task 4.6.3.2) and data collection until 31st March 2014 task (4.6.4.2). All related tasks for 4.6 were adequately met during this reporting period and 2,521 participants were successfully enrolled into the study, with 2066 providing at least baseline data. For the 1555 women expected to deliver their babies during the time while the study was still active, information was received for 464 pregnancy outcomes.

Another major achievement during this reporting period was the delivery of the final data set in April 2014 (Milestone task 4.10). This enabled to meet two report deliverables; (1) Report on results of data collected directly from pregnant women including comparative evaluation with data from other databases (Task 4.12) and (2) Report on transferability of methodology to other target populations and pharmacovigilance situations (Task 4.16) via the development of the WP-4 study report and the generation of a number of manuscripts. We also met our final outstanding deliverable of a report describing 1) the user requirements and formats for consumer-based tools, 2) the assessment of the efficiency, usefulness of and satisfaction with these tools, and 3) recommendations on future development to facilitate the collection of drug utilisation (Task 4.15). The task was completed via the conduct of the study and also by performing qualitative interviews on pregnant woman, as recommended by the External Advisory Board. All of this work can be reviewed in the WP-4 study report and also via the publication of a manuscript 'Balancing the Interests of Patient Data Protection and Safety: An EU-based case study of a multi-stakeholder, direct-to-patient study of medication use and life style factors in pregnancy' in the Journal of Medical Internet Research (JMIR) Medical Informatics, with a publication dated April 2015. Other manuscripts are in progress and several presentations at major conferences have been accepted.

In summary, all required activities for the last period were delivered on time, recognizing the adjustment of the consortium-agreed extension of study closure from August 2014 to April 2015 and with no major risk or deviation.

Objective 2 (Signal detection, Work package 3)

Work Package 3 has completed its planned activities and worked on the communication of its deliverables through publications and presentations.

The 3.02 sub-package *Risk estimates compared with disproportionality statistics* has completed its analyses, but a delay in the completion of this milestone led to a delay in the external communication, which has not been finalised but is under way.

The 3.08 sub-package *Subgroups and risk factors* has completed its analyses and presented its findings at the 30th International Conference on Pharmacoepidemiology and Therapeutic Risk Management in Taipei, Taiwan, Oct 2014, the PROTECT Symposium in London, United Kingdom, Feb 2015, and at the Drug Information Association Annual EuroMeeting in Paris, France, April

2015; beyond that, a full scientific paper has been drafted and is currently undergoing local review by partners before submission to a journal.

The 3.09 sub-package *Signal detection in clinical trials* has completed its analyses, which pursued two distinct approaches to signal detection in clinical trials. The evaluation of methods based on extreme value modeling has already been published as a scientific paper, whereas the evaluation of hierarchical Bayes models has been presented at international scientific conferences; the outstanding dissemination activity is the publication of the hierarchical Bayes evaluation as a full scientific paper.

Each of the above Signal Detection sub-packages have completed their analyses and contributed to the PROTECT recommendations for Good Signal Detection Practices. The overall objectives of the project have thus been met, but will be further enhanced by the completion of the three outstanding dissemination activities.

The PROTECT recommendations are being used by EMA in its revision of the Guidance on Statistical Signal Detection from EudraVigilance.

Objective 3 (Pharmacoepidemiology, Work package 2)

WP2 has completed all planed milestones and deliverables in terms of data analyses, except one (see below). However, writing up of results from WG1/WG2 into papers and final recommendations to be included in the PDS supplement is still in progress and will be delivered beyond the end date of the project.

Milestone 2.5.7 (Apply PERR adjustment for AED pairs in Dutch databases) was not completed. Based on results from simulation study (Milestone 2.5.6, described in Uddin et al. Performance of prior event rate ratio adjustment method in pharmacoepidemiology: a simulation study. Pharmacoepidemiol Drug Saf. 2014 Nov 20. doi: 10.1002/pds.3724) the application of PERR adjustment in the AED pair B2A-AMI was deemed scientifically inappropriate. The simulations showed that violation of one of the assumptions underlying the PERR adjustment method (i.e., that prior outcomes should not affect future exposures) is very influential. Since this assumption is likely to be violated in studies of B2A-AMI, this task was not completed.

Delays with the Danish National Registries results were reported to PROTECT coordinators as red risks in previous periodic reports. DKMA did not deliver the descriptive results of suicidality outcome or the cohort results of AED/Suicidality and CCB/Cancer pairs. For AED/Suicidality, comparison of results with a replication study in CPRD from WP6 was possible. However, for CCB/Cancer, no studies were performed in WP6 therefore only one study in a single database (CPRD) was delivered for this AE-drug pair. All other risks reported previously were solved during the last period.

Objective 4 (Benefit-risk assessment, Work package 5)

WP5 has continued to deliver during year 5.

The final set of umbrella recommendations from the four year research has been published on the PROTECT website in November 2013, with a generic roadmap to the key points and selection of methodologies.

Patient/public involvement continues, the planned work are mainly completed with the exception of an ongoing new survey to people living with multiple sclerosis (MS).

Re-development of the benefit-risk website is complete but updates are being made on regular basis with news and related contents. It has been re-launched with a new URL (http://www.protectbenefitrisk.eu/) to make it more accessible to the public. The website is stored at the Imperial College Webfarm server for future sustainability. Through the website, all the findings from PROTECT have been made available through a user-friendly interactive interface, including the recommendations, methodologies, visualisation, case studies and patient/public involvement. Resources for training and links to contact experts who worked in the project are also available.

Four peer-reviewed publications were published in 2014; a number of manuscripts have been submitted and several others are under preparation. Multiple presentations on the findings from WP5 had taken place, and several more are planned up to September 2015. WP5 continues to hold face to face meetings for the whole work package on an approximately quarterly basis and Management Team meetings on a monthly basis up to February 2015.

Objective 5 (Replication studies, Work package 6)

The overall scope of the replication studies performed by WP6 in relation to WP2 has not been changed since the 4th report. Since the last report, the major achievements were mainly related to:

- confronting results with WP2 studies results
- writing common manuscripts for some of the drug/adverse events pairs in the one hand or standalone manuscripts for the negative control studies and other drug/adverse events pairs in the other hand
- the process of co-authors reviewing, submitting to journals and answering journal reviewers comments.

Results were compiled in papers prepared by WP6 members alone or together with WP2 members.and submitted to the Pharmacoepidemiology and Drug Safety and presented at the 30th International Conference on Pharmacoepidemiology and Therapeutic Risk Management in Taipei, Taiwan, Oct 2014.

The following studies had to be cancelled:

- antidepressants and benzodiazepines prescription and the risk fracture in the CPRD; self control case series; reason: this study had to be performed by a private partner (GSK) which withdrew from WP6 to give priority to other studies performed by the same investigator for WP2 (already stated in 4th report);
- calcium channel blockers and the risk of cancer in the E3N database (case-control study); reason (already stated in the 4th report): feasibility of access to this database was not considered adequate;
- calcium channel blockers and the risk of cancer in Marketscan (cohort study); reason: this study had to be performed by a private partner (GSK) which withdrew from WP6 to give

priority to other studies performed by the same investigator for WP2 (already stated in 4th report);

- antiepileptics & suicidality in CPRD (outcome validation study); reason: this study aimed to match cases of suicidality identified in CPRD with results of a previous study done by GSK to validate cases of suicidality based on other sources of data; however, given the different periods covered by the two studies, there were not enough cases to allow meaningful analyses (already stated in 4th report).

The studies performed by WP6 in relation to WP5 included two groups

1. In WP5/workstream 6-1, this activity consisted in testing and extending the WP5 recommendations on Benefit/Risk assessment methods in real-life settings, accounting for data heterogeneity and scarcity, using primary sources of preference elicitation (panel experts, patients) and factoring in the time factor in the analysis (time horizons of outcomes, repeated assessment over time as new data get available).
   Two case studies, derived from WP5 examples, were planned:
   - Efalizumab, for which the data analysis was redone and refined so that data heterogeneity and difference in time horizon could be addressed. The three assessment times (2004,2008 and 2009) as per the real case were evaluated using a Panel experts mimicking a regulatory bodies and involving regulators and medical experts was used for MCDA-based weights elicitation
   - Rimonabant, for which it was planned to address data heterogeneity and use primary data from obese patients collected in the VISUALIzE study. It was hoped that at least about 100 patients could be analysed, but as of February 2015, only 17 patients had been recruited. This recruitment rate did not allow to conduct the analysis within the PROTECT project timelines.

2. WP5/6-2: This workstream included three activities:
   a. The Visualizing Uncertainty Among Laypersons and Experts (VISUALizE) study (D.6.11); it aimed to assess the applicability and acceptability of visual methods developed by WP5 among patients diagnosed with atrial fibrillation, diabetes and breast cancer. Activities included the design, programming and launch of an online research website displaying 18 questionnaires (3 disease areas, 3 languages for patient and healthcare experts). Collaboration of 52 clinical organizations in the UK and the Netherlands to recruit and administer the questionnaire was obtained. Its achievements are described in section 1.4.

   b. Implementation of B/R into regulatory practice.
   This activity was initiated to facilitate the further implementation of the results from PROTECT in regulatory practice. Initially, it was foreseen that this could be achieved through interaction with the CHMP and the PRAC and through the organization of training sessions at individual National competent authorities. However, while our initiative to facilitate the uptake of the methodology developed in PROTECT was supported by the EMA, it was also recognized that there was clear overlap with the activities conducted as part of EMA's B/R methodology project. To create synergy between these two related initiatives,

the UMCG and EMA agreed to second Douwe Postmus as a National Expert to the Agency.

At the time this extension activity was launched, the CHMP had just completed their second pilot study on the use of the Effects Table, but a decision regarding the implementation of this table still had to be made. As the Effects Table forms the basis for the quantitative methods described in the training material and implemented in the ADDIS software, it was not possible to start interacting with the PRAC before the Effects Table had been officially adopted by the CHMP. As the Effects Table has only become effective since February 2015, it was not possible to complete this activity within the timelines set for the PROTECT extension activities. Interaction with the PRAC on the updating of the B/R balance in response to new safety signals is currently ongoing within the context of the PRAC-BR steering group.

This activity also included provision of training but the EMA preferred to adopt a "train-the-trainer" approach where assessor training at the NCAs will be provided by the countries' senior assessors and CHMP members rather than EMA staff or external stakeholders. To kick-off this activity, the EMA and CHMP agreed to have a workshop dedicated to discussion and brainstorming, followed by a phase where the training tools will be developed based on the agreed principles during the workshop. During this workshop, which took place in January 2015, the B/R section of the day-80 template was discussed and improved. Currently, the day-80 template is revised based on the input from the workshop. The set of training tools that will subsequently be developed may incorporate some elements of the training material developed in the context of ROTECT, such as the discussion regarding the issue of double counting in the selection of the most important favourable and unfavourable effects.

c. An additional activity performed by WP5/6-2 is a new public release of the MCDA web interface of the ADDIS software (D.6.13). This tool is freely accessible at https://mcda.drugis.org and contains the regorafenib case study from the training material as an example problem.

Objective 6 (Training and communication, Work package 7)

The operational process for the identification of **training programme deliverables** and their transfer to the EU2P was established in Year 3. It consisted in identifying potential training topics from the PROTECT publication tracking list, and presenting each topic to the corresponding EU2P domain or training module coordinator. This procedure has been continued during the last period. It should be emphasised, however, that several PROTECT partners were themselves members of the EU2P project and involved in the development or maintenance of a specific training module.

The **Platform of Training Opportunities** was launched in December 2010. During the whole period, there have been between 9 and 12 positions offered. Despite initial interest by many members of the Consortium, only two institutions offered training positions (FICF and Mario Negri). Until 24 April 2015, the Platform had received 13,859 visits by 2,693 users; the main

topics of interest were drug utilization research (around 900 visits), training in clinical pharmacology (around 400 visits), and selection of the epidemiological strategy for specific drug safety issues (around 200 visits). Seventy-nine applications were submitted by 57 potential candidates, of which 12 were EU2P students. The backgrounds were Pharmacy (34), Medicine (5), Biosciences or Biology (11), and other (7). The countries of origin were UK (5), Germany (6), Italy (3), Spain (14), The Netherlands (4), France (2), Sweden (2), and others from Afghanistan, Argentina, Australia, Brazil, Colombia, India, Nepal, and Switzerland. The main fields of interest were pharmacovigilance and case-population research (20), drug utilisation (10), and "collaboration with an ongoing study" (14).

1.4. Significant achievements since last report

Objective 1 (Data collection, Work package 4)

Since the last report, WP4 completed study enrollment (2,521 women) and data collection (of the 2,521 women, 2066 provided any data). It performed the cleaning of data, the preparation of the final data set, the statistical and interpretation of the data and the completion of the WP4 Study report.
In addition to the work originally planned, it performed Qualitative Interviews on pregnant woman following advice from the External Advisory board.
In terms of dissemination activities, two posters were presented at at the 30th International Conference on Pharmacoepidemiology and Therapeutic Risk Management in Taipei, Taiwan, Oct 2014:

- Seeking patient-reported Information on Medication Use, without health care professional intervention: PROTECT Pregnancy Study Results.
- Internet advertisement methods provide the highest levels of recruitment to a pilot study of self-reported medication use and pregnancy outcome.

Other presentations included those at the EUROmediCAT Conference in Poznan, Poland (Best poster award), at the Final PROTECT Symposium (EMA) and in a webinar for the US Agency for Healthcare Research and Quality's Registry of Patient Registries program. A manuscript was published in the medical journal JMIR Medical Informatics

Objective 2 (Signal detection, Work package 3)

Sub-package 3.02 *Risk estimates compared with disproportionality statistics* has completed its analyses, but a delay in the completion of this milestone led to a delay in the external communication, which has not been finalised but is under way. The sub-package team continues to work and is now in the process of writing up the study for publication as a scientific paper.

Sub-package 3.08 *Subgroups and risk factors* has completed its analyses and presented its findings at the 30th International Conference on Pharmacoepidemiology and Therapeutic Risk Management in Taipei, Taiwan, Oct 2014, the PROTECT Symposium in London, United Kingdom, Feb 2015, and at the Drug Information Association Annual EuroMeeting in Paris, France, April 2015; beyond that, a full scientific paper has been drafted and is currently undergoing local review by partners before submission to a journal.

Sub-package 3.09 *Signal detection in clinical trials* has completed its analyses, which pursued two distinct approaches to signal detection in clinical trials. The evaluation of methods based on extreme value modeling has already been published as a scientific paper, whereas the evaluation of hierarchical Bayes models has been presented at international scientific conferences; the outstanding dissemination activity is the publication of the hierarchical Bayes evaluation as a full scientific paper.

Each of the above Signal Detection sub-packages have completed their analyses and contributed to the PROTECT recommendations for Good Signal Detection Practices. The overall objectives of the project have thus been met, but will be further enhanced by the completion of the three outstanding dissemination activities.

Objective 3 (Pharmacoepidemiology, Work package 2)

Throughout the project, WP2 was consistently present at international Pharmacoepidemiological congresses and meetings. Notably, WP2 was widely represented in the past two annual International Conference on Pharmacoepidemiology (ICPE), sponsored by the International Society for Pharmacoepidemiology (ISPE) in Montreal (2013) and Tapei (2014), and midyear ISPE meetings in Munich (2013), Rotterdam (2014), and Bordeaux (2015). In both annual congresses combined, WP2 presented 25 posters, 12 oral presentations on various studies on methods to control for confounding and led two symposiums, i.e. "Improving Consistency in Findings from Pharmacoepidemiological Studies" and "Impact of methodological choices on findings from pharmacoepidemiological studies: final results of the IMI-PROTECT" where results from WP2 and WP6 studies were presented jointly and comparisons between results were discussed. At the midyear meeting in Munich a symposium on PROTECT WP2, WP5 and WP6 was organized by WP2 co-leads.

During the last period, the analysis of all results showed that developing a common protocol for PE studies with great detail shall reduce methodological differences and interpretation by researchers. This requires a solid infrastructure for communication between sites conducting the same study. Conducting analysis in parallel in multiple databases instead of pooling of databases shall show heterogeneity and help exploring its sources. The large number of PE studies conducted in WP2 was the result of a close collaboration between 8 public and 9 private partners from the PROTECT consortium during 5 years.

WP2 widely disseminated its results on the extensive review of methods to control for confounding, focusing on instrumental variable and propensity score analyses. The guidance documents on IV analysis "Instrumental variable analysis in randomized trials with non-compliance and observational pharmacoepidemiologic studies" and the "Practical guidance for applying PS in PE studies" will contribute to the improvement of the use of those methodologies.

WP2 produced the inventory of "Drug consumption databases in Europe", publicly available in the IMI PROTECT website. It is a comprehensive and structured source of information on drug consumption in Europe. It comprises two documents. The master document is a detailed report of the available information, methods to retrieve it, a description of the validity of national drug consumption data and a discussion section. The country profile document summarizes the main results by country. Information is provided for 35 countries i.e. Armenia, Austria, Bosnia and Herzegovina, Belgium, Bulgaria, Croatia, The Czech Republic, Denmark,

Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and United Kingdom. This inventory has been acknowledged by the PROTECT consortium as a very useful tool to the different stakeholders involved in pharmacovigilance and drug safety.

Objective 4 (Benefit-risk assessment, Work package 5)

WP5 has continued to deliver during the 5th reporting period.

The final set of umbrella recommendations from the four year research has been published on the PROTECT website in November 2013, with a generic roadmap to the key points and selection of methodologies.

Patient/public involvement continues, the planned work are mainly completed with the exception of an ongoing new survey to people living with multiple sclerosis (MS).

Re-development of the benefit-risk website is complete but updates are being made on regular basis with news and related contents. It has been re-launched with a new URL (http://www.protectbenefitrisk.eu/) to make it more accessible to the public. The website is stored at the Imperial College Webfarm server for future sustainability. Through the website, all the findings from PROTECT have been made available through a user-friendly interactive interface, including the recommendations, methodologies, visualisation, case studies and patient/public involvement. Resources for training and links to contact experts who worked in the project are also available.

Four peer-reviewed publications were published in 2014; a number of manuscripts have been submitted and several others are under preparation. Multiple presentations on the findings from WP5 had taken place, and several more are planned up to September 2015. WP5 continues to hold face to face meetings for the whole work package on an approximately quarterly basis and Management Team meetings on a monthly basis up to February 2015.

Objective 5 (Replication studies, Work package 6)

In relation to the replicability of the results obtained by WP2, the major achievements since the last report, were mainly related to

- confronting results with WP2 studies results
- writing common manuscripts with WP2 for some of the drug/adverse events pairs on one hand or stand-alone manuscripts for the negative control studies and other drug/adverse events pairs on the other hand;
- the process of co-authors reviewing, submitting to journals and answering journal reviewers comments.

In relation to the reproducibility of WP5 results, achievements included:

- For WP5/6-1: testing and extension of the WP5 recommendations on Benefit/Risk assessment methods in real-life settings, accounting for data heterogeneity and scarcity, using primary sources of preference elicitation (panel experts, patients) and factoring in the time factor in the analysis (time horizons of outcomes, repeated assessment over time as new data get available); the case studies derived from WP5 example addressed data

heterogeneity and the difference in time horizon specifically for efalizumab (https://eroombayer.de/eRoom/PH-GDC-PI-SID/IMI-PROTECT/0_106909).

- WP5/6-2: In the VISUALizE study, the collaboration of 52 clinical organizations in the UK and the Netherlands to recruit and administer the questionnaire is a major achievement. This large European based study will provide robust data addressing the study objectives of assessing the 'best' graphical tools to communicate benefits and risks of medicines to patients and healthcare experts. Version 7 of the report is available on the PROTECT eRoom (https://eroombayer.de/eRoom/PH-GDC-PI-SID/IMI-PROTECT/0_1068fa).
- Another achievement is the new public release of the MCDA web interface of the ADDIS software. This tool is freely accessible at https://mcda.drugis.org and contains the regorafenib case study from the training material as an example problem.

Objective 6 (Training and Communication, Work package 7)

Activities of Work Package 7 have been described in section 1.3. and consisted of the maintenance of the Platform of training opportunities.

Following a decision of the Steering Committee, activities related to the maintenance of lists of reports, manuscripts and publications were performed by WP1 (co-lead by EMA and GSK). This transfer was a logical step as EMA was maintaining the PROTECT website were reports, presentations and publications are made publicly available.

1.5. Scientific and technical results/foregrounds of the project

Objective 1 (Data collection, Work package 4)

The pilot study in pregnant women showed that it was possible to obtain useful information about medications used during pregnancy from self-reported data. Specifically, the following results were found:

1. Demographic characteristics and health status of pregnant women at study entry

   The internet-based recruitment method generally attracted well-educated women, and nearly all respondents were Caucasian. PROTECT women were slightly older than the national averages for pregnant women.

2. The length of time and consistency with which pregnant women recruited via internet would provide the data requested.

   The methods used in this pilot study were not sufficient to promote long retention. We experienced substantial drop out of women after recruitment, either because they lost interest and more likely because the questionnaires were rather lengthy. Less than one third of the women, who were eligible to deliver during the time while the study was actively collecting data, actually provided information about pregnancy outcome. It appears that additional measures need to be taken to promote frequency and consistency of reporting.

3. The usage, accuracy and completeness of self-reported prescription drug use, based on comparisons of self-reported data with data from external sources (pharmacy data bases and electronic health records) in countries where such resources exist, or with national data.

   The accuracy and completeness of self-reported prescription use varied by the number of medications a woman was believed to be taking, based on external data. Women who were

taking a single medication for a chronic disease generally were able to report that information accurately but the quality of the reporting decreased in proportion to the number of medications taken. Women who were recruited earlier in pregnancy reported more accurately than those who were recruited later in pregnancy, possibly due to the fact the women recruited earlier provided more follow-up data than those who responded later. Many women found it difficult to identify and report their medication use by indication for use and instead chose to provide free-text data about medications when asked if there were any other medications they would like to report. Finally, data collection did not distinguish between prescription and non-prescription data, which vary by country, and made comparison of self-reported data with external data difficult.

4. The usage of over-the-counter products, as well as homeopathic and herbal medication use in pregnancy.

   Notwithstanding the difficulty of distinguishing between prescription medications and medications which could be obtained without prescription (see above), many women reported using non-prescription and herbal medications during pregnancy.

5. The validity of self-reported pregnancy outcomes, to the extent such data could be verified.

   Most women reported having healthy babies. For those who reported "visible" birth defects, the reported defects were difficult to classify in medical terms. These results suggest that clinical validation of birth defects would be needed in order to characterise teratogenic effects.

6. The effect, if any, of the frequency of data collection on the completeness and accuracy of reporting

   Women who chose to respond every two weeks, not surprisingly, provided more information than those who chose to respond every four weeks. It appears that the more frequent reporters also stayed in the study longer than those who chose to respond less frequently.

7. The extent to which women would report "sensitive" information about lifestyle and other risk factors for congenital effects

   It appears that women were willing to report use of alcohol, smoking tobacco and marijuana, and use of other illicit/recreational drugs.

8. The amount of loss to follow up and reasons for discontinuation

   Nearly two-thirds of women who initially volunteered for this study dropped out before the study was completed. Interviews with focus groups suggested that the primary reasons may have been due to the length of the questionnaires and the lack of any inducement or reimbursement for study participation. No women in the focus group reported that the questions were too personal or too intrusive or though some were concerned about the security of personal data.

Other questions of interest included

- Can we get data earlier in pregnancy than traditional routes?

  We learned that women can indeed be recruited early in pregnancy and the method of recruitment directly affects the stage of pregnancy at which women volunteer. For example, in order to boost recruitment, we used mail lists of women who had joined

pregnancy clubs, recognising that most of these women would be likely to have successfully completed their first trimester of pregnancy.

- How representative are the women?

  Respondents were slightly older and better educated than the general population and were not ethnically diverse in countries where there is substantial ethnicity diversity, such as the UK. There was broad geographic participation in each country. In addition, the results obtained are consistent with other surveys and national data on medication use in pregnancy.

- How important are data not captured by EHR or pharmacy databases?

  The pilot study showed that self-reported information via the internet was particularly useful for learning about short-term and intermittent medication use, and for information about lifestyle habits including use of alcohol, tobacco and recreational drug use. Information was also provided about medications that had been prescribed but not filled, and prescription medications borrowed from others that had been used by respondents.

- Is the information of sufficient quality to be used for pharmacovigilance?

  It appears that self-reported data from women via the internet represents a meaningful supplement to existing data available from electronic health records and from national registers on prescribed medication and birth outcomes.

Objective 2 (Signal detection, Work package 3)

The scope of IMI PROTECT Work Package 3 – Methods for Signal Detection was ambitious, with 12 sub-packages covering a wide spectrum of research questions, including the use of ontologies in pharmacovigilance, signal detection in spontaneous reports, signal detection in electronic medical records, and signal detection in clinical trials data.

An important strength of PROTECT's signal detection research was the execution of standardized analysis protocols across multiple spontaneous reporting sets, for several of the major studies. For these studies, databases of pharmaceutical companies, national regulatory authorities, and international organizations, such as the European Medicines Agency and the World Health Organisation, were considered and compared.

Standard adverse event ontologies are a basis for statistical signal detection. However, different terms can be used to describe the same suspected ADR, and it is not known whether grouping together related terms or treating them separately is preferable for timely statistical signal detection. PROTECT WP3 sought to determine to what extent use of standard MedDRA® groupings could expedite the detection of disproportional reporting for historical safety signals. The study found no overall benefit in conducting signal detection using MedDRA HLTs or SMQs compared with using individual PTs. Some relatively minor gain in time to signalling was seen when closely related (in a clinical sense) ADR terms were grouped together, and this area is recommended to be explored in future research.

Parallel to this, so-called knowledge engineering techniques were developed as a basis to proposing novel groupings of adverse event terms based on semantic definitions of each term. Such groupings

may go beyond and provide alternatives to the standard groupings available in an adverse event terminology, such as MedDRA Higher Level Terms (HLT) or Standardised MedDRA Queries (SMQ). PROTECT created an ontology for MedDRA preferred terms, OntoADR, with formal definitions of MedDRA preferred terms. The formal definitions were either inherited from mapped SNOMED clinical terms or defined in semi-automatic or manual processes. Knowledge engineering techniques can be used to derive novel groupings of adverse event terms based on semantic definitions of each term. Such groupings may go beyond and provide alternatives to the standard groupings available in an adverse event terminology, such as MedDRA Higher Level Terms (HLT) or Standardised MedDRA Queries (SMQ). Based on these definitions, groups can be defined dynamically, based on the relevant dimensions for a specific topic of interest.

A tangible outcome of PROTECT is the structured database of the ADR information in section 4.8 of the Summary of Product Characteristics (SPC) for all European centrally authorised products, in MedDRA. This database is updated and posted on the EMA web-site. It may reduce the need for manual inspection of SPCs when the focus of the monitoring is detection of new risks. This use has already been tested successfully at the EMA and the UMC and implemented into their corresponding signal detection processes.

Spontaneous reports remain the most relied upon data source for pharmacovigilance, and statistical signal detection algorithms are becoming widely used as core components of many organisations' routine signal detection. Methodological advances to address challenges with single drug- single event associations such as masking, effect modification, and confounding and to explore more complex patterns such as suspected drug-drug interactions have so far had limited impact on practical pharmacovigilance. Also, while there are now many assessments of performance characteristics of different proposed measures and thresholds, there has been little comparison across multiple spontaneous report databases, within a single study.

In a broad study across spontaneous report databases from pharmaceutical companies, national and international pharmacovigilance organisations, PROTECT evaluated the performance of common statistical signal detection algorithms. The choice of signal detection criterion (e.g. threshold on the number of reports, measure of disproportionality, and/or statistical significance) was much more important than the choice of disproportionality measure itself. Performance of any single algorithm might be very different between one spontaneous report database and another but the relative performance of two algorithms was generally similar in different databases.

A related study across nearly the same range of databases, explored the use of stratification and subgroup analysis in disproportionality analysis for spontaneous report databases. It showed that whereas subgroup analyses provided substantial benefits over crude analyses, stratified analyses did not. This is an unexpected finding that has not been reported elsewhere, but that did reproduce across the included data sets in the study at hand. It is important because several organisations use routine stratification, but very few consider subgroup analyses.

PROTECT also explored the impact of masking on statistical signal detection in pharmacovigilance. An example of masking would be when attention to a real or perceived safety issue in the medical community or in public media increases the reporting rate for that drug-event pair to such an extent

that the overall reporting of the event in the database as a whole is raised to the point where it becomes more difficult to identify statistical signals for other drugs with the same event. Under the conditions of the study (assuming that each ADR is masked by exactly one drug), it was rare that the masking affected whether a drug-event pair was considered to be disproportionally reported or not; the drug-event pairs that were affected in this way primarily involved rarely reported ADRs.

Adverse drug interactions harm large numbers of patients every year, and spontaneous reports may convey important information on previously unknown interactions. PROTECT research showed that statistical interaction measures with additive baseline models outperformed those with multiplicative baseline models, typically available in standard statistical software. This was true for both established and emerging adverse drug interactions – for emerging adverse drug interactions, the statistical interaction measures with multiplicative baseline models performed worse than chance.

Duplicate individual case reports distort statistical screening and can mislead clinical assessment. Many organisations rely on rule-based detection but probabilistic record matching is an alternative. PROTECT research showed that probabilistic record matching performed better than rule based screening, and should be considered as an alternative to such methods. Specifically, probabilistic record matching demonstrated a high predictive value above that of rule-based screening, and is expected to improve efficiency and accuracy of duplicate management.

A comparison between estimates of association from formal epidemiological studies and proportional reporting ratios (PRRs) in spontaneous reporting data for a set of known ADRs found a correlation, at a point in time before the ADR was first publicly recognised. This study suggests that it may be possible to use the PRR at the early phase of the analysis of a new safety signal as an indicator of the likely strength of the association, should the signal be confirmed.

At present, signal detection is predominantly based on spontaneous reports, but the use of longitudinal electronic health data in pharmacovigilance is an area of active research. PROTECT performed research on statistical signal detection in The Health Improvement Network (THIN) database of longitudinal electronic health records from general practices in the UK. A process for structured clinical and epidemiological assessment of temporally associated prescriptions and events in electronic health records was developed and evaluated. It showed that important potential safety signals can be identified in these data, whereas clinical and epidemiological review of highlighted statistical associations is crucial to attain an acceptable false positive rate. Conversely, a retrospective evaluation did not detect any of about 500 historical safety signals in THIN, prior to the initial signal at the EMA. In many cases this was due to the drug not being reliably captured in primary care data, and on a few occasions to the drugs not having yet been marketed in the United Kingdom. In contrast, some of the positive controls could be detected in VigiBase, even when the analysis was restricted to spontaneous reports from the UK. This shows that comprehensive surveillance for early safety signals requires broad population coverage as well as effective ascertainment of a wide spectrum of newly marketed drugs and adverse events.

Before approval of a drug, information on adverse events from clinical trials constitutes the primary basis for safety analysis and signal detection. PROTECT explored extreme value modeling as a basis

for predicting drug toxicity in subsequent phases of clinical development and evaluation. A retrospective analysis showed that extreme value analysis of phase 2 data would have highlighted the risk for liver toxicity with ximelagatran, a compound eventually withdrawn from all markets on account of this risk. A study evaluating different approaches to adjust for multiplicity found that Bayesian Hierarchical Models can improve signal detection performance through borrowing strength from related adverse events in the clinical trial dataset. This must be weighed against the more complex computational requirements of Bayesian modeling.

Objective 3 (Pharmacoepidemiology, Work package 2)

The overall objective of WP2 was to develop, test and disseminate methodological standards for the design, conduct and analysis of PE studies, applicable to different safety issues using different data. Since the beginning of the project, WP2 was organized in three working groups (WG). Summary statements for each WG are below.

Working group 1 (WG1): Databases

WG1's objective was to explain differences in drug-adverse event (AED) associations due to choices in methodology and databases by comparing results from PE studies. WG1 started with the selection of key AED pairs to be studied in different European databases. Selected AED pairs were those that represented the majority of decisions taken on drug withdrawal or major summary of product characteristics changes. The selection was based on literature research, EMA records, input from EFPIA Partners and expertise available in the Consortium. Six AED pairs were selected i.e. antibiotics and acute liver injury (AB/ALI); antiepileptics and suicide (AED/suicide); antidepressants/benzodiazepines and hip fractures (AD-BZP/Hip); inhaled long acting beta 2 agonist and acute myocardial infarction (B2A/AMI); calcium channel blockers and cancer (CCB/Cancer). In parallel, the basic characteristics of the partners' databases available for conducting PE studies were described. Partners' databases used were the Danish national registries (DNR); the Dutch Mondriaan database (Mondriaan); the British databases (CPRD and THIN); the Spanish BIFAP project database (BIFAP) and the German Bavarian claims database (Bavaria). These results were published in a scientific paper describing the background and rationale of WP2 (Abbing et al. Bridging Differences in Outcomes of Pharmacoepidemiological Studies: Design and First Results of the Protect Project Current Clinical Pharmacology. 11 Nov 2013).

Common protocols for each AED pair, 6 in total, were written prior to the conduct of the studies. Major efforts were put in the harmonization of the methodology to be used within each AED pair. Processes for requesting and preparing the data from the different databases were established and detailed data specification documents, including definitions of exposures, outcomes, and confounders for each database were written. A procedure to ensure the blinding of results from individual database analyses was developed. All protocols were registered in the ENCePP online registry thus are publicly available, in line with the PROTECT external advisory board recommendation, to enhance transparency and the dissemination of results.

WG1 conducted descriptive studies on the prevalence of use of the 6 selected drugs and the incidence of the adverse events in all databases. Plus, a total of 30 association studies in individual databases using different designs (i.e. 12 cohort studies, 8 nested case control studies, 5 case crossover studies and 5 self-controlled case studies) were conducted. Results from studies using the same design within the same AED pair were compared across databases in order to analyze and explain the discrepancies found despite applying the same methodology. Comparisons with replication studies conducted by WP6 were also performed. All comparisons included extensive sensitivity analyses on the main methodological issues.

WG1 has published 7 papers so far, 8 other papers have been submitted to journals and 10 more are in preparation. The final outcome of WG1 will be in the form of a journal supplement to include 20 papers from WP2/WG1 and WP6 related papers under the title "Improving consistency and understanding of discrepancies of findings from pharmacoepidemiological studies: the IMI-PROTECT project". The supplement will compile the methodological lessons learned drawn from all studies into a final recommendations section to be included at the end of the supplement. The WP2 co-leaders have signed an agreement with the journal Pharmacoepidemiology and Drug Safety (PDS). A guest editor has been identified and the majority of manuscripts are submitted and under review or currently responding to reviewer comments. The contents of the supplement have been defined. The supplement will include: a preface/introduction section with a description of the main objectives of WP2 and WP6; a selection of studies performed by WP2 and WP6 i.e. descriptive and association studies for antibiotics and liver injury, antiepileptics and suicide, antidepressants and hip fracture, benzodiazepines and hip fracture, beta2agonists and myocardial infarction, and negative control studies for antibiotics and myocardial infarction; and a final section with the challenges, lessons learned and recommendations. Currently, 1 submission has been accepted for publication in the PDS supplement. The estimated date of release is late 2015/early 2016.

Working group 2 (WG2): Confounding

WG2's objective was to evaluate and improve methods to control confounding by testing them in simulation studies and applying them in real-life data sets. WG2 developed a protocol to conduct simulation studies and WG1 partners' databases were used for empirical studies. WG2 research was organized considering 5 scenarios:

Scenario 1: An adverse event has multiple potential confounders. Two simulation studies to determine how to select confounders and how to model continuous confounders were conducted. Two papers were published.

Scenario 2: A relatively rare adverse event has a large battery of measured potential confounders. Three simulations and three real-life data studies on how to assess the quality of propensity score models and balance measures were conducted. Six papers were published.

Scenario 3: The treatment is time varying with a measured time-dependent confounder meeting criteria for an intermediate variable. Two studies with real-life data to compare Marginal Structural

Modeling (MSM) to ordinary methods to control for confounding were conducted. Two papers are available (1 published).

Scenario 4: There is a strong possibility of unmeasured confounding. Instrumental variable (IV) analysis was studied as ultimate solution to unmeasured confounding. Several studies were conducted: Two studies to identify the key assumptions; one study to quantify the key assumptions using balance measures; two studies to apply IV methods in real-life data. In addition, the prior event rate ratio (PERR) method was also proposed as a solution to unmeasured confounding. One study on how sensitive the PERR method is to violations of the key assumptions was conducted. Results determined that not all confounders may be measured, but somewhere information is available. Therefore a study on how to incorporate external confounder information was conducted. Five papers were published and 3 are still under consideration.

Scenario 5: An adverse event for which data come from multiple databases. One simulation and one real life data studies on how to combine data from different databases, looking at possible sources of bias and possible solutions were conducted. Two papers are available.

WG2 has published 13 papers. Several other WG2 papers will be included in the PDS supplement and lessons learned from applying methods to control for confounding in real-life data will add to the recommendations section. Two guidance documents will be delivered as final outcomes of WG2. One guidance document on IV analysis entitled "Instrumental variable analysis in randomized trials with non- compliance and observational pharmacoepidemiologic studies" has been published and another document on PS analysis "Practical guidance for applying PS in PE studies" is been prepared in the form of a chapter for a Handbook in Experimental Pharmacology. In addition, insights gained through the work of WG2 will be implemented in the upcoming revised ENCePP Guide on Methodological Standards.

Working group 3 (WG3): Drug utilisation (DU)

WG3's objective was to provide guidelines on how to identify, assess validity and use national drug utilisation data to estimate public health impact of AEs.

WG3 elaborated the inventory of Drug Consumption Databases in Europe, published in the IMI-PROTECT website (http://www.imi-protect.eu/drugConsumption.shtml). The inventory has been updated yearly during the duration of the project and the last version available is from February 2015. Three papers related to the inventory, one of them focusing on inpatient drug utilisation, have been published. Graphical results related to drug utilisation data for selected European countries are also available as part of the WG3 output.

WG3 conducted systematic literature reviews of the 6 selected AED pairs in WG1. Two papers on the review of studies on antiepileptic use and suicide and one on the review of studies on the association between bronchodilator treatment and myocardial infarction in COPD were published. A manuscript on macrolides- and amoxicillin/clavulanate- induced liver injury: systematic review and meta-analysis, was submitted. In addition, the assessment of the quality in systematic literature reviews of

adverse effects was studied and a paper on the validity of the scale developed to assess the quality is in preparation.

As part of the analysis of the public health impact of AED pairs, WG3 published several drug utilization studies across different European countries i.e. "Cross-National comparison of antiepileptic drug use: Catalonia, Denmark and Norway, 2007-2011"; "Sales of macrolides, lincosamides, streptogramins and amoxicillin/clavulanate in 10 European countries, 2007-2010"; "Excess risk of hip fractures attributable to the use of antidepressants in 5 European countries and the US"; "Potential impact of benzodiazepine use on the rate of hip fractures in five large European countries and the United States"; "Utilization and off-label prescriptions of respiratory drugs in children". The paper "Usage of tiotropium Respimat® vs. HandiHaler® in a real-life setting - comparison of patient characteristics' and TIOSPIR trial generalizability" is under review and "Trends in prescribing of long-acting beta2-agonists and inhaled corticosteroids after the SMART trial" has been accepted for publication.

WG3 developed a protocol for the calculation of the Population Attributable Fraction (PAF) across the AED pairs selected in WG1. These results are available in the report "Measuring public health impact of adverse drug reactions" delivered in February 2015. In collaboration with WG1, an assessment of the degree of comparability of the national databases in measuring drug exposure compared to healthcare utilization databases was conducted. One paper, using the key drugs selected in WG1 as examples, is under review.

Finally, the report "Drug utilisation studies in the PROTECT project: summary of achievements, and reflections from the collaboration with the EURODRUG Group" was delivered to the PROTECT Consortium in February 2015. It was a collaboration between WG1 and WG3 of the PROTECT project and the EURODURG Group (Björn Wettermark, Monique Elseviers, Robert Vander Stichele, Ria Benko, Vera Vlahovic-Palcevski). Its main objective is to discuss the applicability of aggregated data in DU studies.

Objective 4 (Benefit-risk assessment, Work Package 5)

Over the course of the IMI-PROTECT project, WP5 has carried out extensive academic and practical investigations into benefit-risk assessment methods. The main deliverables are available in the forms of study reports and a project website (http://www.protectbenefitrisk.eu/).

Of particular note are the following achievements, where innovative approaches have been used to evaluate methodologies and communicate our findings:

- A new taxonomy of methods for benefit-risk assessment;
- An up-to-date review of those methods, covering both matters of principle and practical application;
- Comparative testing of methods on real-life scenarios;
- A review of visualisation methods, considering the application of graphical display principles to benefit-risk problems;

- Confirmation of the principle that quantitative benefit-risk modelling of medicinal products is possible and desirable;
- A review of the role of patient and public involvement in the benefit-risk assessment process including comparative evaluation of preference elicitation methods;
- Effective collaboration amongst pharmaceutical companies, regulators, and academics, working together in teams to arrive at a common consensus, ensuring a variety of viewpoints are represented.

The experience of WP5 has been distilled into a clear set of practical recommendations for benefit-risk decision processes and supporting tools, and these are organised around the five stages of a generic benefit-risk assessment roadmap:

I. Planning: This stage encourages stakeholders to focus on critical issues related to benefit-risk assessment, including the purpose and context of the assessment. Clear documentation of discussions allows future analyses and updates to utilise the same foundations.
- Useful methodologies included frameworks, such as the Benefit-Risk Action Team (BRAT) and Problem, Objectives, Alternatives, Consequences, Trade-offs, Uncertainty, Risk and Linked decisions (PrOACT-URL) frameworks that organise data, with tree diagrams and structured tables providing useful means of visualisation.

II. Evidence gathering and data preparation: This stage identifies data sources and extracts evidence relevant to the benefit-risk assessment, and may include aggregation of multiple sources of evidence, which may require the use of estimation techniques. It encourages the systematic handling of missing data and requires engagement of clinical, statistical, epidemiological, and database expertise.
- Useful methodologies include Indirect/Mixed Treatment Comparison (ITC/MTC) and Probabilistic Simulation Method (PSM), and visualisation techniques such as structured and colour-coded tables, and network graphs to enhance the communication of data.

III. Analysis: In this stage, the data are evaluated, quantifying the magnitudes of benefits and risks, and perhaps weighing and/or integrating favourable and unfavourable effects as required by a given approach.
- Useful methodologies for analysis include metric indices which provide numerical representations of benefits and risks (Number Needed to Treat / Number Needed to Harm (NNT/NNH), Impact numbers), quantitative frameworks which model benefit-risk trade-off and balance benefits and risks (Multi-Criteria Decision Analysis (MCDA), Stochastic Multi-criteria Acceptability Analysis (SMAA)), and utility survey techniques which elicit stakeholders' preference information (Discrete Choice Experiment (DCE)).
- Visualisations recommended for the analysis stage include visualisation techniques specific for eliciting value preferences (tree diagram, method-specific visualisations such as MACBETH grid, Analytic Hierarchy Process (AHP) table, swing-weighting 'thermometer' scale, drop-down list), and visualisations for presenting analysis results (tables, forest/interval plots for qualitative or partially quantitative analyses; 'Difference display' (MCDA), and stacked or grouped bar charts for quantitative analyses).

IV. Exploration: This stage assesses the robustness and sensitivity of the main results to various assumptions and sources of uncertainties, considers impact or added value of risk minimisation measures, and likely requires both statistical and clinical input.

- Useful methodologies include ITC/MTC, utility survey techniques (DCE, AHP, Swing-weighting, MACBETH), PSM, and SMAA. Preferred visualisation techniques include the box, distribution, scatter, and forest/interval plots; tornado diagram; and most importantly, techniques that are interactive with the user.

V. Conclusion and Dissemination: This is the point at which, after considering all the information in the previous four stages, a conclusion is reached. The results and consensus from the benefit-risk assessment are then explicitly communicated to a wider audience, providing a transparent audit trail of the whole assessment process and bringing all aspects together in a holistic fashion. The content of the communication and visualisation methods used should match the needs of the intended audience.

Below are publicly available reports:

Final recommendations

- Recommendations of the IMI-PROTECT Benefit-Risk

Reviews

- Review of methodologies for benefit-risk assessment of medicines
- Review of visual representation methods of benefit risk assessment of medication (Part 1)
- Review of visual representation methods of benefit risk assessment of medication (Part 2)

Case studies

- Efalizumab case study on plaque psoriasis + Supplement 1 + Supplement 2
- Natalizumab case study on multiple sclerosis (Wave 1)
- Natalizumab case study on multiple sclerosis (Wave 2)
- Rimonabant case study on weight loss + Supplement (Wave 1)
- Rimonabant case study on weight loss (Wave 2)
- Rosiglitazone case study on type 2 diabetes
- Telithromycin case study on chest and throat infection
- Warfarin case study on stroke prevention in patient with atrial fibrillation

Objective 5 (Replication studies, Work Package 6)

Replication studies in relation to WP2

The overall scope of the "replication studies" of WP6 has not changed drastically since inception. All planned analyses were completed and results were compiled in papers submitted to the Pharmacoepidemiology and Drug Safety. Papers were either prepared by WP6 members alone or together with WP2 members.

The table below summarizes the initial objectives of WP2/6 work and the methods used:

Objective 1 **Replication study in same database**	Is the study replicable when conducted independently in the same	- CPRD - Danish Psychiatric, Somatic Hospital Discharge &	- Cohort study

	database?	Mortality Registers (DMR)	
Objective 2 **Replication study in different database**	Do the results have external validity?	- LabRx/Premier - MarketScan and Medicare - E3N - PGRx - UPOD	- Nested case control - Population case control - Cohort - Descriptive study
Objective 3 **Negative control study**	Does a study using the same protocol provide absence of evidence of an association where the exposure is such that the expected result is one of no association?	- LabRx/Premier - PGRx	- Nested case control (AMI) - Population case control
Objective 4 **Use of alternative outcome definition**	What is the impact of different levels of certainty of the outcome (e.g. definite, probable, possible) on the effect estimate?	-CPRD - PGRx - DMR	- Population case control
Objective 5 **Validation of outcome**	Has the outcome of interest been validated through clinical record review? What is the impact of validation on the effect estimate?	- CPRD - LabRx/Premier - UPOD - DMR - CPRD	- Population case control - Nested case control - Cohort study
Objective 6 **Assessment of confounders**	Has confounding been adequately taken into consideration? Are there additional confounders that need to be assessed? How does better control for confounding impact the effect estimate?	- UPOD - PGRx - DMR	- Descriptive study - Population case control - Cohort study

The table below shows the studies that had been planned and completed for each drug-adverse event pair.

Table: WP2/6 studies for each drug-adverse event pair

Drug-event pair	Database	Design	Status	Who?
Antibiotics & ALI	**LabRx premier**	Case-control	Completed	Sanofi
	UPOD	Validation study	Completed	Utrecht University

		Case-control	Completed	Utrecht University
	CPRD	Case-control	Completed	Takeda
Antibiotics and AMI (negative control)	**LabRx**	Case-control	Completed	Sanofi
	PGRx	Case-control	Completed	LA-SER
Antidepressants/BZD & hip/femur fracture	**CPRD**	Case-control	Cancelled	-
Beta2 agonists & AMI	**PGRx**	Case-control	Completed	LA-SER
	LabRx	Cohort	Completed	Sanofi
Ca Chan.Bloc. & Cancer	**E3N**	Case-control	Cancelled	LA-SER
	Marketscan	Cohort	Cancelled	-
Antiepileptic & Suicidality	**Danish register**	Cohort	Completed	Aarhus University
		Validation study	Completed	Aarhus University
	PGRx	Case-control	Completed	LA-SER
	CPRD	Validation study	Cancelled	-

The following table shows the objectives of replication studies per drug/event pair and data source that had been performed.

WP6 Research Plan for WP2 studies replication: Objectives per drug/event pair and data source

Drug / Adverse event pair	WP6 Partner	Data sources	Obj 1 same database	Obj 2 different database	Obj 3 negative control study	Obj 4 alternative outcome	Obj 5 validation of outcome	Obj 6 confounders
(1) Antibiotics & ALI	TAKEDA	CPRD	X			X	X	
	SANOFI	Invision Datamart		X			X	
	Utrecht U	UPOD		X			X	X
(2) Antiepileptic & suicidality	LA-SER	PGRx		X		X	X	X
	AARHUS	Danish Register	X			X	X	X
(3) Beta2 agonists & AMI	SANOFI	Invision Datamart		X				
	LA-SER	PGRx		X		X		X
(4) Negative control ATB & MI	SANOFI	Invision Datamart			X			
	LA-SER	PGRx			X			

Replication studies in relation to WP5

WP5/6-1: This activity consisted in testing/extending the WP5 methods in real-life setting. A MCDA benefit-risk assessment was used to address the needs identified by WP5. These needs and how they were addressed are summarised below. See the full report for more detailed explanations (https://eroombayer.de/eRoom/PH-GDC-PI-SID/IMI-PROTECT/0_106909).

- **WP5 recommendation: need for a consistent approach to benefit-risk assessment**

To address the issue of consistency, we developed a hierarchical benefit-risk tree that is generic at the criteria level, i.e., the same relative weights are applied for "Clinical efficacy" vs "PROs" vs "Safety" across different assessment. These weights capture the relative preferences of assessors for these broad concepts based on their individual value systems.

Outcomes (i.e., sub-criteria) under the "Clinical efficacy" and "PROs" branches of the benefit-risk tree were selected to represent the most meaningful outcomes from clinical and patient perspectives. The most relevant clinical outcomes are usually those recommended by the EMA and are expected to be consistent for a therapeutic area and across comparators, which was the case for PASI75 and PGA for psoriasis. For PRO criteria, variations do exist as well as absence of data for comparators. Most commonly used and validated PRO measures should be used and were selected for this study.

Under the "Risks/Safety" branch of the benefit-risk tree, instead of including specific safety outcomes, which will invariably vary between different interventions and diseases, we developed an entirely generic structure for this branch, consisting of 'AEs' (i.e., excluding serious and fatal AEs), 'Non-fatal serious AEs (sAEs)' and 'Fatal AEs'. The rationale for this design is that what matters for benefit-risk assessment are the consequences of the "risks" or AEs (e.g., were they fatal, severe, mild?) rather than the specific nature of the "risk" itself (e.g., a death from PML or from an infection should be given the same weight). Nevertheless, in the MCDA evidence matrix for AEs, sAEs and fatal AEs, decision-makers are provided with detailed information on the nature of AEs and the source of data so they can fully apply their own judgement when scoring. This "generic backbone" design of the benefit-risk tree allows transferability and consistency of benefit-risk assessments across different interventions. In addition, the use of constructed rather measured scoring scales (explained below) further allows transferability of this approach across different benefit-risk assessments.

- **WP5 recommendation: need for an approach compatible with comparative effectiveness/safety**

Data on healthcare interventions is generally comparative in nature, as an intervention is always assessed in context and compared to some existing path of care. This calls for an assessment that is directly based on the comparative data (e.g., vs placebo, vs usual care) reported in the literature.

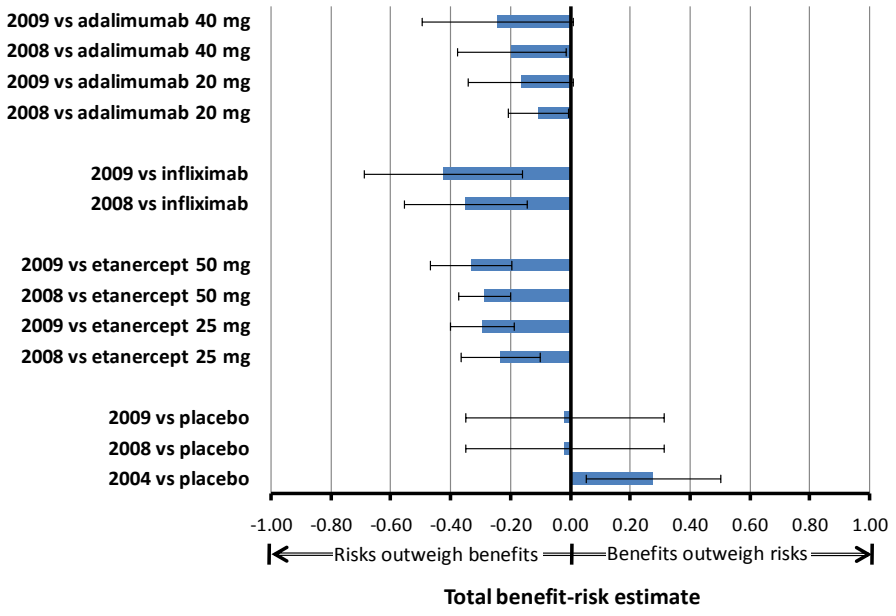
On a more fundamental level, the objective of benefit-risk assessment of therapeutics differs in an important aspect from the objective of standard MCDA. The latter compares the absolute performance of *several options* for a *specific decision problem* to select the best among them. In contrast, the objective of EMA's benefit-risk assessment is to assess whether *a specific medical intervention* has an *acceptable risk-benefit balance* (in the context of other available interventions for the indication) to authorize its use.

Therefore, the MCDA approach that we propose differs from the classical MCDA in that it is fundamentally comparative in nature, as the decision-makers have to judge directly for each decision criterion whether the therapeutic is 'better', meaning providing more benefit or less

risk, or 'worse', meaning less benefit or more risk, than the comparator (ideally current standard of care). An acceptable benefit-risk balance would then correspond to a situation in which the intervention has an overall positive benefit-risk balance compared to the current standard of care.

Furthermore, in addition to the importance of comparators, the panellists commented that the broader context of the intervention is also important in licensing decision-making. In real-life decision-making, criteria such as disease severity, unmet clinical needs and quality of evidence, impact the acceptability of the benefit-risk profile of an intervention. These could easily be added to the MCDA model using a similar approach as that proposed here.

Assessment versus active comparators involves considering that risks (AEs) may have a positive contribution to the benefit-risk balance (i.e., if there are fewer AEs for the intervention assessed than for the comparator) and that benefits (efficacy or PROs) may have a negative contribution (i.e., if the intervention is less efficacious than the comparator). To be able to take that into consideration without adding too much complexity, we broke down the assessment of efalizumab into four comparative MCDA exercises (1 placebo, 3 active comparators) and provided scales that include positive and negative scores. Then we designed a comprehensive visualization to see at a glance the benefit-risk balance of efalizumab in the context of all its comparators and across time (see figure below).



- **WP5 recommendation: need to avoid complex mathematical transformations to reduce opportunity for bias**

The MCDA approach tested in WP5 required each outcome to be numerically transformed into utility scores (i.e., scaled on a scale of 0–100). This step is complex. First, it requires arbitrary decisions on the two endpoints of the scale. (For example, in the efalizumab assessment, the maximum [utility score 100] for the outcome PASI75 was fixed to 60% of patients. As the basis for this choice is not given, one is left to wonder why, for example, 100%, meaning that

all patients experience the PASI75 endpoint, was not chosen to represent the maximum.) Second, if the relationship between outcome and utility score is deemed not linear, the software performs a non-linear transformation, which further adds to complexity.

Multiple computations and transformations increase the 'mental distance' between the reflection of the decision-maker and the data. Therefore, a framework aiming at supporting the deliberative process should keep computational complexity at a minimum to have face validity and be useful in practice.

To address this issue, we propose scoring scales that capture the intuitive judgement of the decision-maker on the data presented before him/her (i.e., 'constructed scales') rather than using numeric scales that represent mathematical transformations of the data itself ('measured scales'). The proposed scoring scales are:

- Identical for all Benefit-Risk criteria, which contributes to consistency across different assessments; and
- Designed for direct comparative assessment spanning "Much better than comparator" (score +5) to "Much worse than comparator" (score -5), with the score 0 representing "No difference."

These scales are easy to use and allow decision-makers to directly express their judgment on the performance of the intervention assessed relative to the relevant comparator. Although such an approach departs from the Keeney-Raiffa principles and has limitations associated with this, it nonetheless provides an acceptable and intuitive first step to transitioning from the current qualitative context of regulators towards a more structured and quantitative process.

- **WP5 recommendation: need for a flexible approach that can deal with uncertainty, lack of data, and heterogeneity of outcome measures**

These issues highlight a limitation of the measured utility scores scales used in WP5, which are based on the direct mathematical transformation from one outcome to one utility score using software with limited flexibility and constraints. To address the issue of uncertainty, constructed scoring scales, designed to capture judgments, can allow assessors providing a range of scores to express the uncertainty of data and their judgments.

Constructed scoring scales can also address the issue of data heterogeneity, as they can accommodate situations in which different types of data, which cannot be reduced to a single measure (e.g., based on different treatment durations), are available for a Benefit or Risk criterion. Also, this flexibility allows both absolute data (e.g., percentage of patients reaching PASI75) and relative data (e.g., odds ratios) to be presented and used in the MCDA evidence matrix.

Lack of data that would be required for benefit-risk assessment is a real-life fact, and much can be done to improve the current situation. However, absence of data for one criterion, such as fatal AEs, does not mean this criterion should not be considered or not included in the benefit-risk tree. Fatal AEs are inherent risks of some medical interventions. However, often they are not observed in licensing clinical trials but after long exposure of a large and diverse population. This highlights the need for long-term, real-world data to confirm/inform trial observations and conduct benefit-risk reassessments, and a long-term vision for the design of the benefit-risk tree.

- **WP5 recommendation: need for a simple weighting method that supports consistency between assessments**

There are many methods for weight elicitation in MCDA, some of which fairly simple and intuitive and based on pragmatic perspectives and other more advanced and based on decision theory.

In swing-weighting, assessors are asked to take into account for each criterion both the difference in performance between the least and most preferred options, and how much they care about that difference in comparison to the difference in performance for other criteria. The rationale behind this approach is to make sure that the weighted preference values have equal units across all criteria, i.e., that their scales are constant. However, this requires executing in one step both "comparative judgements about the ranges of effects and clinical judgements about how much they matter relative to each other", which can be cognitively demanding.

A point to consider in the benefit-risk context is that because the swing-weighting method requires defining performance ranges for each criterion, it is most applicable to a 'classical' MCDA decision problem, in which the goal is to select the best among a defined set of options. For such an application, the MCDA model can be constructed around the attributes of these options and the range of their performances. To quote from the recommendations (pg90): "*...the scales are defined locally for a given model, not globally across all models. In addition, the number of criteria and scale ranges can affect the values of normalised weights."* However, swing-weighting may be less applicable for assessing whether *any one* therapeutic product meets the pre-established standard for a positive benefit-risk balance which needs to be consistent and portable across assessments. A method that results in stable weights across different benefit-risk assessments may be more suitable for the needs of EMA committees who need to demonstrate a certain consistency.

To address some of the these issues in the benefit-risk context, a direct weighting method is proposed that involves distributing weights across different criteria and sub-criteria of the benefit-risk tree (hierarchical point allocation) *independently* of the performance (scoring) of the intervention.

Direct weighting methods have several potential advantages. They are intuitive methods which may provide a truer expression of assessors' perspectives. Also, they may allow better communication between assessors and facilitate examining the effects of changing weights.

Direct weighting methods have the limitation that they cannot guarantee that weighted scales will always be equal across different criteria. This problem is mitigated in a comparative assessment that uses constructed scoring scales and judges not the absolute performance but the difference in performance between the product and its comparator. When scoring, the assessors must then decide to what degree that difference is significant (e.g., "much better", a "little bit better" or "much worse"?), applying their own mental scaling based on their judgment of the evidence.

- **WP5 recommendation: need to establish a clear audit trail**

For full transparency and to establish an audit trail, it is critical to directly link judgment on the data (scores) and the data on which it is based. To address this issue, we developed the by-criterion evidence matrix as the basis for benefit-risk assessment. This matrix directly juxtaposes a description of the evidence with the score on that evidence. The description of the

evidence contains the actual outcomes data as well as a brief description of the data sources, which provides a richer context to the data than solely the numbers. This design accomplishes two objectives:

- It facilitates judgement on the available data by closely linking the evidence and the score and providing context.
- It ensures that an audit trail is directly established, on both the data that was used as well as the judgments on that data.

In addition, to facilitate understanding of the evidence presented and the clinical and quality-of-life consequences of AEs, we developed detailed glossaries in appendices to describe rare AEs that may be less familiar to assessors.

- **WP5 recommendation: need to include diverse stakeholders and their different perspectives**

Although the need to account for different points of view and include a wide range of stakeholders in decision-making is acknowledged, WP5 recommends arriving at a consensus for the importance (i.e., weights) of the criteria during the decision conference.

However, a participatory process that aims at encouraging a diversity of perspectives is unlikely to result in consensus. Rather, different stakeholders should have the opportunity to independently express their opinions on the importance (weight) of each benefit-risk criterion. These weights are to a great degree subjective, as they are rooted in the personal value system of the individual (e.g., whether more weight should be given to fatal AEs versus benefits), and therefore are expected to vary. In the proposed approach, the diversity of stakeholders around the table is reflected in the diversity (range) of weights given, and the group's average weight represents the overall opinion of the group. The panel session for this study included clinicians, regulators and methodologists, and the diversity of the group was reflected by the variations in weights observed across panellists. Of note, weights were very stable over time for each panellist.

- **WP5 recommendation: need for pragmatism and efficiency**

These comments highlight the need for a simple, flexible approach that can be easily implemented by users that have little prior experience with MCDA.

When designing such an approach, trade-offs between needs, feasibility, cognitive burden and the resources required to operate need to be considered. Therefore we intended to design an approach that would be pragmatic and flexible and would allow as much portability as possible between different assessments.

- **WP5 recommendation: need to achieve precision in weights**

Stability over time in weights is important and needs to be demonstrated. Therefore, to explore whether weights represent a stable expression of assessors' perspectives, a test-retest trial is part of the proposed methodology.

- **WP5 recommendation: need for appropriate visualisations**

The proposed approach aimed at an effective visual design to support the cognitive process of assessment and to communicate the results. This included a benefit-risk tree that directly

contains the definitions of each criterion as well as bar charts to report weights and different types of bar charts for the benefit-risk balance, including:

- Visual representation of positive and negative effects (stacked bar)
- Overall relative benefit-risk balance

WP5/6-2: The study was conducted as per the protocol. The primary objective of testing the applicability and acceptability of visual methods was assessed among patients diagnosed with atrial fibrillation, diabetes and breast cancer. Healthcare professionals practicing in these therapeutic areas and medical assessors who review regulatory submissions were also included in the study. The study was launched in three countries in the EU (United Kingdom, the Netherlands and France) and will provide comprehensive data on the 'best' method of communicating benefits and risks of medicines with patients and healthcare professionals. The study is also designed to test two methods of collecting patient preferences for treatment outcomes. The results from this part of the study will provide recommendations on the 'best' method to use in collecting preference data. There were no deviations from the original study design. Results of the study can be summarised as follows (for Version 7 of the full report, see https://eroombayer.de/eRoom/PH-GDC-PI-SID/IMI-PROTECT/0_1068fa).

- **Primary objective**: *Comprehension of benefits and risks using several different graphical formats.*

  The format that was best understood was the table format for patients with atrial fibrillation and diabetes. Patients with breast cancer best understood the drug vignette.

- **First secondary objective**: *Impact of presentation format and order on perception*

  Among breast cancer patients, the perception of the drug was most positive when the information was displayed by the survival curves and pictograms and least positive among the atrial fibrillation and diabetes patients when the information was displayed by the bar graphs.

  Among the experts, the perception of the drug was most positive when presented by the survival curves (presenting the benefits) in combination with the waterfall plot (presenting the risks). The perception score was least positive on the information displayed by the waterfall plot only (presenting both benefits and risks).

- **Second secondary objective:** *Impact of mood states on perception (only patients)*

  In patients with atrial fibrillation, being excited, sad or irritated was significantly related to the perception of the safety of the drug. Patients with these mood states had an increased likelihood of having a negative perception of the drug compared to patients with a neutral mood. These results were not found in patients with breast cancer and diabetes.

  In patients with breast cancer, being excited was significantly related to an increased likelihood of having a positive perception on the willingness to take the drug (instead of placebo) for their disease, whereas and being tense was significantly related to an increased likelihood of having a more negative perception on the willingness to take the drug for the disease In patients with

atrial fibrillation and diabetes mood states were not significantly related to the perception of the willingness to treat their disease with the drug.

- **Third secondary objective:** *Analyses of preferences*

  Discrete Choice Experiment:

  Patients with atrial fibrillation chose the prevention of stroke as the most important attribute while for experts fatal bleeding was the most important attribute; all other attributes were given the same order of importance by both groups. For diabetes, both patients and experts indicated preventing cardiac disorders as most important attribute of a treatment. However, the order of the remaining attributes differed. For breast cancer, the order of importance of all the attributes was the same for patients and experts, the percentage of patients dying within 30 months was indicated as the most important attribute.

- **Fourth secondary objective:** *Analyses of stated preferences*

  Among patients and experts of all three disease areas, the table format was most often selected as easiest to understand and most helpful in making a decision about which drug to prescribe.

- **First exploratory objective:** *Differences: Textual vs. graphical*

  Differences in comprehension

  Patients with atrial fibrillation have the highest comprehension when the benefits and risks are presented in a table. In patients with breast cancer the textual format was the best format regarding comprehension. In contrast, for patients with diabetes all graphical formats had a better comprehension score than the textual format.

  Differences in perception

  In all patient groups the presentation formats were not significantly related to the perception on the safety of the drug.

  In patients with breast cancer the survival curves in combination with the pictograms was significantly increased likelihood of having a negative perception on the willingness to take drug (instead of the placebo) for their disease. Patients with diabetes had a significantly increased likelihood of having a positive perception towards the drug when the information was presented in a bar graph (with the outcomes of the drug on the x axis). In patients with atrial fibrillation the presentation formats were not significantly related to the perception on the willingness to treat their disease with the drug.

  Differences in confidence

  Patients with atrial fibrillation had a significantly increased likelihood of having a lower confidence in understanding the presented information, when the treatment outcomes were presented in a stacked bar graph.

Patients with breast cancer had a significantly increased likelihood of having a lower confidence in understanding the information when the information was presented in survival curves and pictograms.

In patients with diabetes the bar graph (with the treatment on the x axis) had a significantly increased likelihood of having a lower confidence in understanding the information. The table format (using numbers out of 1000) had a significantly increased likelihood of having a higher confidence on understanding the information.

The development of the training material and the extension of the ADDIS software were completed according to the work plan (see https://mcda.drugis.org). As such, two important hurdles hindering the further uptake of the BR tools from WP5 into daily regulatory practice have been cleared. However, due to differences in agenda's in terms of implementation progress, timelines and priority setting, the interaction with the CHMP and the PRAC has so far been limited to the improvement and implementation of the Effects Table, which is only one of the tools from our BR assessment toolkit

1.6. Potential impact and main dissemination activities and exploitation of results

Objective 1 (Data collection, Work package 4)

Research is increasingly emphasising the importance of in-utero exposures to the future health of the individual. This includes factors such as nutrition and stress levels of the mother as well as individual life style factors. However, post thalidomide, exposure to medicinal products during pregnancy is one of the major concerns of in-utero exposure. Most drug exposures will probably not cause harm but the lack of evidence on the relative safety of drugs means that any exposure is likely to cause concern and worry to the pregnant woman which in itself may be damaging to the foetus. Termination of pregnancy may be an option chosen by some women because of the unknown potential for harm and the devastating effects such harm could have.

Some drug exposure during pregnancy cannot be prevented, either because the risks to the mother and/or foetus of the disease being untreated are greater than potential harms of medicines or because exposure has taken place before the mother has realised that she is pregnant. But not all medicines in the armamentarium of treatments for a particular disease will have the same effect on the foetus. Information on which medicines are "safe" during pregnancy, as well as which have the potential to cause harm, has immense societal benefits for European citizens both in protecting the health of the foetus, and in reducing the individual, familial and societal, emotional and financial burdens of caring for individuals damaged by harmful exposure. The hallmark of a civilised society is that citizens, who are unable to look after themselves due to physical or mental disability, are cared for, but that care is expensive. Therefore any research that contributes to the welfare of the foetus and therefore the future health of the citizen has immense societal benefit.

Considering these needs, this WP was designed to collect real-world evidence about medications used during pregnancy and to evaluate the utility of self-reported data for pharmacovigilance. We piloted whether collecting information on drug exposure during pregnancy directly from pregnant women would produce data suitable for research, and also whether it was possible to collect data in the early stages of pregnancy when the potential for harm to the foetus is greatest.

We found that it was possible to recruit pregnant women without the direct intervention of health care professionals although we also found that paid advertisements were necessary to alert women to the study. We were also able to recruit women earlier on in pregnancy than might be possible using some of the more traditional methods of data collection. We were able to collect details of non-prescription medicine use as well as herbal and homeopathic drug use – all of which have proved difficult by other means. Women were also willing to provide details of life-style choices such as alcohol, smoking and recreational drug use which are frequently not accurate or non-existent in medical records. This means that information on other risk factors and potential confounders can be collected to increase the validity of the observational research and the strength of the evidence.

This study conducted the research in four EU countries using the predominant language of each country. These countries had different health care systems and traditions and joined the EU at different times. The EU is a highly desirable locale for this type of research since, unless the drug is a major teratogen, large numbers of exposed women will be needed to show the potential for harm or conversely that the medicine is safe. With a population of 500 million, the EU can provide the necessary numbers. The existence of pan-European health care systems along with national centres of excellence mean that medical outcomes of interest are likely to be identified and recorded. The EU population is ethnically and genetically diverse so there is potential for the effects of different genomes to be studied including the interactions between genes, medicines, life style and other factors which again is useful for global research.

Our results suggest that the methods used are scalable and can be exported to other countries and languages. Extending this system across the EU would provide a resource that would be extremely attractive for biopharmaceutical research.

Objective 2 (Signal detection, Work Package 3)

A significant outcome of the Signal Detection Work Package in the final year of PROTECT are the recommendations distilled from its research efforts, and compiled in *The PROTECT Recommendations on Good Signal Detection Practices*. These recommendations are intended to provide actionable advice for pharmacovigilance professionals, particularly those with an interest in research and development of methods for quantitative signal detection. In total, they comprise 39 separate recommendations based on empirical research, and a further 27 recommendations calling for future research, reflecting the boundaries of the research completed under PROTECT.

As part of that, the completion of the evaluation of stratification and subgroup analysis for statistical signal detection in spontaneous reports was an important milestone. Its outcome was unexpected, with results suggesting that the common practice of routine stratification in statistical signal detection may be misguided and should perhaps be replaced by routine subgroup analysis.

PROTECT concluded a first-of-its-kind development and evaluation of a comprehensive approach to prospective signal detection in longitudinal electronic health data. The study brought to light 91 potential safety signals ranging from life-threatening medical events such as multiple organ failure to those that are less serious, but important for patients and for adherence, such as epiphora. At the same time, a key finding is that clinical and epidemiological review is essential to effective safety surveillance of electronic health data: three out of four medical events could be dismissed as unlikely to have been caused by the drug in question. The basis for these decisions included the identification

of earlier signs and symptoms suggesting that the disease preceded the initiation of treatment, and likely association with the underlying disease, including protopathic biases.

Objective 3 (Pharmacoepidemiology, Work Package 2)

Throughout the project, WP2 was consistently present at international Pharmacoepidemiological congresses and meetings. Notably, WP2 was widely represented in the past two annual International Conference on Pharmacoepidemiology (ICPE), sponsored by the International Society for Pharmacoepidemiology (ISPE) in Montreal (2013) and Tapei (2014), and midyear ISPE meetings in Munich (2013), Rotterdam (2014), and Bordeaux (2015). In both annual congresses combined, WP2 presented 25 posters, 12 oral presentations on various studies on methods to control for confounding and led two symposiums, i.e. "Improving Consistency in Findings from Pharmacoepidemiological Studies" and "Impact of methodological choices on findings from pharmacoepidemiological studies: final results of the IMI-PROTECT" where results from WP2 and WP6 studies were presented jointly and comparisons between results were discussed. At the midyear meeting in Munich a symposium on PROTECT WP2, WP5 and WP6 was organized by WP2 co-leads.

WP2 research showed that developing a common protocol for PE studies with great detail shall reduce methodological differences and interpretation by researchers. This requires a solid infrastructure for communication between sites conducting the same study. Conducting analysis in parallel in multiple databases instead of pooling of databases shall show heterogeneity and help exploring its sources. The large number of PE studies conducted in WP2 was the result of a close collaboration between 8 public and 9 private partners from the PROTECT consortium during 5 years.

WP2 widely disseminated its results on the extensive review of methods to control for confounding, focusing on instrumental variable and propensity score analyses. The guidance documents on IV analysis "Instrumental variable analysis in randomized trials with non- compliance and observational pharmacoepidemiologic studies" and the "Practical guidance for applying PS in PE studies" will contribute to the improvement of the use of those methodologies.

WP2 produced the inventory of "Drug consumption databases in Europe", publicly available in the IMI PROTECT website. It is a comprehensive and structured source of information on drug consumption in Europe. It comprises two documents. The master document is a detailed report of the available information, methods to retrieve it, a description of the validity of national drug consumption data and a discussion section. The country profile document summarizes the main results by country. Information is provided for 35 countries i.e. Armenia, Austria, Bosnia and Herzegovina, Belgium, Bulgaria, Croatia, The Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and United Kingdom. This inventory has been acknowledged by the PROTECT consortium as a very useful tool to the different stakeholders involved in pharmacovigilance and drug safety.

Objective 4 (Benefit-risk assessment, Work Package 5)

The benefit-risk assessment of a medicine or therapeutic option is a complex task in which many conflicting factors and viewpoints may need to be considered. To decide whether a treatment is appropriate, one must weigh up its favourable effects against its unfavourable effects along with the uncertainties around both. Such decisions have two main components: objective evidence of the treatment effects; and relative importance of the effects, which may vary according to stakeholder perspective. A sound assessment process should therefore (i) include the objective evidence (quantitative and/or qualitative in nature), (ii) recognise differences among stakeholders on the relative importance of each effect, and (iii) bring all this information together in a transparent and logically sound manner to determine whether or not a given treatment is justified, either for an individual patient or at the population level. The structured recommendations from WP5 could address these concerns.

Furthermore, WP5 demonstrated that collaborations between diverse stakeholders can be successful when there is a common goal, where multiple viewpoints, interests and priorities allow key methodological issues to be formally and transparently discussed to reach mutual understanding. This model of public private partnership can encourage expertise exchange and staff mobility between companies and countries, and therefore improve socio-economic benefits for the European citizens. WP5 also demonstrated that such successful collaborations have led to increased reputation of partners worldwide, which in turn could increase the competitiveness of Europe and help to establish Europe as the most attractive place for biopharmaceutical research and development.

WP5's contributions include reviews of frameworks and methodologies for benefit-risk analysis, and of associated graphical representations, provide guidance for selecting appropriate tools. There are particular stages of the benefit-risk assessment process where involvement of patients can provide added insight into the relevance of effects and improve the result and decision making by ensuring the correct focus is maintained. Case studies conducted reveal that formal, structured and transparent methodologies can be applied to real-world problems and contribute to the benefit-risk judgement by making the process more explicit and communicable. These structured approaches, when applied appropriately, can better contribute to the health of European citizens by ensuring the medicinal products available on the European market are based on robust and transparent benefit-risk assessment using the best available evidence.
Partners of WP5 will continue to champion the use of structured benefit-risk assessments of medicinal products within the individual organisations as well as externally, whether as individual or joined efforts. WP5 plans to further the application of structured benefit-risk assessment, through adoption and demonstration, beyond the scope of PROTECT to reach wider audience and trigger appetite among other scientific communities, patients; organisations, healthcare professionals and policy makers.

Objective 5 (Replication of studies, Work Package 6)

In relation to replicability of WP2 results, results were compiled in manuscripts submitted to the Pharmacoepidemiology and Drug Safety, either as stand-alone documents prepared by WP6 members or as part of the special issue together with WP2 members.

In relation to replicability of WP5 results, the VIZUALiSE study will provide comprehensive data on the 'best' method of communicating benefits and risks of medicines with patients and healthcare professionals. The study is also designed to test two methods of collecting patient preferences for treatment outcomes. The results from this part of the study will provide recommendations on the 'best' method to use in collecting preference data.

The development of the training material and the extension of the ADDIS software will support the further uptake of the BR tools from WP5 into daily regulatory practice. The day-80 template of the CHMP assessment report is being revised and the set of training tools that will subsequently be developed may incorporate some elements of the training material developed by PROTECT within Work Package 6, such as the discussion regarding the issue of double counting in the selection of the most important favourable and unfavourable effects

Objective 6 (Training and Communication, Work Package 7)

The **Platform of Training Opportunities** was launched in December 2010. During the whole period, there have been between 9 and 12 positions offered. Despite initial interest by many members of the Consortium, only two institutions offered training positions (FICF and Mario Negri). Until 24 April 2015, the Platform had received 13,859 visits by 2,693 users; the main topics of interest were drug utilization research (around 900 visits), training in clinical pharmacology (around 400 visits), and selection of the epidemiological strategy for specific drug safety issues (around 200 visits). Seventy-nine applications were submitted by 57 potential candidates, of which 12 were EU2P students. The backgrounds were Pharmacy (34), Medicine (5), Biosciences or Biology (11), and other (7). The countries of origin were UK (5), Germany (6), Italy (3), Spain (14), The Netherlands (4), France (2), Sweden (2), and others from Afghanistan, Argentina, Australia, Brazil, Colombia, India, Nepal, and Switzerland. The main fields of interest were pharmacovigilance and case-population research (20), drug utilization (10), and "collaboration with an ongoing study" (14).

The operational process for the identification of **training programme deliverables** and their transfer to the EU2P was established in Year 3. It consisted in identifying potential training topics from the PROTECT publication tracking list, and presenting each topic to the corresponding EU2P domain or training module coordinator.

1.7. Lessons learned and further opportunities for research

The mix of partners in the public private partnership brought together a diverse range of expertise which helped with planning and executing the project. Lessons learned, recommendations and further opportunities for research are discussed first in general and then for each work package as appropriate, as the work programme includes distinct activities with different further opportunities for research.

General aspects

Lessons learned

The collaborations in a public-private partnership (PPP) have been an added value to achieve the objectives of the projects. They were useful on the following aspects:

- building a network of and having access to a range of cross-functional experts / practitioners from a variety of backgrounds and independent of current affiliation;
- providing mutual insights into best practice and "applied practice" across PPP participants;
- providing access to a range of data sources;
- fostering mutual understanding of different perspectives and building trust , ultimately enabling awareness about pitfalls, constraints, and shared challenges;

- driving harmonisation of approaches;
- better understanding of IT tools;
- improving ability to build a shared toolset which can be cross-applied.

Recommendations that could be useful for a PPP

The following are recommendations based on experience. They may apply both at project level and at IMI governance level:

- there is a need for project management and administrative support from day 0; it should embedded in the project proposal;
- attempts should be made to minimise bureaucracy and certain documentation requests, and particularly to strive for a "leaner approach" to financial controlling (e.g. flat rates per capita as opposed to hourly "billing" etc.);
- provisions should be made for attrition / change of affiliation for individual participants to the extent possible during project lifetime;
- links should be fostered with scientific journals / publishers etc., possibly as part of advisory boards;
- publicity and training should be increased.

Objective 1 (Data collection, Work package 4)

Lessons learned

The study carried-out by WP4 for Objective 1 was led in each country by either an academic lead or a respected health institute which provided a local focus and contributed to the perceived trustworthiness of the study. This regional representation helped adapt the protocol to local needs, e.g., lowering the age of participation from 18 to 16 in the UK, and using methods for informed consent that were acceptable in each country for each modality. With regard to informed consent, in two countries, "electronic signatures" were acceptable as recorded by IVRS or internet, whereas in one country, electronic signatures confirming informed consent were only accepted by Internet and not by IVRS and in another country, volunteers were required to print and mail informed consent.

There were some aspects of the public private partnership which required careful management. For example, because of the strict regulation regarding the interaction between the pharmaceutical industry and patients, there was an understandable reluctance to offer any inducements to the pregnant women to encourage participation and retention in the study, even with regard to promoting attention to the study, so there were few resources available to support raising study awareness. Prior expectations on the public side that part of the contribution of the industry would include marketing expertise were unmet since it was mostly the research divisions of the industry that was engaged in this work package.

The data protection requirements were particularly difficult in such a multinational project. For the WP4 study, the EU data protection supervisor advised that all the partners in WP4 were "joint controllers" which required legal agreements between all parties to share liability, which caused complications – partly because applications were also to national bodies – not all of whom recognised joint controller status initially. These legal agreements slowed progress as did complicated negotiations in Poland with regard to informed consent, leading to a 22 month delay in starting recruitment in Poland as well as delay in other countries. The provisions within the new data

protection regulation could be a major complication in future public-private partnerships because of the financial implications for all if one partner fails in their duties.

Further opportunities for research

Initially, WP4 was the only work package in PROTECT which involved primary data collection to assess feasibility and whether the data would be of sufficient quality for research purposes. As such it has made a major contribution to our knowledge on what information is best collected from which source. There is still a large unmet need for research into drug exposures during pregnancy and the effect on the foetus.

Expanding this pilot into a full, on-going study would make an important contribution to the health of EU citizens but there would need to be additional incentives to keep women involved and there would need to be a software update to reflect drug approvals and withdrawals and adaptations to other languages used in the EU. It might also be very useful to change the type of questions used to elicit responses about medication use, since much of the data was reported as free text fields, rather than in response to questions about medication use for specific conditions. It might also be desirable to adapt data collection so it could be used on more platforms, such as smartphones and tablets, and to divide the baseline and follow-up questionnaires into smaller parts, administered more frequently, to reduce respondent burden. There seems to be no need for interactive voice response systems which is good since our work showed it is not ideal as a data collection method, nor did women choose to use it. Linkage to national databases could be used to augment the information on pregnancy outcomes and reduce the amount of information the women were required to input. However, analyses from the Danish data showed that it was difficult to match self-reported medication use, which included both prescription and non-prescription medications, with prescription registry data, both because of medications that could be obtained either by prescription or over the counter, and also because substantial information on the drug was entered as “free text” which made accurate matching of entity and dose challenging. More research is needed in this area.

Whereas an EU wide system collecting information from pregnant women could be of great importance for health, from the experience of this pilot, a public private partnership with multiple partners might not be the ideal way to move this forward given the issues raised above, and alternative funding options should be considered.

Objective 2 (Signal detection, Work Package 3)

Lessons learned

The public-private partnership was important in providing perspectives from both private and public partners and the different environments used in signal detection. It also brought additional data sets for analysis and this was critical to the success of our project and ensured the relevance of the recommendations to a greater part of our community.

A challenge for both regulators and companies is that the project was often added on top of the day job, in contrast to academic institutions who will recruit new resources with the grant money; academic institutions on the other hand often do not have the same insights into the real needs and challenges of the field. A critical success factor for the future would be to secure and protect the

engagement of all scientists. This is particularly challenging for the EFPIA companies who make in-kind contributions.

Further opportunities for research

- *Addition of dates in the PROTECT SPC-ADR database*

A preliminary task in setting up the subpackages of IMI PROTECT concerned with signal detection was compiling a large database of established adverse drug reactions against which the truth or falsehood of statistical signals could be evaluated. The majority of these ADRs were taken from the structured database of SPC Section 4.8 for centrally authorised products that was the output of yet another PROTECT subpackage. The number of ADRs in the reference database was 23,742. This size of dataset provided statistical power to investigate subgroups of drugs but meant that additional ADR specific data that were not already coded in the SPC dataset would have required considerable effort to obtain.

One such additional piece of information that would have been useful is the date at which each ADR was added to the SPC. With this date it would have been possible to determine whether statistical signals preceded the type II variations and hence whether they might have provided useful early warning of the ADR. These dates are potentially available to regulators in text documents but natural language processing techniques may make it possible to link many of these terms to text documents.

- *Coding of non-centrally-authorised products*

The PROTECT SPC dataset is a useful research tool but possibly of greater use as in the routine process of signal detection. In this context it allows automatic identification of signals corresponding to ADRs that are already listed in section 4.8 of the SPC and hence avoids the work of making this determination by hand. To enhance its usefulness to Member States, coding of a much wider range of products would be required. Some very preliminary work on this topic was done in PROTECT SP3.1 but much more detailed work is required if a process for extracting such data from, possibly differing, national SPCs in a number of EU languages is to be developed.

It is unclear how much added value for testing signal detection methods would be provided. There is no clear evidence that signal detection in CAPS is different from that in other products. However, this question might, of itself, be worth answering using a moderately sized subset of NAPs in order to investigate the robustness of research results.

- *Identification of synonyms in MedDRA PTs*

In PROTECT, no standard level of the MedDRA hierarchy was found to give superior signal detection performance compared to the preferred term level. However, a finding of the SPC recoding and also of earlier work (Alvarez) was that some terms within the PT level are essentially the same in as much as it is unlikely that a reporter, with limited clinical information, might choose them interchangeably. Such terms will undoubtedly delay the signalling of some ADRs because a threshold is placed on the numbers of reports in the definition of an SDR. Identifying and equating such ‘effective synonyms’ for the purpose of signal detection was identified as a promising way to improve signal detection and

should be considered in any future development work. The mapping of synonyms to the ADR database would provide more complete identification of listed ADRs.

- *Product coding levels*

An unresolved issue of signal detection is how to treat multi-constituent products. It is likely that a product containing substances s1 and s2 will have ADRs in common with one containing s1 alone. However, whether these products should both be included when performing signal detection for s1 is not a matter that can be decided by principle. In practice it might vary from substance to substance or depend on the nature of s2. However, in screen for potential ADRs , we would like to know what general rule is likely to render the signal detection most effective. Of course, a general rule might be more complex than a decision to always combine or always keep separate. However, whatever rule is chosen it should be based on empirical evidence that demonstrates that it works.

Strategies for combining adverse event codes were tested in PROTECT but no equivalent work was carried out for substances.

Allied questions relate to combining over classes of products and also to comparing classes of products. Some work has been done in this area (H1N1 vaccines, anything using ATC?) but general rules for screening large numbers of products are still lacking.

- *Examining restricted background distributions*

The question of whether disproportionality statistics will discriminate ADRs better against background distributions of adverse events obtained from restricted sets of products has been previously investigated. However, larger studies with more categories of background are required to draw conclusions for signal detection over a wide range of products.

- *Using additional variables in signal detection in spontaneous data*

Disproportionality analysis is based solely on aggregate numbers of reports, disregarding report quality and content: previous studies have shown that simultaneously accounting for aspects such as number of informative reports, free test descriptions, geographic spread, recent reports… significantly improves the accuracy of automated screening of individual case reports compared to disproportionality analysis alone.

Utilising such a model can be expected to reduce the number of false alerts and uncover drug safety issues that would otherwise go undetected or being detected with a delay.

- *Subgrouping in signal detection*

Carrying out signal detection separately in various subgroups of ICSRs can improve performance. How much of this performance change is due to exploiting information carried by the subgrouping variable can be explored by comparing the results to those obtained with random splits of the ICSRs. The ‘best’ use of subgrouping variables seems to depend on the nature of the signal detection algorithm and this, and the very large number of possible subgroup calculations possible, suggest that a systematic investigation may be worthwhile.

- Determinants of precision in signal detection in SR databases

An observation from the PROTECT project was that increasing size of database appeared to change the precision of signal detection using disproportionality. Some further investigation of this phenomenon may help to explain the mechanism of signal detection and improve efficiency.

A possibly related effect was that precision reduced for specific products with time on market. Several competing explanations were hypothesized for this effect and further investigation could be considered.

- Matching signal detection to available resources

The nature and number of adverse drug reactions that remain undiscovered cannot be known. Hence it is difficult to set limits on the resources, particularly human resources, to allocate to signal detection activities. However, for any fixed resource allocation, the most efficient way to use them can be calculated and this may vary from organisation to organisation. For example, if very limited resources are available maximisation of positive predictive value is paramount.

The investigations of signal detection performance in PROTECT and other projects can be applied to many pharmacovigilance settings but the way that the findings are implemented will vary. An important additional piece of information is a cost-effectiveness calculation regarding the best way to use these results in different settings.

- Follow-up of signal detection

An important topic for subsequent calls would be to explore, evaluate and advance the steps that follow statistical signal detection, especially, the signal validation, the manual clinical causality assessment and signal prioritisation.

Objective 3 (Pharmacoepidemiology, Work Package 2)

Opportunities for further research

An important constraint in the decision-making process in pharmacovigilance is the timely investigation of a safety signal once it has been detected. This delay may have an important public health impact. If the signal is refuted, patients may have been deprived of a safe and effective treatment if precautionary measures have been taken, and, if the signal is confirmed, patients may have been exposed to a harmful drug. In addition, uncertainties about the time required to investigate potential safety issues may lead regulators to request active surveillance studies at the time of authorisation. Therefore, public health, regulatory authorities and industry would benefit from a system facilitating the conduct of high quality observational safety studies. A key element of such system is the ability to combine data or results from several population-based health care databases. This approach is particularly important for rare, severe or long-term adverse reactions and investigation of safety issues for several products of a same class with different drug utilisation patterns in different countries.

A further project would be to test the feasibility of an international collaborative network for evaluating safety signals of global relevance in a wider range of epidemiological datasources such as the Canadian Network for Observational Drug Effect Studies (CNODES).[5] PROTECT and CNODES have both developed a collaborative population-based approach whereby drug-event associations are analysed separately in each database using a common protocol and common data analysis methods. Outputs of each study may be combined by meta-analysis.[i] The model of a collaborative population-based approach would differ from the distributed data model, in which electronic data are maintained and controlled locally and a common data model is used to execute standardised programs and share the output of these programs in a summary form (Mini-Sentinel) or to transmit aggregated data electronically to a central data warehouse for further analysis (EU-ADR).[ii] The following objectives could be pursued: 1) to test the feasibility of using and developing common methodology and study protocols to investigate safety issues of global relevance; 2) to test the consistency of results on a same safety issue obtained in each network based on a same protocol; 3) to test the feasibility and constrains of a global network to study real life safety issues in terms of sharing of information, timeliness of results and interactions with regulatory authorities and industry; 4) to develop core principles for a potential sustainable global network Lessons drawn from this testing would be used to address potential collaboration with other networks.

Objective 4 (Benefit-risk assessment, Work package 5)

Lessons learned

Input from the different public/private stakeholders has been key in the WP5 work in terms of evaluating benefit risk techniques with regard to the different uses that these may be put to and it has been crucial for WP5 to have active input from industry, regulators, academia and patient groups. WP5 recognises that the teams' experience at the start of using the methods was limited or non-existent; the teams learned as they applied the methods, and they found that the Wave 2 case studies were completed more expeditiously than the Wave 1 cases. As might have been expected in using unfamiliar technology, experience in using the methods reduced the time it took to complete an assessment.

Throughout PROTECT WP5's case studies, there are two limitations that have been encountered time and time again, regardless of the particular methodologies being employed. These relate to the benefit-risk time horizon and the extent of publicly available data.

None of the methodologies identified are designed to quantify changes in benefits and risks over time (except perhaps some health indices for specific disease areas, e.g., QALYs, but even these handle time in a rigid, pre-defined way). The standard approach is simply to focus on a particular time period of interest, and update the analysis when warranted by additional information or at designated time periods. This approach is probably sufficient for many purposes, but it would be interesting to see if methods can be developed that explicitly model the dynamic nature of benefits and risks over time.

The researchers were frequently frustrated by the lack of publicly available effects data, particularly at the individual patient level. Changing this situation may require significant political will; the Wave 2 rosiglitazone case study team went as far as to recommend "that the European Commission

[5] CNODES is part of the Drug Safety and Effectiveness network (DSEN) and part of the Canadian Institutes of Health Research (CIHR). CNODES research teams and collaborating centres form a coordinated national network of 69 researchers and databases from 7 Provinces with the aim to create the capacity to respond in a timely manner to the drug safety queries of regulators.

investigate this issue of data availability and take steps to ensure that patient-level data about clinical studies of medicinal products are properly archived and made accessible." Related to this is the problem of heterogeneity of the outcome measures reported in clinical trials. Establishment of a standard reporting template that facilitates the extraction of data, including measures of uncertainty, would be a great step forward.

Most recently, there have been a number of steps forward with regards to data availability. An Institute of Medicine workshop on sharing clinical research data took place in October 2012, and a summary of the findings are now available. The European Medicines Agency published a draft policy on the publication and access to clinical trial data in June 2013. The European Federation of Pharmaceutical Industries and Associations (EFPIA) and the Pharmaceutical Research and Manufacturers of America (PhRMA) launched their joint principles for responsible clinical trial data sharing in July 2013. Efforts being undertaken by GlaxoSmithKline to provide access to anonymised patient-level data has also been recently published (Nisen and Rockhold, 2013).[6]

Opportunities for further research

WP5 identified some challenges in structure benefit-risk assessment that had been anticipated assessors may have to address in the future:

1) Pre-marketing / Licensing
   - *Limited Evidence*

     Limited evidence does not mean that there is no value in carrying out a benefit-risk assessment during the early stages of the product life cycle. Benefit-risk assessment is always a dynamic process to be undertaken throughout the use of a treatment, rather than as a single determination. A decision regarding the benefit-risk balance should be based on the best evidence available at that point in time and may later change as new evidence becomes available. Benefit-risk assessments under limited evidence should be investigated further, e.g. when it applies to difficult-to-study subgroups.

   - *Relevance of outcome measures*

     The outcomes recorded in clinical trials may not have been measured in a way that is optimal for the purpose of a benefit-risk assessment. Surrogate measures, and varying definitions and quality of measurements are some key issues. Where such concerns exist regarding the relevance of outcome measures, these should be documented carefully and fully so that the decision makers using this evidence can make informed decisions on the relevance of certain outcome measures. This documentation should also be revised and addressed in future periodic assessments. The relevance of outcomes should be fully studied.

2) Post-marketing
   - *Long term follow-up data*

     Where trials have followed up participants beyond the original trial period, they can provide a useful source of data on a treatment's long term effects. However, analysts and reviewers

[6] Nisen, P. and Rockhold, F. (2013). Access to patient-level data from GlaxoSmithKline clinical trials. New England Journal of Medicine, Aug 1, 369(5): 475-478.

of a benefit-risk analysis will recognise that the controlled nature of a clinical trial breaks down at the end of the original study period, and data from that point on is more akin to that from an observational study. Assuming the trial had a positive result, the control subjects will often have been switched to the active treatment after the end date, meaning that long-term control data may not be readily available. Such extensions to comparative clinical trials may also encounter more issues with compliance and confounders. As with any analysis, the sources and degrees of uncertainty, and their likely impact should be clearly documented.

- *New evidence of efficacy and safety*

  Where the new evidence has come from a study that is not sponsored by the company and does not fall under the EMA's clinical data transparency regime, only the published summary results may be available. Integrating this information into a benefit-risk assessment based mainly on the company's own data may present challenges. Some techniques reviewed and tested in WP5 may be viable options but need to be fully tested.

  Where the new evidence specifically relates to a different patient group from that for which the product was originally licensed, a separate benefit-risk assessment may be required for these patients. This is not simply a case of changing the data in the existing assessment; for different patient groups, the decision context will vary and so the entire assessment should be revisited from the bottom up. Assessors will need to consider whether it is appropriate to assume that the efficacy and safety profile is similar between the different groups, and should be studied further.

- *Observational / surveillance data*

  Epidemiological studies, registry reports, and spontaneous reports may provide important data on emerging risks. As is true of clinical trials, there is potential bias associated with each source. The source of data can also be considered in terms of quality of evidence, e.g., CDC hierarchical system. Aggregating the evidence with that observed in clinical trials may also be problematic. Statistical methods may exist to deal with these issues, but this remains a relatively specialised field and not all assessors may have the resources for such approaches (or consider it appropriate to use such complex techniques for the decision at hand).

- *Well-established products*

  Products with a long history on the market have the advantage of cumulative data. However, this may mean that the BR assessment of mature products may need to be done using multiple data sources of varying quality, including missing information. The regulatory paradigm was likely not as robust as it is today, resulting in less comprehensive documentation of evidence at time of approval. In addition, there are practicalities, such as the loss of archived information, that impact the ability to introduce data into a benefit-risk assessment. Some sort of sensitivity analysis may deal with this issue, but the best practices for this scenario are still evolving and needed.

Objective 5 (Replication of studies, Work Package 6)

<u>Lessons learned</u>

In WP5/6-1, the MCDA-based approach tested allowed pragmatic quantification and visualization of the relative benefit-risk balance and its contributors over time and across comparators (placebo and active comparators). Advanced statistical analysis and longitudinal modelling allowed providing

effect estimates when relevant evidence was limited, which is often the case in real-life decisions at the regulatory level, especially at point of licensing but also post licensing. The overall structuring aspect of this framework was deemed promising by the panellists, with the expandability to add knowledge over time and expand branches as necessary. The modular aspect of the proposed framework could easily allow integrating additional criteria on top of benefit and risk, such as disease severity, unmet needs and risk management plans, to design holistic MCDA approaches

The scope of the WP5/6-2 project was ambitious in that the objective of capturing data from patients, health care professionals and medical assessors in 3 EU countries has never before been carried out. This required coordination of institutions and ethical review boards in several countries and this is only possible if there is close collaboration between partners in the respective countries. In addition the support of the EMA as a multilingual/culture institution aided the success of the design and implementation of the project.

Opportunities for further research

It has been shown that combination of pragmatic MCDA and advanced statistical analysis into a ready-to-use decision matrix supports the applicability of MCDA-based approaches for real-life decision making to further transparent, consistent and comprehensive benefit-risk assessments. Implementation of an MCDA-based risk-benefit assessment framework in real-life decision-making could proceed in a step-wise fashion taking into account the context of committees and organisational constraints, such as the resources available and training requirements. Further research on how to integrate MCDA-based approaches in real-life decision-making is warranted.

There is clearly scope for developing the science around how to improve communication of benefits to patients. From the results of the VIZUALiSE study, patients appear to be very willing to be engaged and this can be seen by the response shown in at least one of the study countries. Also work on the cognitive and behavioural aspects of eliciting patient preferences and the impact of these aspects on the elicited preferences need further exploration. Furthermore, elicitation of preferences at a single time point was explored but this may not be fully reflective of the changing nature of patient's views as their disease develops.