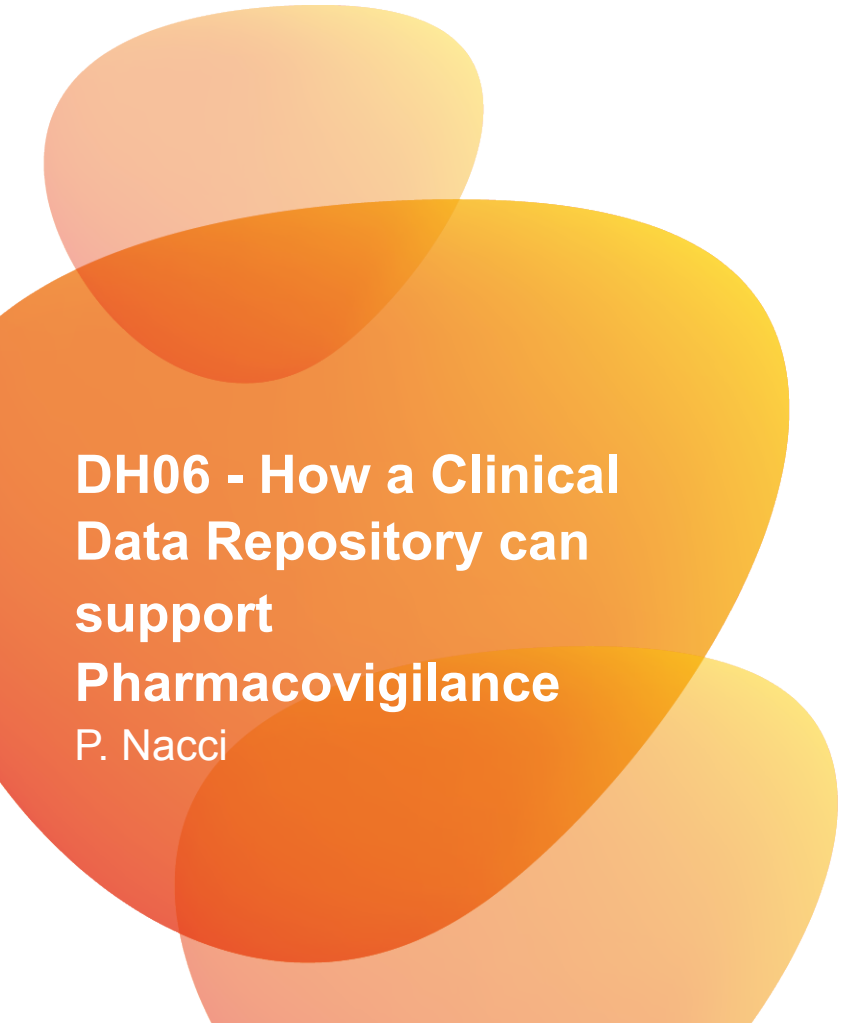




DH06 - How a Clinical Data Repository can support Pharmacovigilance

P. Nacci



Introduction

NB The non-flu part of the Novartis Vaccines business unit was acquired by GlaxoSmithKline (GSK) in March 2015

- Soon after the 2009 flu pandemic crisis was over Novartis Vaccines initiated a project to retrieve, recode and pool all legacy studies for which electronic data were available in-house within the framework of a Clinical Data Repository (CDR)
- All new EDC trials would be run in the new environment
- The software product chosen as the hub of the CDR was SAS Drug Development (SDD)
 - SDD Version 4.x was evaluated, but version 3.5 was actually delivered
- A slightly adapted variant of CDISC SDTM 3.1.2 was selected as the standard data structure for long-term storage, analysis and reporting
 - Proper management of study metadata was included
- EDC panels were redesigned to collect CRF data in CDASH format, making automatic remapping into SDTM feasible

- Most legacy study data have now been remapped, for a theoretical total of around 400 studies, including flu ones
- Study metadata have also been retrieved
- Re-mapping of legacy data is done using custom-written SAS programs
- When applicable (re-)coding of adverse events, concomitant medication, medical history, etc. is performed in Oracle Thesaurus Management System (TMS)
- Our home-grown Standard Reporting Software (SRS) suite of validated SAS programs was adapted to run in the new environment using SDTM data structures
- New ADaM-compliant reporting programs have been developed, validated and put in production during the past year
- Last but not least, all our existing processes were migrated to the new platform, and several new ones mapped and properly documented

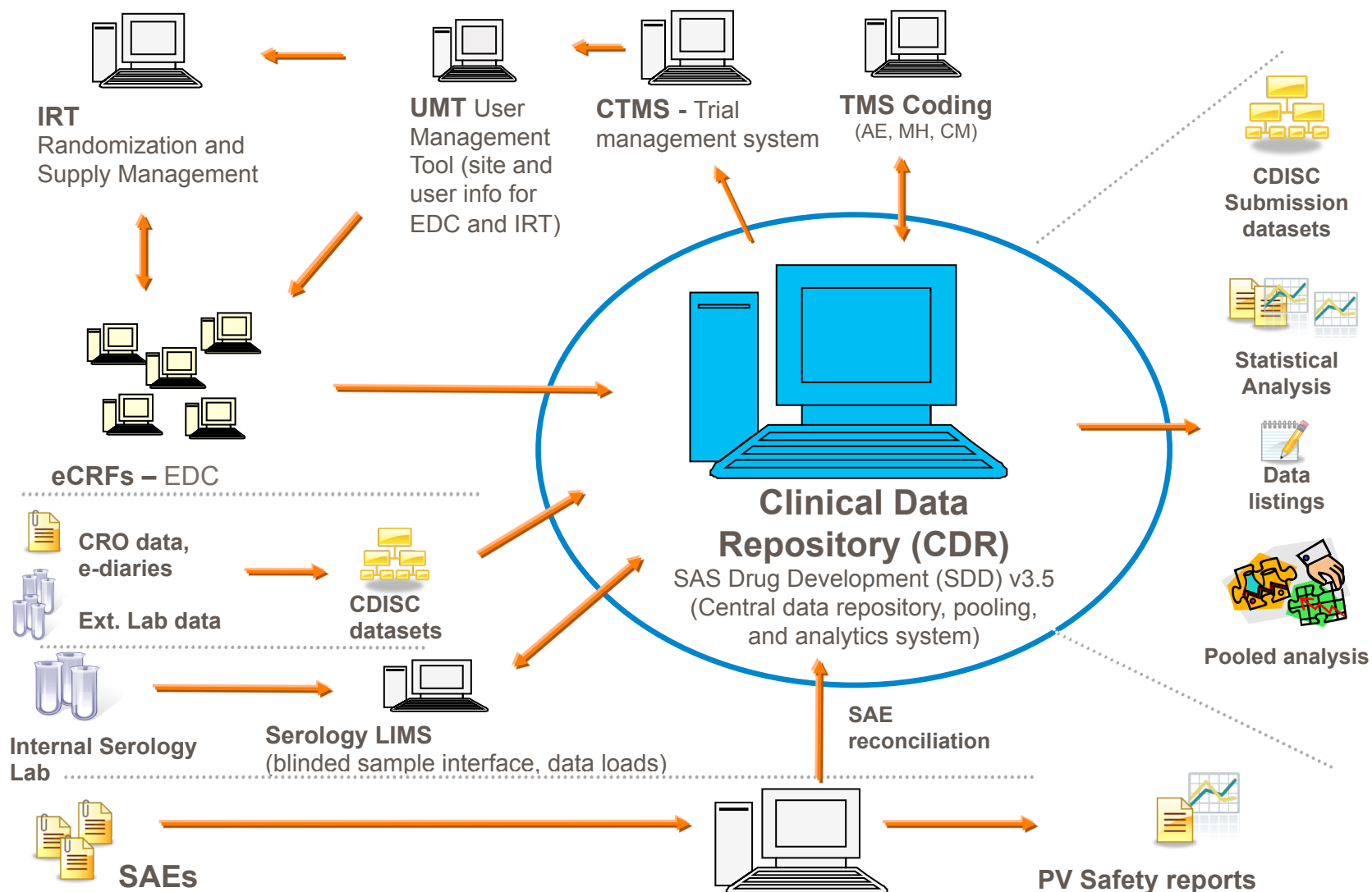
The first phase of the development of CDR encompassed several projects:

1. Design, configuration, and validation of SDD 3.5
2. Definition of new CDISC-based data standards (CDASH, SDTM)
3. Several interfaces & database connections
4. New centralized coding system
5. Re-write of SOPs, Work Instructions, and SOP related documents
6. Legacy data conversion (roughly 400 studies, from 1992 onwards)
7. Internal Standard Reporting Software SAS programs conversion

Further developments on which we are actively working:

- Definition of an ADaM standard and relative SAS analysis programs
- Migration to SDD 4.x
- Development/refinement of processes to better support regulatory submissions from the clinical trial data point of view (e.g., creation of define.xml, better bookmarking, etc.)
- Data visualization packages
- Support for other internal functions in addition to Clinical

CDR Overview



- Almost four years after CDR went live, all new clinical studies are now managed completely within it
 - EDC data get pushed automatically into the system at regular intervals (min. 20 minutes)
 - SDTM data are then obtained directly from CDASH dumps nightly, so that the latest version is always available online
 - Deviations from the standard eCRFs are dealt with using a set of study-specific 'rules', and previous approval by a technical committee is necessary
- Data from almost **300** legacy studies (including flu ones) are now available for use, together with those from **~70** CDR-native studies
- The SRS suite has been tested on real SDTM data, necessary fixes have been applied and new enhancements, including compatibility with ADaM data structures, are either already available or actively being worked on

- A utility to create and maintain complete and ongoing (dynamic) ‘oceans’, including the capability to update all AE and (partly) MH verbatims to the latest MedDRA version, has been fully validated, and is run nightly
- Another utility to create (static) ‘data poolings’ for analysis has been validated too, and is in full use
 - Several data poolings have been created and released for analysis



Supporting Pharmacovigilance

“All medicinal products in the EU are subject to a strict testing and assessment of their quality, efficacy and safety before being authorised. Once placed on the market they continue to be monitored so to assure that any aspect which could impact the safety profile of a medicine is detected and assessed and that necessary measures are taken. This monitoring is called **pharmacovigilance.**

Pharmacovigilance is the process and science of monitoring the safety of medicines and taking action to reduce the risks and increase the benefits of medicines. It is a key public health function.”

- Similar definitions are available in the USA, from WHO, etc.
- WHO established its Programme for International Drug Monitoring in response to the thalidomide disaster, detected in 1961
- Historically this has been managed in the pharmaceutical industry by means of dedicated databases containing serious adverse reactions (SARs), organized in ‘cases’
 - MedDRA-coded data are kept homogeneous by always using the latest dictionary version
- As far as I know, analysis of clinical trial data is mostly performed at the single study level, or at most using *ad hoc* poolings of selected studies, often never reused again

- We currently use Argus, which is one of several such databases
 - One of the tools included by default allows the formulation of complex searches on reported medical terms using all levels of MedDRA terminology, up to and including Standardised MedDRA Queries (SMQs), which are basically validated, version-specific standard sets of MedDRA terms
- Over time we have developed a standard internal form, which the physicians in the PV department must always fill in to fully specify the search scope and criteria
- A new ticket is created by the requester in a tracking system with the completed form attached, and the resulting output is added to the same ticket before closure, so that all requests and their outcomes are fully logged
- The ultimate goal is to ensure that the implementation path is clear for the operators and that final results are in line with expectations

- As with all such complex systems, CDR was initially just a resource black hole: plenty of work by many people went into setting it up, defining and documenting its processes, uploading data, etc. but very little came out of it
- Things changed once the first lot of SDTM-remapped legacy data, including 153 studies, was finally made available
- The first tests replicated previous calculations used, e.g., for the yearly updates of Investigator Brochures (IB)
 - Study-level actual exposure numbers were obtained by a single PROC FREQ, and the few discrepancies investigated and quickly solved
- In the case of extension studies we were able to easily identify subjects who had been already exposed to the vaccine of interest in the parent study, thus avoiding double counting them
- The identification of all studies where administration of a certain vaccine component/formulation was planned was also possible

- The next step was to identify (all studies including) subjects who have reported at least one adverse event (AE) which was coded to specific MedDRA Preferred Terms (PTs)
- Several different scenarios were then available, e.g.,
 - All AEs of interest reported in the selected studies
 - All AEs reported by subjects exposed to a certain vaccine component/ formulation, etc.
- To reach these goals two programs were prepared, and several specific SAS macros were developed along with them, in addition to the general-use ones already available as part of our SRS suite
- A detailed explanation of how the new SAS macros work can be found in the paper

This is the high level structure of the first program, calculating both planned and actual exposure to a certain vaccine component or formulation:

- Create a pooled TE domain from complete and ongoing oceans
- Identify all studies with at least one planned arm containing the vaccine component/formulation of interest
- Identify ongoing, blinded studies using the TS domain from ongoing ocean only, since they cannot be used to calculate exposure
- Create two more pooled SDTM domains (DM/SUPPDM, EX/SUPPEX), excluding data from studies identified above to reduce size
- Add variables in supplemental domains to those in parent ones

Calculating Planned and Actual Exposure (2)



- Attach recoded actual exposure information to single subjects using ACTARM/ACTARMCD
 - NB Subjects not exposed to any study medication are dropped
- Only data for subjects exposed to the vaccine(s) of interest are kept, and any double records from extension studies deleted
- The exposure numbers can then be easily calculated and reported by study and sex/race/age group/country/etc.

This is the high level structure of the second program, sharing with the previous one the first six steps, not repeated here:

- Create a set of pooled SDTM domains (plus their supplemental domains, as applicable), keeping data only from exposed subjects (to either the desired vaccine component/formulation or comparator)
- Create a dataset containing the list of MedDRA terms to be used in the search
- Compile the list of all PTs (AEDECOD) included in the terms specified above according to the designated dictionary version
- Subset the AE domain by AEDECOD, then 'sync' SUPPAE to remove orphaned records

- It is then possible to use this information to define four different scenarios:
 - Only AEs of interest in subjects exposed the desired vaccine(s)
 - All AEs reported in subjects exposed to the desired vaccine(s) and reporting at least one AE of interest
 - Only AEs of interest in all subjects exposed to a study vaccine in studies where at least one vaccination arm contains the desired vaccine(s)
 - All AEs in all subjects exposed to a study vaccine in studies where at least one vaccination arm contains the desired vaccine(s)



Data Visualization and CDR

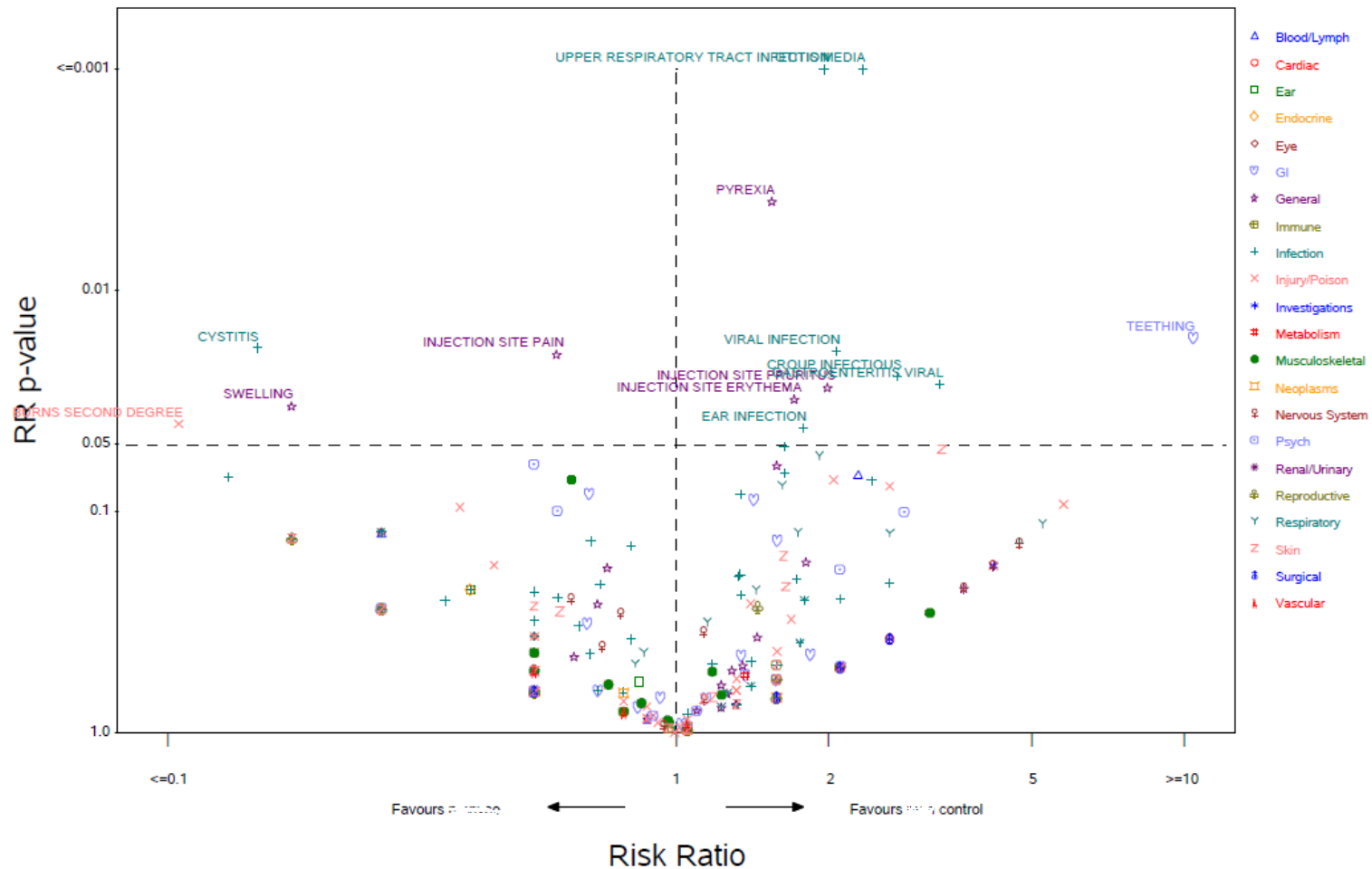
- Graphical methods are widely regarded as an excellent way to both explore data for unknown/unexpected correlations as well as analyze large quantities of them together
- Given the number of studies now in CDR this is for us a logical step, which we envisioned in our plans from the very beginning anyway
- I had actually hoped to devote a sizeable amount of time describing our experiences with some commercially available data visualization packages (e.g., JMP Clinical, Spotfire, JReview, etc.) in conjunction with CDR, but technical and contractual issues have made it impossible to do it in time for this paper
- The one aspect linked to graphics that we were able to investigate in my team, even if still to a much lesser extent than initially planned, was that of statistical techniques aimed at safety signal detection

- Several SAS programs were developed, aiming at presenting safety data in an aggregated way, e.g., using volcano plots like the one on the next slide, or forest plots
- Research into the use of Bayesian methods to perform safety signaling is also underway
- An internal document was recently finalized, detailing some of the principles we want to follow in these exercises
- Work is still ongoing on these topics, and its developments might be the topic of another paper sometimes in the future

Safety Signaling: A Volcano Plot



P-risk (Risk Ratio) Plot of Treatment Emergent Adverse Events at PT Level



Conclusions

-
- Building on the experience accumulated during the 2009 flu pandemic, it was decided by our management that all electronically available data from selected studies should be retrieved, remapped and stored in one physical location using common, non-proprietary database structures
 - Once that goal was achieved and data from more than two hundred studies had been pooled it became clear that new opportunities to exploit the amassed data had become suddenly feasible
 - When I started playing around with them, showing some early results to colleagues who were not previously considered as possible clients, they immediately expressed a keen interest

-
- One group of such clients is in Pharmacovigilance, and so we set up a small team to investigate any unmet needs
 - The two SAS programs I described in detail above, as well as other ones I only alluded to, are the initial answers to these needs:
 - Even if still quite rough, the results they produce were very well received, and have shown great promise in reducing the time needed to perform repetitive task, like gathering data for Risk Management Plan (RMP) updates or revising IBs

-
- EU definition of Pharmacovigilance:
http://ec.europa.eu/health/human-use/pharmacovigilance/index_en.htm
 - Data oceans in CDR:
<http://www.phusewiki.org/docs/Conference 2014 DH Papers/DH10 .pdf>
 - Volcano plots: [https://en.wikipedia.org/wiki/Volcano_plot_\(statistics\)](https://en.wikipedia.org/wiki/Volcano_plot_(statistics))

Questions?



Pantaleo Nacci
GSK Vaccines S.r.l.
Via Fiorentina, 1
53100 Siena, Italy
Email: pantaleo.x.nacci@gsk.com

