

## PhUSE 2017

### Paper DH04

# EMA Policy 0070: Data Utility in Anonymised Clinical Study Reports (CSRs)

Jean-Marc Ferran, Qualiance ApS, Copenhagen, Denmark

Sarah Nevitt, University of Liverpool, Liverpool, Great Britain

## ABSTRACT

EMA Policy 0070 “Phase 1” has been effective since January 2016 and requires CSRs and other regulatory documents from central applications to be published in an anonymised format.

While sponsors are favouring eliminating or minimising risk of re-identification of patients in data made public, EMA Policy 0070 requires sponsors to demonstrate that they have prioritised data utility in the anonymised CSRs published within the scope of the policy.

The concept of data utility is not clearly defined in the associated anonymisation guidance and can be interpreted from the perspective of different data consumers with different needs and level of skills to interpret regulatory clinical documents.

This paper will elaborate on the concept of data utility as outlined in EMA Policy 0070 and explore how different stakeholders (researchers, patients, health professionals, etc.) may use this wealth of information that has recently been made available in the context of central applications.

# PhUSE 2017

## DEFINITIONS

Abbreviation /Term	Definition
AE	Adverse Event
CCI	Company Confidential Information
ClinicalTrials.gov	US Clinical Trials Database
CSR	Clinical Study Report
DIA	Drug Information Association
EMA	European Medical Agency
EU	European Union
EudraCT	European Clinical Trials Database
FDA	Food and Drug Administration
HTA	Health Technology Assessment
IPD	Individual Patient Data
LPLV	Last Patient Last Visit
PII	Personal Identifying Information
SAE	Serious Adverse Event

## INTRODUCTION

Clinical Study Reports (CSRs) represent a wealth of information related to design, conduct and analysis of trials, in addition to more comprehensive trial results compared to publicly available databases such as journal manuscripts of clinical trials and clinical trial registries. Doshi 2013 [1] refers to CSRs as an “hitherto mostly hidden and untapped source of detailed and exhaustive data on each trial.” and addresses the concept of “compression factor” defined as the ratio of CSR page length compared to the page length of the same trial as published in scientific journals which ranged from 1 up to 8,805 based on the review of 78 CSRs.

Previous work has highlighted the impact of selective outcome reporting [2, 3], in that the data and results published within a journal manuscript may be incomplete or misleading, and the biases that originate from this selective reporting. Increasingly, researchers undertaking secondary analyses of clinical trial data such as systematic reviews and meta-analyses are seeking access to previously confidential regulatory documents as a means of assessing and reducing the impact of any selective reporting bias, generating more complete and reliable information and investigating clinical questions which could not have previously been considered using published data sources alone [4 - 15].

In the context of Policy 0070 “Phase 1”, where anonymised CSRs are made public, a myriad of data recipient groups could be considered together with various objectives for reviewing and using the information within these anonymised CSRs. The EMA external guidance [16] does not clearly identify these different data consumer groups and their intentions or objectives. However, from an anonymisation perspective, such considerations could help to define better anonymisation approaches, in addition to ensuring that adequate data utility for the purposes of the data recipients is retained.

The concept of data utility appears to be an important criterion for the anonymised CSRs to meet the objectives of EMA Policy 0070 “Phase 1”. Reference is made in several sections of EMA Policy 0070 CSR anonymisation external guidance and in its Anonymisation Report template that the data controller must demonstrate that data utility has been considered and optimised.

Despite the apparent importance of this criteria, the EMA external guidance does not define or quantify the utility of anonymised CSRs and “Data Utility” is also absent within section “3. Definitions” of the EMA external guidance. A definition from the Organization for Economic Co-operation and Development (OECD) is: “A summary term describing the value of a given data release as an analytical resource. This comprises the data’s analytical completeness and its analytical validity. Disclosure control methods usually have an adverse effect on data utility. Ideally, the goal of any disclosure control regime should be to maximise data utility whilst minimising disclosure risk. In practice disclosure control decisions are a trade-off between utility and disclosure risk.” [17].

For any public data release, there are intended data consumers and other data consumers (not primarily intended) that may also benefit from the data. The objective of this paper is to itemise who and how anonymised CSR data could be used by intended data consumers by reviewing possible data consumer groups, their purposes and the potential data utility associated with their purpose.

The research presented in this paper was conducted between February and October 2017.

It is outside the scope of this paper to discuss the redaction of Company Confidential Information (CCI) and its implication on data utility and not intended data consumers such as hackers or bodies following non-scientific purpose.

## PhUSE 2017

### INSIGHTS FROM POLICY 0043 REQUESTS

EMA Policy on access to documents (related to medicinal products for human and veterinary use) also known as “Policy 0043” has been effective since 2010 and enables individuals and parties to request access to documents. The scope of Policy 0043 is not limited to regulatory documents such as CSRs but can provide some insights on how much data has been requested by different bodies. Requesting documents through Policy 0043 is a controlled process and requesters must identify themselves as part of the request process that also includes a possible appeal in case of a rejection. The purpose of the request is however not required to be documented as part of the request. Table 1 below summarises the number of requests and number of pages released per affiliation and was made public in 2016 [0]. A breakdown of the information within Table 1 relating to the type of documents requested was not available.

**Table 1 – Policy 0043 documents requests per affiliation**

#### ***Affiliation (per initial requests and appeals in 2016)***

Affiliation	Number of requests received	In %	Number of pages released <sup>6</sup>	In %
Not-for-profit organisation	7	0.17	642	0.17
EU Institution (EC etc)	1	0.12	139	0.04
Regulator outside EU	2	0.24	0	0.00
EU NCA	4	0.49	103	0.03
Patients or Consumer	55	6.68	36388	9.55
Healthcare professional	24	2.92	16294	4.28
Academia/Research institute	66	8.02	120323	31.59
Legal	91	11.06	38463	10.10
Media	38	4.62	5960	1.56
Pharmaceutical industry	449	54.56	148013	38.86
Consultant	86	10.45	13044	3.42
Other	n/a	0.00	1542	0.40
<b>Total</b>	<b>823</b>	<b>100</b>	<b>380,911</b>	<b>100</b>

We, Jean-Marc Ferran (JMF) and Sarah Nevitt (SN), had an opportunity to discuss the data presented in Table 1 with EMA representatives and they confirmed that requesters under “Legal” and “Consultant” are mostly professionals from or contracted by the pharmaceutical industry. The category “Pharmaceutical industry” both includes companies from the innovative and generic industry. No distinction was made.

Grouping together the first 7 rows from “Not-for-profit organisation” to “Academia/Research institute”, the number of requests received represent 18.6% while “Legal”, “Pharmaceutical Industry” and “Consultant” represent 76% of requests. However, when comparing the number of pages released, the two groups are approximately equivalent with 52.3% and 45.6% respectively. This shift is mainly explained by the large number of pages in regulatory documents that are often requested and released to Academia/Research institutes (31.59%) while other subgroups are more likely to request other documents (e.g. meeting minutes) that represent few pages. It is indeed possible to

## PhUSE 2017

request a set of documents (e.g. all CSRs used in a submission) within one single request. It must also be noted that requests from "Patients or Consumers" represent 6.68% of all requests and 9.55% of pages released.

It may not be possible to directly extrapolate these figures from Policy 0043 (a controlled process covering all types of documents) to Policy 0070 (public access to regulatory documents part of a central application), but the affiliations described within Table 1 provide an indication of who the data consumers of Policy 0070 may be. While the Pharmaceutical Industry themselves and Academia / Research Institutes are likely to be the main data consumer of Policy 0070 documents, it cannot be excluded that "highly literate" patients or patient representatives may also benefit from such data being made public (almost 10% of all pages released in 2016).

NOTE: We (JMF and SN) were provided with the data in Table 1 and further interpretations from discussions with EMA representatives towards the end of the present research that was conducted and is summarised in this paper. Further research on Policy 0043 requests, particularly relating to the types of documents requested by each data consumer group could help understanding better similar concepts in Policy 0070.

### DATA CONSUMER SCENARIOS

Our first consideration during this research focused on determining data consumer groups that may use Anonymised CSRs and their purpose, noting that documents are complex and lengthy and require expertise to read, understand and process.

Conceptually, we considered two scenarios. The first scenario is built on the assumption that only highly skilled individuals or groups may invest time and efforts in working with such documents and that other potential data consumers would benefit from the findings of this highly skilled group and indirectly consume anonymised CSRs. The second scenario assumes that various data consumer groups would have the skills, educate themselves or contract other skilled professionals to consume these documents for various purposes.

The following subsections describe in more detail the subgroups and mechanism in terms of how knowledge can be derived and who are the primary and secondary data consumers on anonymised CSRs made public.

#### SCENARIO 1: SIMPLIFIED "SKILL-BASED" SCENARIO

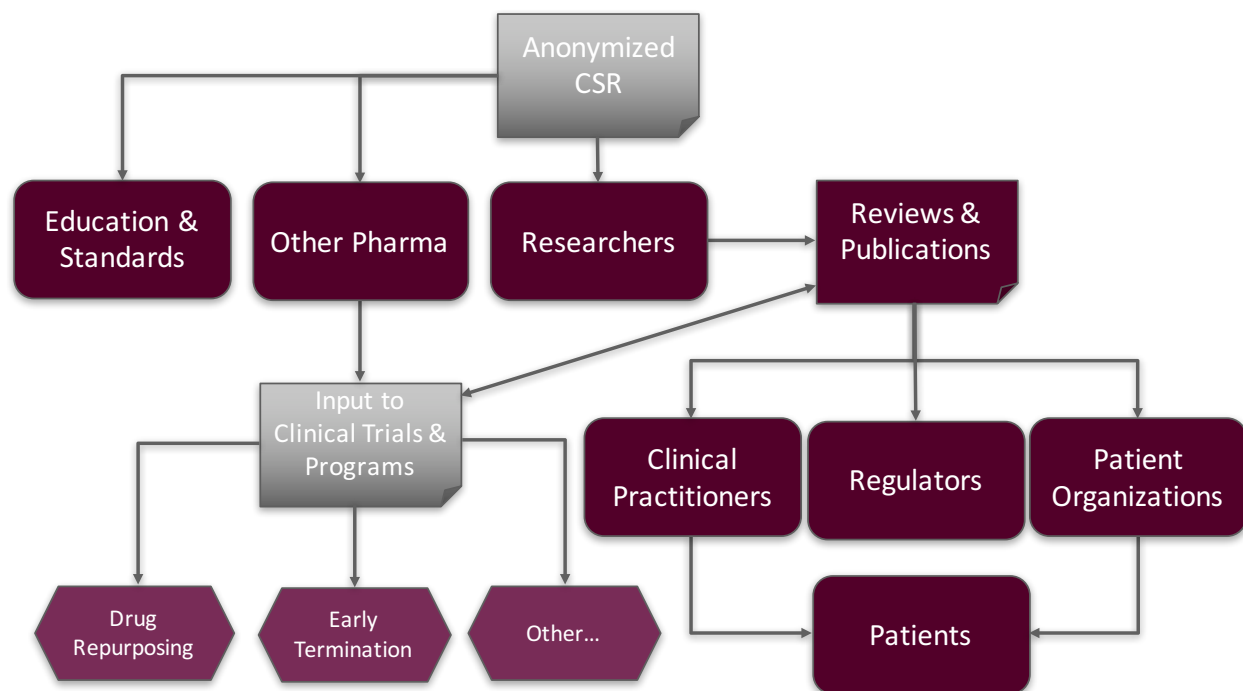


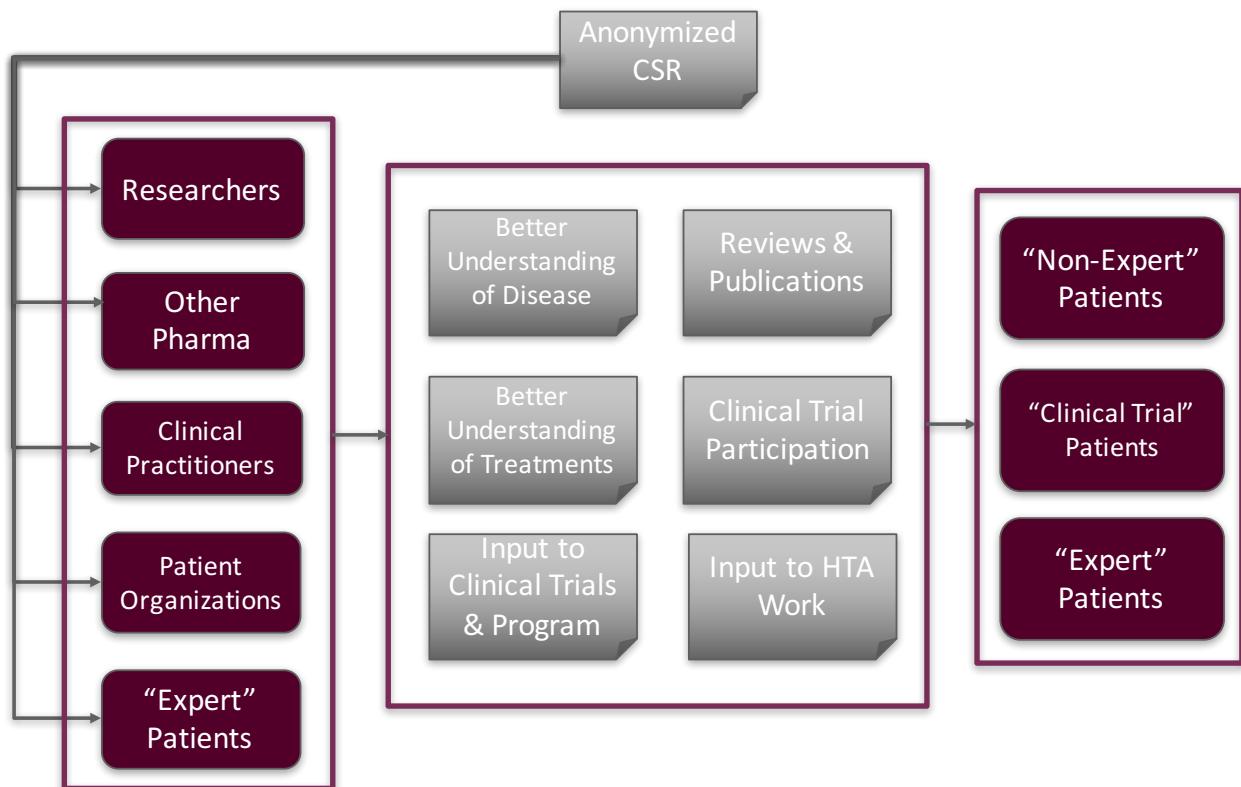
Figure 1 - Simplified "Skill-based" Scenario Flow

## PhUSE 2017

Figure 1 represents the first scenario where only highly skilled individuals and bodies would use anonymised CSRs. Researchers would conduct novel or secondary analyses such as systematic reviews and meta-analysis to derive knowledge that would be published and disseminated to Clinical Practitioners and Patient Associations who will then translate the findings to patients.

Other Pharmaceuticals would be able to consume the documents relating to similar compounds in order to inform their clinical trials and clinical programs. In the particular case of drug repurposing (i.e. the application of existing compounds to new indications), access to CSRs from previous studies is key. Also, within the scenario of the rejection or withdrawal of a central application of a competitor, access to CSRs could provide valuable lessons for further submissions or even lead to re-investment of resources within an alternative, perhaps more promising, clinical program. Similarly, other Pharmaceuticals and Regulators may also benefit from published novel research conducted by researchers, who in turn, may benefit from advances in clinical trial design and clinical programs conducted within the Pharmaceutical Industry.

### SCENARIO 2: VERSATILE USE ACROSS DATA CONSUMERS



**Figure 2 – Versatile use across Data Consumers Scenario Flow**

In a second scenario illustrated in Figure 2, patients would form three group: “Expert” Patients, “Clinical Trial” Patients and “Non-Expert” Patients. Policy 0043 data shows that “Patients or Consumers” have requested nearly 10% of all pages released in 2016 and it is assumed here that “Expert” Patients who are highly literate will be using Anonymised CSRs to better understand their disease and available treatments. In this scenario, Clinical Practitioners and Patient Associations would also use such reports to also advise on clinical trial participation, clinical trial designs and use in HTA process. On the right-hand side, the three different types of patients (that may overlap, e.g. a patient is participating in a clinical trial and is highly literate about the diseases) would benefit directly or indirectly from all the knowledge derived by other bodies.

The paper will explore how plausible these two scenarios are through review of potential uses of the anonymised CSRs across data consumer groups.

## RESEARCHERS AND RESEARCH GROUPS

From previous work which has made use of regulatory documents for secondary analyses [4 -15], it is expected that the research community, such as research groups comprised of academic and clinical researchers, and medical statisticians, will be the primary users of the anonymised CSRs.

A short summary [18] of advantages and disadvantages in using CSRs versus published articles for Cochrane review highlights the extra information and opportunities as well as challenges working with CSRs for academics.

Jefferson [19] refers to a survey conducted among Cochrane authors (31,901 authors – not all active) between June and September 2016 about their experience (if any) using regulatory data (defined as CSRs and other regulatory documents). Initial results of the survey showed that among 156 respondents that only 10% used or requested regulatory data, 5% considered using regulatory data and 85% have not considered using regulatory data. Among the 10% of respondents who have used or requested regulatory data, 80% believes that regulatory data should be used on Cochrane reviews while this number falls to 38% and 32% for the ones who considered regulatory data and the ones who have not considered regulatory data. Source of data requests that were reported were pharma (11) and regulators (7). CSRs were used by 8 authors to supplement published trial data in meta-analyses. But 67% of respondents who accessed and included data in their reviews mentioned barriers when using data: limited data sharing/restricted access and expertise and skills required. Other interesting data shows that 32% of respondents had no understanding of the regulatory process and documents developed and only 12% of authors knew where to access regulatory data. Further results of this survey will soon be published [20].

These results demonstrate that using regulatory data is rather new for the academic community but researchers who are requesting or using regulatory documents to complete Cochrane Reviews consider access to these documents important and values for their analyses. The field is merely a “new born” and more examples of how regulatory data for secondary analyses can be used should emerge in the coming years.

Results of the survey also suggested that the availability of further guidance on how to interpret and use regulatory data in secondary analyses would help to promote the use of CSRs in Cochrane Reviews [20,21]. An “*Interim guidance on the inclusion of Clinical Study Reports and other regulatory documents in Cochrane Reviews*” is being developed.

Researchers and research groups may wish to use the anonymised CSRs for a range of different research purposes and using variety of techniques and methods. Some of the research purposes are outlined in this section; in general terms, these purposes fall into two categories, data and method comparison and novel analyses, although some research purposes may overlap these categories.

### DATA AND METHOD COMPARISON: DETAILED ASSESSMENT OF STUDY METHODS AND EVALUATION OF RISK OF BIAS

A potential purpose of gaining access to complete information regarding trial design and conduct may be used by researchers to inform the design of future trials.

A more common reason for using CSRs, as discussed in the introduction of this paper, to access to detailed trial methodology and comprehensive results within anonymised CSRs allows for assessment of bias in the trial design and any selective outcome reporting bias in journal manuscripts [2, 3]. For example, by comparing the information in CSRs and other regulatory documents to published trial reports for 20 trials of Gabapentin, Vedula *et al* [15] identified selective outcome reporting for trials of off-label use of gabapentin which threatens the validity of evidence for the effectiveness of off-label interventions. Eyding *et al* [4] also discovered that published data overestimated the benefit of reboxetine versus placebo by up to 115% and reboxetine versus SSRIs by up to 23%, and also underestimated harm compared to the information presented in CSRs and other regulatory documents.

Similarly, by comparing publications of Orlistat trials to their corresponding CSRs, both Scroll *et al* [9] and Hodkinson *et al* [10] identified that journal publications provided insufficient information on harms outcomes of clinical trials and in some cases, inconsistent numbers of adverse events were reported across different documents relating to the same trial. Further examples of research using CSRs and other regulatory documents can be found in Appendix 1.

### DATA AND METHOD COMPARISON / NOVEL ANALYSIS: USE OF NARRATIVES, INDIVIDUAL PARTICIPANT LISTINGS & EVALUATION OF HARM

Summary of AEs are required to be made public in ClinicalTrials.gov and EudraCT. However, a number of examples of under-reporting or misleading reporting of harms in publically available reports of clinical trials compared to the more detailed harms information available in CSRs have been published (see Appendix 1).

## PhUSE 2017

These publications highlight that review of individual narratives and participant listings (Note that participant listings are out-of-scope of Policy 0070 “Phase 1”) for Serious Adverse Events and other Adverse of Events of interest would help researchers understanding further the safety profile of the drug beyond review of summary statistics of dictionary-coded events and verify safety data in CSRs. We explore here 2 publications from Emma Maund and colleagues [13-14].

Maund 2014 [13] is a methodological paper relating to different conclusions that can be drawn from reading summary tables of AEs which are usually dictionary coded, compared to reading verbatim descriptions in the narratives. Their illustrative example shows that coded events and narratives suggest different numbers of events related to suicide and the authors conclude that in this case the narratives are more informative and coded events in summary tables may be misleading and may not fully capture the true nature of the event. The paper states that *“narratives of adverse events can provide additional information, including original investigator reported adverse event terms, which can enable a more accurate estimate of harms”*. *“Using the patient’s trial identification number we were able to reconcile data reported in the patient listings with those in the narrative. Secondly, using data (treatment assignment, coded term, and timing of event) from the patient listings and narratives, we were able to reconcile data from these two formats with the data in summary tables.”*

Maund 2016 [14] is a clinical paper (a meta-analysis) relating to the benefit and harms of a drug for a particular indication. The authors are interested in some specific harms, related to suicide and violence, due to FDA concerns about the association between this drug and these SAEs. The authors use data from CSRs (summary tables and narratives) to perform their analyses and the discussion compares the results of this analysis to a Cochrane review of the same topic conducted only with data available in the public domain (i.e. from trial publications). The two analyses come to overall the same conclusions (i.e. that clinical benefits of the drug in question do not outweigh the potential harms) but the two analyses consider different outcomes sets and show slightly different results for common outcomes. Notably, the analysis of CSR data has allowed more detailed considerations of specific adverse events which would not have been possible without access to CSRs and narratives. For example, in Maund 2016 [14], *“one patient had a “nervous breakdown,” which was coded as mental disorder, and another patient reported “feeling drugged,” which was coded as somnolence. In addition, 5 patients, all receiving duloxetine, experienced a total of 8 events that were mentioned only in the narrative text.”*

The ability to follow a patient through narratives, conserve sequence and distance between events, findings and interventions and access the investigator reported terms seem essential for the work described above. Availability of demographics and medical history among others could also support more detailed analyses.

### **NOVEL ANALYSIS: USE OF UNPUBLISHED SUMMARY DATA FOR SYSTEMATIC REVIEWS AND META-ANALYSES**

Researchers performing systematic reviews and meta-analyses, such as Cochrane Reviews, can take one of two approaches; either using summary (aggregate) data only or performing a re-analysis of individual participant data (IPD). Using IPD has many advantages for meta-analysis and allows more complex research questions to be considered. Researchers can request access to IPD from pharmaceutical trials via data sharing platforms such as CSDR [22] and YODA [23]. However, an IPD approach to meta-analysis is very resource and time consuming and may not be necessary if published and unpublished aggregate data can answer the clinical question [24].

Therefore, to increase the precision and reliability of systematic review and meta-analysis results, researchers may wish to use unpublished summary (aggregated) data from one or more CSRs within systematic reviews and meta-analyses. In 2014, Jefferson et al [15] reported on the first Cochrane review to be based on all relevant full clinical study reports of a drug, augmented by regulatory comments.

Summary statistics and details of the statistical analyses of primary and secondary efficacy endpoints are required to be made public in ClinicalTrials.gov and EudraCT 30 days after submission approval and 6 (paediatric trial) to 12 (adult-only trial) months after Last Patient Last Visit (LPLV) for any trial conducted in Europe respectively and would be available in such registries outside an EMA central application. However, previous work has shown that such publically available information may not be sufficient [5], particularly for the objectives of an original systematic review or meta-analysis. Additionally, the format that the summary results are provided in may not readily allow the inclusion of the information within meta-analysis (for example, where a measure of precision of the treatment effect is not published).

A CSR would typically contain more details about the choice of statistical method, interpretation of results and the full set of endpoints, results and statistics at all time points measured and therefore may provide a useful supplementary source of data for systematic reviews and meta-analysis. For example, a review of 101 CSRs conducted by Wieseler et al [5] shows that CSRs provided complete information on 78% to 100% of benefit outcomes (compared to 20% to 53% from publically available sources), CSRs also provided considerably more information on harms and on patient-relevant outcomes such as outcomes describing morbidity, mortality, and health-related quality of life (HRQoL). This research group emphasise that it is essential that sufficient information is available to patients and clinicians on



## PhUSE 2017

benefits, harms and patient-relevant outcomes when new drugs become available [6,7].

We describe two examples of projects that Sarah Nevitt (née Nolan) has been involved in to demonstrate the rationale and reasons for conducting an IPD analysis compared to using unpublished summary data (e.g. from CSRs) in secondary analysis. Both projects were conducted as Cochrane Reviews, but the level of analysis involved was very different.

The first project is an IPD meta-analysis of Epilepsy trials [26]. The primary outcome of this analysis is time-to-treatment withdrawal which is a complex outcome and often reported differently across studies, so to perform a meta-analysis, the definition of this outcome had to be standardised. Often, free text or 'verbatim' reasons for an individual withdrawing from treatment were required within the IPD, to ensure the correct classification of reasons for withdrawal within the re-analysis of the IPD. Furthermore, an objective of this analysis was to examine differences in treatment response across different patient subgroups (such as different epilepsy types and different ages). Such information will never be available to the required level in the public domain, so IPD and statistical expertise to perform the appropriate analysis to address this question is required.

The gain in knowledge provided by IPD in this analysis allowed for two drugs which were considered to be approximately equivalent in terms of their effectiveness as anti-epileptics, to be separated in a number of respects. For example, one of the drugs was shown to have an advantage over the other in terms of effectiveness (i.e. time to treatment withdrawal), while the opposite effect was shown when considering only the efficacy of the drugs in controlling seizures (i.e. time to 12-month remission of seizures). Such findings allow a more personalised approach medical decision making to specific subgroups of patients depending on their priorities and requirements of an anti-epileptic medication.

The second project is a Cochrane review of Mannitol (a new drug) for Cystic Fibrosis [25]. Regulatory objectives within clinical trials often focus on efficacy of new compounds (e.g. lung function within Cystic Fibrosis) as demonstration of efficacy over placebo or a standard treatment is usually required for a new product to be licenced. However, the focus of the Cochrane Collaboration is often around patient important outcomes such as Quality of Life or Burden of treatment which are often more meaningful indicators of clinical status and improvement to patients.

While the main published sources of new regulatory trials (journal articles, Clinical trials.gov etc.) focus on efficacy outcomes such as lung function, very little information was reported regarding on Quality of Life or other patient reported outcomes. Therefore, the manufacturer of Mannitol (Pharmaxis), were contacted by SN and the Cochrane Review team and the manufacturer provided unpublished summary data (aggregate data), allowing for all of the patient reported outcomes relevant to the Cochrane Review to be reported in a great level of detail. To clarify, it was not CSRs specifically that were made available by the manufacturers, but the level of information required for the Cochrane review would have certainly been available in CSR.

The unpublished data was then synthesised within the Cochrane review using 'standard' meta-analysis methodology which would not necessarily require the support of a statistician. SN became involved in the project due to her experience of requesting unpublished regulatory data, but if this information had been available in the public domain, a statistician's expertise would not have been needed for this project.

### EXAMPLES OF USE OF CSRS BY ACADEMIC RESEARCHERS

In a presentation given at DIA 2017 by Tom Jefferson in Glasgow [19], the following journal publications were listed as examples of academic work using CSRs in secondary research:

- Eyding *et al* 2010 [4]
- Le Noury *et al* 2015 [8]
- Schroll *et al* 2016 [9]
- Hodkinson *et al* 2016 [10]
- Jefferson *et al* 2014 [11]
- Maund *et al* 2014 [13]
- Vedula *et al* 2009 [15]

A summary of the regulatory data sources, methods and conclusions of these manuscripts is provided in Appendix 1. These manuscripts are based on CSRs that were gathered before the implementation of EMA Policy 0070. Most of these manuscripts tend to demonstrate publication bias, reporting inconsistencies and/or provide new knowledge about drug efficiency and benefit/harm ratio using meta-analyses.

We are providing this summary from a methodological perspective on use of regulatory data in academic research. It must be noted that this is a 'selective sample' of academic work which has mostly shown changes in conclusions, particularly regarding harms of drugs, when re-analysing clinical trial data using CSRs and the re-analysis

## PhUSE 2017

approaches taken by some of the academic research groups have been challenged by the pharmaceutical companies in questions and response articles published on journal websites, linked to the academic work in question. This sample should not be considered completely reflective of academic objectives for accessing regulatory data or a comprehensive list of all research using CSRs (which is likely much wider as indicated by the number of requests from Table 1). The selective nature of the sample summarised here must be taken into account when interpreting the findings in the context of all published research making use of unpublished regulatory information.

We (JMF and SN) attempted to make contact with the authors of the publications listed above to further explore the use of CSRs in the projects and the potential impact that EMA Policy 0070 may have on the data utility of anonymised CSRs in secondary research. Full details of the correspondence with authors is provided in Appendix 2.

In summary, all of the authors stated that their analyses would not have been possible without access to CSRs. None of the authors raised any specific concerns about anonymised or redacted CSRs (in line with EMA Policy 0070). In fact, one research team had used CSRs publically available from a sponsor website which were redacted and this redaction did not impact upon the analysis from the author's recollection. Furthermore, none of the authors stated that their team had any difficulties in interpreting the information from the CSRs; the only problems related to 'illegible' text or the format of the documents which prevented electronic searching.

All of the authors stated that some or all of their analyses or research would not have been possible if narratives and/or appendices (with participant listings) were removed from anonymised CSRs under EMA Policy 0070. One author stated that: "Anonymised CSRs are ok, but the current EMA policy redacts important information about when the adverse events appeared as well as what they were. Newer CSRs does not have individual adverse event listings and the EMA are not even in possession of them." Another author with knowledge of Policy 0070 stated that: "I have actually looked at data that are released under the EMAs new policy 0070, and they do provide fully redacted CSRs. So yes, I would say you could use these provided the drug is centrally licenced. But redactions may permit what data can be used, and they may not be of use for creating IPD datasets without the subject IDs and other patient-level information."

It should be emphasised that these observations are anecdotal and rhetorical as these projects were based on CSRs that were obtained before the implementation of EMA Policy 0070. However, these observations and the rationales of the type of analyses being conducted using CSRs do raise some potential issues relating to data utility of documents anonymised under EMA Policy 0070 "Phase 1". The full extent and any impact of such issues will not become apparent until sufficient research projects are conducted and published using anonymised CSRs prepared in line with EMA Policy 0070.

### COMPETITORS / OTHER PHARMACEUTICALS

As outlined conceptually in Figure 1, other pharmaceuticals (including competitors) would be able to consume the documents relating to similar compounds in order to inform their clinical trials and clinical programs. Competitors can gain knowledge about similar drugs and design better their studies.

Gaining insight into similar compounds' clinical trial results may also help competitors to stop "bound-to-fail" clinical programs earlier and reinvest resources on promising drugs. In the scenario of the rejection or withdrawal of a central application of a competitor, access to CSRs could provide valuable lessons for further similar submissions. The case of drug repurposing also require access to all possible data from previous studies in other indications.

Bonini et al. 2014 [30] refers to "optimising future study designs with regard to population selection and sample size, choice of outcomes, definition of clinically relevant differences for various end points, or identification of biomarkers for better disease phenotyping". Policy 0070 "Phase 1" also provide access to data and information on more endpoints, subgroup, full set of analyses on same endpoints in comparison with EudraCT and ClinicalTrials.GOV. European citizens can already request CSRs through EMA Policy 0043 but this is a lengthier process (up to several months).

### CLINICAL PRACTITIONERS

As outlined conceptually in Figure 2, clinicians such as general practitioners, nurse specialists, consultants etc. may wish to review details within CSRs not published in the public domain and understand better the safety and efficacy profile of the drugs they prescribe and to better inform their patients on drug choices and clinical trial participation.

At the time of the draft EMA Policy 0070 being on review, several of the comments came from practising clinical practitioners but this was only a small number of the people who could have actually commented.

## PhUSE 2017

The comments at the time were all supportive of the policy and referred to recent examples in the media where initial data conclusion, and drug approvals, were overturned when additional data was further analysed.

In recent follow up with a sample (n=5) of these practitioners they reflected on their reasons for submitting comments and that they were statements of support. Furthermore, they confirmed that they have no directly interest in producing research and meta-analysis to investigate data reports but rather want reassurances that researchers are able to access reports and data to continue to produce additional analyses.

### PATIENTS / PATIENTS' ASSOCIATIONS

Lay Summaries are probably more accessible to Patients but it cannot be excluded that “Highly Literate and Expert” patients (typically in chronic diseases) would be able to make use of anonymised CSRs and have had the opportunity so far to request relevant CSRs through Policy 0043. As part of this research, we have not been able to get further insight on how Patients would use anonymised CSRs.

A review of the comments sent from patients' and consumers' associations on the initial draft version of EMA Policy 0070 in 2013 (also covering sharing of IPD, known now as “Part II” and not finalised yet) [27] shows that:

- The initiative is received very positively
- There are concerns around current Informed Consent forms that patients signed
- There are concerns around possibility of wrong secondary analyses
- They do not believe there are CCI in clinical trials data
- Patients privacy is of utmost importance

A number of patients' & consumers' organisations have also been contacted to provide input:

- BEUC (<http://www.beuc.eu/>): Participated in an interview
- ECPC (<http://www.ecpc.org/>): No answer
- EURODIS (<http://www.eurodis.org/>): Participated in an interview.
- Genetic Alliance (<http://www.geneticalliance.org/>): Participated in an interview.
- EPF (<http://www.eu-patient.eu/>): Answered it was too early to reflect on how patients would be using these reports but would like to contribute in the future.

Organisation “Understanding Patient Data” (<https://understandingpatientdata.org.uk/>) was also contacted and communicated that there was no plan for the moment to develop guidelines for patients to help them understand and use CSRs.

BEUC (The European Consumer Organisation) is an umbrella organisation for EU consumers association. We spoke with a representative from the Health& Food department who underline the importance of making such data available in the public domain to ensure transparency of the information flow from a consumer's perspective and increase trust in regulators.

EURODIS and Genetic Alliance represent patients with rare diseases. These associations also contribute to clinical research and public affairs.

According to a representative of EURODIS, CSRs provide much more detailed data compared to public registries. However:

- Data from Phase I/II are essential to recommend rare disease patients which clinical studies to join in the lack of a variety of available treatments. This would require phase I/II clinical studies CSRs to be available early while Policy 0070 “Phase 1” makes the submissions of anonymised CSRs at the time of the drug submission and would not address this need.
- Placebo data can be used to understand better the disease from this controlled and carefully monitored population.
- CSRs may help to inform better Academic Strategic Clinical Trials aiming at learning when to start and how to use better the available treatments.
- Indirect comparison of drugs could be supported
- CSRs could be used to support discussions between Community Advisory Board. (Patients representative interacting with sponsor on methods and logistics, etc.) and sponsors on Clinical Trials throughout the entire compound Life Cycle.
- In the case of drug repurposing (from e.g. a frequent to a rare disease area), having access to all previous data is of the utmost importance and would speed up the process.

## PhUSE 2017

- In general, rare diseases require other policies than Policy 0070 to address earlier and faster access to data including access to IPD.
- It was noted that patients with rare diseases considering the often genetic roots of the disease are concerned over privacy as data breach may also affect their relatives.

According to a representative of Genetic Alliance:

- CSR data can be used to build stronger cases in HTA work and provide stronger arguments towards investment and research.
- Patient Association giving evidence based on full range of treatments to a level of professionalism that matches what is produced by other stakeholders would be a significant step forward and help patients having a higher impact.
- Privacy may be a secondary concern for certain people who have rare disease in contrast with developing adequate treatments for their diseases but this opinion may only be UK specific.

## DISCUSSION

### SUMMARY OF FINDINGS AND IMPLICATIONS

CSRs have previously been considered as an 'untapped' source of detailed information relating to design, conduct and analysis of clinical trials. The value of the information within CSRs is becoming increasingly recognised within the academic research community, particularly within the Cochrane Collaboration and an *"Interim guidance on the inclusion of Clinical Study Reports and other regulatory documents in Cochrane Reviews"* is being developed.

It is already compulsory that aggregated data from primary and secondary outcomes of clinical trials is public in the EudraCT database within 6 to 12 months from LPLV for most studies conducted in Europe. While such data may be suitable and sufficient to support some secondary analyses such as meta-analysis, anonymised CSRs provides complete information and data on study and statistical methods, interpretation of results and the full set of endpoints' results and statistics and would certainly enable verification of numerical results and assessments and conduct meta-analyses using data from all endpoints.

EMA Policy 0070 "Phase 1", where anonymised CSRs are made public, is likely to further facilitate the secondary use of the information within CSRs. However, little consideration has been given to the data utility of the anonymised information within CSRs under Policy 0070. The objective of this paper was to identify data recipient groups with a range of purposes and to investigate the data utility of anonymised CSRs associated with various purposes.

Based on the number of requests made under EMA Policy 0043, we anticipate that researchers or research groups and the Pharmaceutical Industry are likely to be the primary recipients of anonymised CSRs under EMA Policy 0070 as described in Scenario 1: Simplified "Skill-based" Scenario. The research examples we discuss within this paper indicate that the objectives and scopes of secondary analysis and novel research that have been conducted using CSR data are vast. Authors of such research have communicated with us their concerns over the type of research that could be conducted in the future if information such as participant listings or narratives are redacted or removed completely under EMA Policy 0070.

It should be noted that not all data consumers would benefit from complex data available in a CSR and it could be justified to focus on most relevant data consumers in terms of data utility if not all needs can be met. In reality, it is likely that clinical practitioners, Patient Associations and patients themselves will indirectly gain knowledge from anonymised CSRs via the disseminated published findings of new research rather than these consumers using regulatory documents directly. However, it cannot be excluded that as in Scenario 2: Versatile Use across Data Consumers, "highly literate" patients or patient representatives would wish to access anonymised CSRs released under Policy 0070; almost 10% of all pages released in 2016 under EMA Policy 0043 were requested by patients or consumers.

Keeping similar conclusions and primary and secondary analyses in the Anonymised CSRs similar to the ones available in the Scientific CSR is of utmost importance. Handling of narratives seems to be the most difficult aspect of the policy from a technical standpoint and various levels of anonymisation would define further different levels of data utility together with the handling of in-text listings. There are examples in the literature on how narratives are used to verify safety conclusions (see Appendix 1 for examples).

Certain free-text fields such as e.g. Adverse Events Reported Terms or Reason for Withdrawal may be instrumental for certain secondary analysis to e.g. verify dictionary coding and conduct re-analysis [13] or map reason for withdrawal consistently across studies in the case of meta-analysis [26]. Further, preserving Subject IDs and Dates in an anonymised format would help using the narratives in particular in order to follow a patient throughout the Adverse

## PhUSE 2017

Events and use sequences and distances to understand. Free-text variables are often redacted in both IPD and documents when a dictionary-coded variable is available and generally better suited for analysis. The PhUSE De-Identification standard [31] recommends as primary rule in the case of pro-active release of data to follow such rational and a secondary rule to “Review and redact PII” in such free-text variables. It is therefore advised to researchers to make it clear in their requests to pharmaceutical companies whether certain free-text variables must be retained even though a dictionary-coded variable is available in the given data domain. In the case of public release of documents, should such free-text variables be retained, processing to ensure no PII remains prior to publication is essential.

Patient Associations raised the concern of the risk of wrong secondary analysis in their initial comments to Policy 0070 received in 2013 [27]. In the case of research request sent directly to pharmaceutical companies and for IPD requests made via ClinicalStudyDataRequest.com (or a single sponsor platform), a clause is present within Data Sharing Agreements specifying that in the case of detection of a new safety signal, the researcher must inform the regulatory authorities and pharmaceutical companies immediately. In the case of Policy 0070, there are no such expectations and any new findings (also related to new safety signals) are discussed through public academic debates.

Further understanding of the safety profile of the drugs and verification of how conclusions of clinical studies are derived is certainly an added value for many stakeholders and data consumer groups. However, several academic publications that were reviewed in the paper and described in Appendix 1 have had their findings challenged by concerned pharmaceutical companies through comments on journals web-sites. Discussion of academic findings and interpretations should always be encouraged but there is a risk that ‘rapid-response’ additional analyses as a challenge to published research may confuse readers and secondary data consumers such as clinical practitioners, patients and Patient Associations who cannot interpret which of the many published results are the correct ones. Bonini et al. 2014 [30] also mentions that “access to clinical data imposes a high ethical standard on anyone using those data, lest inappropriate reanalyses breed unjustified concern about the efficacy or safety of marketed drugs.” We (SN and JMF) suggest that communication between academic research groups and pharmaceutical companies regarding interpretations of regulatory data and results from their different perspectives during the research projects before publications within journals may provide the most informative novel results and in turn, provide the most benefit to readers and data consumers.

### LIMITATIONS

The work presented within this paper is on a small number of interviews and e-mail communications conducted with Researchers, Patient Associations, Doctor Associations rather than a systematic survey across a significant population of data consumers. Therefore the findings presented within this paper represent only tendencies to explore further. Particularly, certain data consumers, their purposes and any associated data utility such as “Other Pharmaceuticals”, “Generics” or “Regulators” should be investigated further.

It must be emphasised that the examples of academic research using CSRs summarised within this paper are a selective sample and do not necessarily represent all research objectives which would make use of anonymised CSRs under EMA Policy 0070. Further, most observations provided to us by data consumers and our interpretations are rhetorical, rather than based on direct experience of anonymised CSRs and the validity of these observations may not become clear for some time.

### FUTURE CONSIDERATIONS

“Phase 2” of EMA policy 0070 on sharing of IPD should provide the next level of data utility that is required to conduct robust secondary analyses. A number of sponsors already provide access to anonymised IPD based on research request for studies based on different criteria (e.g. EFPIA commitment [29]). “Phase 2” of the policy that is planned in the future should in principle systemise the access to anonymised IPD for studies part of a central application in EU regardless of the outcome of the application. The needs of the research community often include access to full patient listings which is out of scope of Policy 0070 “Phase 1” and may be addressed in “Phase 2” of the policy.

In addition more guidance on using regulatory documents is required to researchers, patients and anyone from the general public in order to make the best use of such publicly available data.

In conclusion, EMA guidance refers to various levels of anonymisation but based on level of risk of re-identification of patients rather than different levels of data utility and this paper can hopefully help to consider the anonymisation problem from both perspectives. This research field is “merely a new born” and more experience from use and findings based on Policy 0070 data should be reviewed in the future.

# PhUSE 2017

## ABOUT THE AUTHORS

### Jean-Marc Ferran, MSc.

Jean-Marc Ferran is an Independent Consultant based in Copenhagen with 15 years of experience in the Life Sciences industry. Prior to starting his company, Qualiance, he worked as a Statistician, Standards Manager and Director of Statistical Programming at Novo Nordisk and Ferring Pharmaceuticals. Jean-Marc leads the PhUSE Data Transparency Working Group and advises companies on how to implement Data Transparency initiatives. He has recently been appointed to EMA Technical Anonymisation Group and Health Canada Stakeholder Reference Group on Public Release of Clinical Information.

### Sarah Nevitt, PhD

Sarah Nevitt (née Nolan) is a Medical Statistician at the University of Liverpool. Sarah is the Statistical Editor of the Cochrane Epilepsy Group and her role also includes working within the Clinical Trials Unit at the University of Liverpool and performing Health Technology Assessment on behalf of the National Institute for Health and Care Excellence. Sarah has an active research interest in data sharing and data anonymisation and recently been awarded a PhD, thesis titled "Data sharing and transparency: the impact on evidence synthesis." Sarah is also a member of the PhUSE Data Transparency Working Group and has recently been appointed to EMA Technical Anonymisation Group.

## DECLARATION OF CONFLICT OF INTEREST

Jean-Marc Ferran: I have worked for various Pharmaceutical companies as a consultant since January 2010 on compound specific clinical projects or cross-compound system-related projects. I have also entered a strategic alliance in January 2017 with d-Wise Inc. (North Carolina) where I contribute as an SME to the design and development of their data de-identification products. The views expressed within this research are my own.

Sarah Nevitt: I have no known conflicts of interest. The views expressed within this research are my own and do not necessarily reflect the views of the University of Liverpool

## ACKNOWLEDGMENTS

Thanks to the Researchers, Patient Associations, Doctor Associations, Industry Representatives and Regulators, who engaged with us, for their time and invaluable input.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Jean-Marc Ferran

Qualiance ApS

Email: [JMF@qualiance.dk](mailto:JMF@qualiance.dk)

Web: <http://www.qualiance.dk>

Sarah Nevitt

University of Liverpool

Email: [Sarah.Nevitt@liverpool.ac.uk](mailto:Sarah.Nevitt@liverpool.ac.uk)

Web: <https://www.liverpool.ac.uk/translational-medicine/staff/sarah-nevitt/>

Brand and product names are trademarks of their respective companies.

## REFERENCES

- [0] Annexes to the annual report of the European Medicines Agency 2016, Annex 19: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Annual\\_report/2017/05/WC500227333.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Annual_report/2017/05/WC500227333.pdf)
- [1] Doshi P, Jefferson T. Clinical study reports of randomised controlled trials: an exploratory review of previously confidential industry reports. *BMJ Open* 2013;3(2):e002496. doi: 10.1136/bmjopen-2012-002496: <http://bmjopen.bmj.com/content/bmjopen/3/2/e002496.full.pdf>
- [2] Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *Bmj*. 2010 Jan 1;340:c365.
- [3] Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PloS one*. 2013 Jul 5;8(7):e66844.
- [4] Eyding D, Lelgemann M, Grouven U, et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ* 2010;341:c4737. doi: 10.1136/bmj.c4737
- [5] Wieseler B, Kerekes MF, Vervoelgyi V, McGauran N, Kaiser T. Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. *BMJ*. 2012 Jan 3;344:d8141.
- [6] Wieseler B, Wolfram N, McGauran N, et al. Completeness of Reporting of Patient-Relevant Clinical Trial Outcomes: Comparison of Unpublished Clinical Study Reports with Publicly Available Data. *PLOS Medicine* 2013;10(10):e1001526. doi: 10.1371/journal.pmed.1001526
- [7] Köhler M, Haag S, Biester K, Brockhaus AC, McGauran N, Grouven U, Kölsch H, Seay U, Hörn H, Moritz G, Staeck K. Information on new drugs at market entry: retrospective analysis of health technology assessment reports versus regulatory reports, journal publications, and registry reports. *BMJ*. 2015 Feb 26;350:h796.
- [8] Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320. doi: 10.1136/bmj.h4320
- [9] Schroll JB, Bero L, Gotzsche PC. Searching for unpublished data for Cochrane reviews: cross sectional study. *BMJ* 2013;346:f2231. doi: 10.1136/bmj.f2231
- [10] Hodkinson A, Gamble C, Smith CT. Reporting of harms outcomes: a comparison of journal publications with unpublished clinical study reports of orlistat trials. *Trials* 2016;17(1):1-11. doi: 10.1186/s13063-016-1327-z
- [11] Jefferson T, Jones M, Doshi P, Spencer EA, Onakpoya I, Heneghan CJ. Oseltamivir for influenza in adults and children: systematic review of clinical study reports and summary of regulatory comments. *Bmj*. 2014 Apr 9;348:g2545.
- [12] Maund E, Tendal B, Hróbjartsson A, Jørgensen KJ, Lundh A, Schroll J, Gøtzsche PC. Benefits and harms in clinical trials of duloxetine for treatment of major depressive disorder: comparison of clinical study reports, trial registries, and publications. *Bmj*. 2014 Jun 4;348:g3510.
- [13] Maund E, Tendal B, Hróbjartsson A, Lundh A, Gøtzsche PC. Coding of adverse events of suicidality in clinical study reports of duloxetine for the treatment of major depressive disorder: descriptive study. *BMJ*. 2014 Jun 4;348:g3555. <http://www.bmj.com/content/348/bmj.g3555>
- [14] Maund E, Guski LS, Gøtzsche PC. Considering benefits and harms of duloxetine for treatment of stress urinary incontinence: a meta-analysis of clinical study reports. *Canadian Medical Association Journal*. 2017 Feb 6;189(5):E194-203. <http://www.cmaj.ca/content/early/2016/11/14/cmaj.151104.abstract>
- [15] Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *New England Journal of Medicine*. 2009 Nov 12;361(20):1963-71.
- [16] [External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use (Revision 2 of 12/04/2017 is considered here): [http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general\\_content\\_001799.jsp&mid=WC0b01ac0580b2f6ba](http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_001799.jsp&mid=WC0b01ac0580b2f6ba)

## PhUSE 2017

- [17] OECD Data Utility Definition: <https://stats.oecd.org/glossary/detail.asp?ID=6905>
- [18] Musini V, Lawrence K, Wright J. Using Clinical Study Reports versus published articles in a Cochrane Review update. In: Challenges to evidence-based health care and Cochrane. Abstracts of the 24th Cochrane Colloquium; 2016 23-27 Oct; Seoul, Korea. John Wiley & Sons; 2016. <http://abstracts.cochrane.org/2016-seoul/using-clinical-study-reports-versus-published-articles-cochrane-review-update>
- [19] Jefferson, « EMA's policy 0070: opportunities, challenges and measuring success », DIA 2017, Glasgow. Slides not publicly available.
- [20] Hodkinson A, Dietz KC, Lefebvre C, Golder S, Jones M, Doshi P, Henegan C, Jefferson T, Boutron I, Stewart L. The use of clinical study reports and other regulatory documents to enhance the quality of systematic reviews: a survey of systematic review authors [submitted]. Summary available at: <http://methods.cochrane.org/methods-innovation-fund-2>
- [21] Beaumier J, Lawrence K, Mintzes B, Jauca C, Wright J, Perry T, Turner E, Puil L. Accessing clinical study reports: potential to reduce bias in systematic reviews. In: Filtering the information overload for better decisions. Abstracts of the 23rd Cochrane Colloquium; 2015 3-7 Oct; Vienna, Austria. John Wiley & Sons; 2015.
- [22] CSDR. ClinicalStudyDataRequest.com <https://clinicalstudydatarequest.com/>.
- [23] The Yale University Open Data Access (YODA) Project. The YODA Project: Forging a unified scientific community <http://yoda.yale.edu/>.
- [24] Tudur Smith C, Marcucci M, Nolan SJ, Iorio A, Sudell M, Riley R, Rovers MM, Williamson PR. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. The Cochrane Library. 2016 Sep 6.
- [25] Nolan SJ, Thornton J, Murray CS, Dwyer T. Inhaled mannitol for cystic fibrosis. Cochrane Database of Systematic Reviews 2015, Issue 10. Art. No.: CD008649. DOI: 10.1002/14651858.CD008649.pub2. <https://www.ncbi.nlm.nih.gov/pubmed/26451533>
- [26] Nolan SJ, Tudur Smith C, Weston J, Marson AG. Lamotrigine versus carbamazepine monotherapy for epilepsy: an individual participant data review. Cochrane Database of Systematic Reviews 2016, Issue 11, Art No. CD001031. DOI: 10.1002/14651858.CD001031.pub3: <https://www.ncbi.nlm.nih.gov/pubmed/27841445>
- [27] Overview of comments from stakeholders received as part of public consultation on EMA Policy 0070 of June 2013: [http://www.ema.europa.eu/ema/index.jsp?curl=pages/special\\_topics/general/general\\_content\\_000556.jsp&mid=WC0b01ac05809f363f](http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000556.jsp&mid=WC0b01ac05809f363f)
- [28] Scaiano 2016: <http://www.sciencedirect.com/science/article/pii/S1532046416300697>
- [29] Principle of Responsible Clinical Trial Data Sharing: <http://transparency.efpia.eu/uploads/Modules/Documents/data-sharing-prin-final.pdf>
- [30] Bonini S, Eichler H-G, Wathion N, Rasi G. Transparency and the European Medicines Agency — Sharing of Clinical Trial Data. NEJM 371;26. <http://www.nejm.org/doi/pdf/10.1056/NEJMp1409464>
- [31] PhUSE Data De-Identification Standard for SDTM 3.2 (2015) <http://www.phuse.eu/data-transparency-download>



## PhUSE 2017

### APPENDICES

#### APPENDIX 1: EXAMPLES OF ACADEMIC WORK USING CSRS IN SECONDARY RESEARCH

Drug under consideration and references	Regulatory data Source(s)	Methods	Manuscript Conclusion*
<b>Reboxetine</b> Eyding <i>et al</i> [4]	13 CSRs and other regulatory documents from published and unpublished trials provided by Pfizer	<ul style="list-style-type: none"> <li>• Data and method comparison (<i>CSRs and protocols compared to trial manuscripts, trial registries and regulatory authority websites</i>).</li> <li>• Data extraction</li> <li>• Systematic review and meta-analysis</li> <li>• AE counts</li> </ul>	<ul style="list-style-type: none"> <li>• Published data overestimated the benefit of reboxetine versus placebo by up to 115% and reboxetine versus SSRIs by up to 23%, and also underestimated harm.</li> <li>• Reboxetine is, overall, an ineffective and potentially harmful antidepressant. Published evidence is affected by publication bias, underlining the urgent need for mandatory publication of trial data.</li> </ul>
<b>Paroxetine and imipramine</b> Le Noury <i>et al</i> [8]	Individual participant data, CSR and Appendices available on GSK website and additional appendices provided by GSK for one trial (Study 329)	<ul style="list-style-type: none"> <li>• Data and method comparison (<i>IPD, CSR and original analysis compared</i>).</li> <li>• Re-analysis</li> <li>• AE counts</li> </ul>	<ul style="list-style-type: none"> <li>• Neither paroxetine nor high dose imipramine showed efficacy for major depression in adolescents, and there was an increase in harms with both drugs.</li> <li>• Access to primary data from trials has important implications for both clinical practice and research, including that published conclusions about efficacy and safety should not be read as authoritative.</li> <li>• The reanalysis of Study 329 illustrates the necessity of making primary trial data and protocols available to increase the rigour of the evidence base.</li> </ul>
<b>Orlistat</b> Schroll <i>et al</i> [9]	7 CSRs provided by Roche	<ul style="list-style-type: none"> <li>• Data and method comparison (<i>CSRs compared to protocols and trial manuscripts</i>).</li> <li>• AE counts</li> </ul>	<ul style="list-style-type: none"> <li>• For one trial, an additional 1,318 adverse events were identified that were not listed or mentioned in the CSR itself but could be identified through manually counting individual adverse events reported in an appendix. The majority of patients had multiple episodes of the same adverse event that were only counted once, though this was not described in the CSRs.</li> <li>• In the orlistat trials, we identified important disparities in the reporting of adverse events between protocols, clinical study reports, and published papers. Reports of these trials seemed to have systematically understated adverse events. Based on these findings, systematic reviews of drugs might be improved by including protocols and CSRs in addition to published articles.</li> </ul>
<b>Orlistat</b> Hodkinson <i>et al</i> [10]	5 CSRs provided by Roche	<ul style="list-style-type: none"> <li>• Data and method comparison (<i>CSRs compared to trial manuscripts</i>).</li> <li>• AE counts</li> </ul>	<ul style="list-style-type: none"> <li>• Journal publications provided insufficient information on harms outcomes of the Orlistat trials and did not specify that a subset of harms data were being presented.</li> <li>• CSRs often present more complete data on harms, including serious adverse events</li> <li>• CSRs could support a more complete, accurate, and reliable investigation, and researchers undertaking evidence synthesis of harm outcomes should not rely only on incomplete published data that are presented in the journal publications.</li> </ul>

## PhUSE 2017

<p><b>Oseltamvir</b> Jefferson <i>et al</i> [11]</p>	<p>83 CSRs obtained from EMA and Roche, of which 23 were used in the systematic review and meta-analysis</p>	<ul style="list-style-type: none"> <li>• Data and method comparison <i>(CSRs compared to trial manuscript)</i>.</li> <li>• Data extraction</li> <li>• Systematic review and meta-analysis</li> <li>• AE counts</li> </ul>	<ul style="list-style-type: none"> <li>• This is a report of the first Cochrane review to be based on all relevant full clinical study reports of a drug, augmented by regulatory comments.</li> <li>• The trade-off between benefits and harms should be borne in mind when making decisions to use oseltamivir for treatment, prophylaxis, or stockpiling.</li> </ul>
<p><b>Duloxetine</b> Maund <i>et al</i> [12-14]</p>	<p>9 CSRs and protocols as appendices obtained from the EMA.</p>	<ul style="list-style-type: none"> <li>• Data and method comparison <i>(CSRs compared to protocols, trial manuscripts and clinicaltrials.gov entries)</i>.</li> <li>• Data extraction</li> <li>• Systematic review and meta-analysis</li> <li>• AE counts</li> </ul>	<ul style="list-style-type: none"> <li>• CSRs contained extensive data on harms that were unavailable in journal articles and trial registry reports. There were inconsistencies between protocol and CSRs.</li> <li>• The listings of adverse events for individual patients and narratives of adverse events within CSRs can provide additional information, including original investigator reported adverse event terms, which can enable a more accurate estimate of harms.</li> <li>• Following re-analysis using data from CSRs, the apparent harms of Duloxetine outweigh the benefits</li> </ul>
<p><b>Gabapentin</b> Vedula <i>et al</i> [15]</p>	<p>20 CSRs (and other regulatory documents such as protocols) provided by Pfizer and Parke-Davis</p>	<ul style="list-style-type: none"> <li>• Data and method comparison <i>(CSRs and other regulatory documents compared to published trial manuscripts)</i>.</li> </ul>	<ul style="list-style-type: none"> <li>• For 8 of the 12 trials reported as trial manuscripts, the primary outcome differed in the published trial compared to the CSRs / protocol</li> <li>• Other sources of disagreement between published trials and regulatory documents included introduction of a new primary outcome, failure to distinguish between primary and secondary outcomes, relegation of primary outcomes to secondary outcomes and failure to report one or more protocol-defined primary outcomes.</li> <li>• This selective reporting of off-label use of gabapentin which threatening the validity of evidence for the effectiveness of off-label interventions</li> </ul>

\*: NOTE: Some of the findings of these research papers have been challenged by the concerned pharmaceutical companies and responses to manuscripts have been published on the journal website.

### APPENDIX 2: CORRESPONDENCE WITH AUTHORS OF MANUSCRIPTS USING CSRS FOR RESEARCH

#### Conversation with Dr Jon Jureidini, 18 August 2017.

Dr Jureidini was involved in the project relating to the manuscript:

*Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. BMJ 2015;351:h4320. doi: 10.1136/bmj.h4320*

Dr Jureidini spoke to Sarah Nevitt regarding this project on 18 August 2017. This appendix summarises the key findings of this work with comments from Dr Jureidini (conversations summarised by SN, comments are not direct quotes from Dr Jureidini).

This was the first study published using data from CSDR. The authors also used CSRs which were available on the GSK website (including appendices) that had been redacted to some extent. The initial reporting of Study 329 has been considered controversial, hence this research team wished to perform a re-analysis with the original data and regulatory documents of Study 329. The original analysis found the drugs to show efficacy with limited harms (only significant adverse events reported). The re-analysis shows that neither drug actually shows any efficacy, and there was an increase in harms from both drugs.

*Comment from Dr Jureidini (on the rationale for the work): Everyone is interested in efficacy of the drugs and p values. This is what the main papers always concentrate on and the harms are always a secondary consideration. We should be more interested in the harms of the data, these are never reported enough and it can be hard to get people interested in harms. To get a full picture of potential harms, you need to look a lot deeper into the data and the CSRs – this will never be fully captured in a trial publication.*

*There were clear issues with the initial publication of Study 329 – not fraudulent, the authors were very honest, almost too honest as it revealed that the efficacy they described was at odds with the data. This became even more evident looking at the CSRs.*

The use of CSRs in this project was for the harms data. The authors recoded some of the narrative descriptions of adverse events. Due to unclear descriptions within the publically available appendices from the GSK website, the authors requested an extra unpublished extra appendix of the study CRFs. This was provided via the CSDR remote data access system.

*Comment from Jureidini: The system was so cumbersome and difficult to access that due to the time involved of recoding and comparing adverse events, we only completed this task for around a third of AEs; the AEs we considered to be the most important. We were criticised for this by reviewers of the BMJ article but as we had no funding and limited resources for this analysis but didn't have a lot of choice.*

*The redaction of information from narratives and appendices didn't really cause any issues as far as I can recall. The problem was that some of the text was illegible – maybe due to the age of the document which wasn't helped by the remote data access system which prevented zooming in, using software to improve quality etc.*

*We had no problem with the content or interpreting the documents etc. The problems were with the quality of the documents and trying to use them in the remote system.*

*This analysis would have been impossible without access to the appendices and it would be 'travesty' if the implication of EMA Policy 0070 means that appendices are completely redacted or no longer shared.*

#### Correspondence with Dr Beate Wieseler, 20 August 2017

Questions were sent to Dr Wiesler by Sarah Nevitt to further explore the use of CSRs in the project relating to the manuscript

*Eyding D, Lelgemann M, Grouven U, et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. BMJ 2010;341:c4737. doi: 10.1136/bmj.c4737*

Dr Wieseler provided the following additional references reporting on the experiences of the research group using CSRs

## PhUSE 2017

Wieseler B, Wolfram N, McGauran N, et al. Completeness of Reporting of Patient-Relevant Clinical Trial Outcomes: Comparison of Unpublished Clinical Study Reports with Publicly Available Data. *PLOS Medicine* 2013;10(10):e1001526. doi: 10.1371/journal.pmed.1001526

Wieseler B, Kerekes MF, Vervoelgyi V, McGauran N, Kaiser T. Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. *BMJ*. 2012 Jan 3;344:d8141.

Köhler M, Haag S, Biester K, Brockhaus AC, McGauran N, Grouven U, Kölsch H, Seay U, Hörm H, Moritz G, Staeck K. Information on new drugs at market entry: retrospective analysis of health technology assessment reports versus regulatory reports, journal publications, and registry reports. *BMJ*. 2015 Feb 26;350:h796.

Dr Wiesler also provided the following responses to the questions by e-mail and is happy for the responses to be reported in this paper:

1. Could you explain the rationale for using CSRs for your research (rather than published trial reports, IPD etc.)?

We are interested in the full evidence base of the interventions under assessment in our Health Technology Assessment reports to be able to provide unbiased assessments for decision making in the German health care system. By using regulatory documents like CTD clinical overviews and clinical summaries and specifically CSRs we want to overcome publication and outcome reporting bias.

Specifically we need full information on trial methods including the study protocol with any amendments and the statistical analysis plan (SAP) to be able to understand the study planning or any changes in the conduct of the study or the analyses. Among other things we use this information to choose relevant endpoints/analyses and assess the risk of bias on a study or endpoint level. Journal publications and registry reports do not provide the level of detail we need. E.g. journal publications do not provide full inclusion/exclusion criteria or full specification of endpoints or statistical analyses (including information on handling of missing or specification of statistical methodology like models used in the analyses.)

Furthermore, we need full numerical summary data on all endpoints (including information on which endpoints/analyses were pre-specified or defined post-hoc) to be able to choose the endpoints/analyses relevant to our assessment and to conduct meta-analyses. Often numerical information in journal publications is too limited for meta-analysis. We are also interested in subgroup or sensitivity analyses which also might not be presented in documents other than the CSR.

2. Were any specific outcomes reported within the CSRs for interest to you?

We are specifically interested in patient-relevant outcomes (i.e. outcomes describing how a patient feels, functions or survives). Often these endpoints are secondary endpoints which are not fully published in journals. Our research has described and quantified to which extent we could add endpoints from CSRs as compared to other sources (*BMJ* 2012;344:d8141, *PLoS Med* 10(10): e1001526; *BMJ* 2015;350:h796)

3. Which sections of the CSRs did you use for your research (e.g. summary tables, narratives, appendices, participant listings, others sections etc.)

We use the full CSR including end-of-text tables and appendices for our work.

4. If you used the narratives, can you explain exactly what you used them for? For example, did you wish to extract participant level information to create a dataset

We rarely use narratives because the information we use in most cases is available also in other data presentations (e.g. listings or summary tables of SAEs). However, if these other data presentations are missing, we might extract information from narratives.

5. Did you have any difficulties in extracting or interpreting information from the CSRs?

We do not have difficulties in extracting or interpreting information from well written CSRs.

6. Were there any key findings of your research specifically relating to the use of CSRs?

The CSRs of reboksetine allowed to assess all evidence available in a situation where only a biased set of studies was published in journals (please see *BMJ* 2010;341:c4737). We have been able to extract substantially more methodological and endpoint information for our assessments from CSRs than from other sources (please see

## PhUSE 2017

references given under Question 2). Furthermore, often we were only able to understand a study included in our assessment fully by using the CSR including the protocol and SAP (please see Z Evid Fortbild Qual Gesundhwes. 2017 Feb;120:3-8 for a description of the example of our assessment of liraglutide).

7. Would you have been able to complete this research without access to CSRs?

Without CSRs we would have missed important information.

8. Would you have been able to complete this research if narratives and/or appendices (with participant listings) had been removed from the CSRs?

We definitely do need appendices including protocols and amendments and the SAP. If we need participant listings depends on the open questions that arise from the assessment (see e.g. the assessment of liraglutide in which we needed the information of when hypoglycemic events occurred in individual patients).

9. Would you have been able to complete this research if you had been provided with anonymised CSRs (in line with EMA policy 0070)?

In rare cases I can foresee that we need to follow study participants through several listings. This would only be possible if there is some kind of ID (not necessarily the original ID from the study).

### Correspondence with Dr Jeppe Schroll, 31 August 2017

Questions were sent to Dr Schroll by Sarah Nevitt to further explore the use of CSRs in the project relating to the manuscript

*Schroll JB, Bero L, Gotzsche PC. Searching for unpublished data for Cochrane reviews: cross sectional study. BMJ 2013;346:f2231. doi: 10.1136/bmj.f2231*

Dr Schroll provided the following responses by e-mail and is happy for the responses to be reported in this paper:

1. Could you explain the rationale for using CSRs for your research (rather than published trial reports, IPD etc.)?

The purpose of our paper was to compare published with unpublished data.

2. Were any specific outcomes reported within the CSRs for interest to you?

We covered reported adverse event and methods on how to handle adverse events. Lots of outcomes were reported that were not available in the publications

3. Which sections of the CSRs did you use for your research (e.g. summary tables, narratives, appendices, participant listings, others sections etc.)

We used summary tables, appendices and narratives and participant listings. For our CSR individual patient listings of adverse events were available.

4. If you used the narratives, can you explain exactly what you used them for? For example, did you wish to extract participant level information to create a dataset

We explored the reason for discontinuation and found divergence from the listing of withdrawals. In the narratives the sponsors' causality assessment was available.

5. Did you have any difficulties in extracting or interpreting information from the CSRs?

It was a great difficult that the documents were not in a "text readable" format. Which made electronic searching impossible. The pages had more than one page number and there was – in out material – no overall table of contents. Only for the specific sections.

6. Were there any key findings of your research specifically relating to the use of CSRs?

We found that a lot of data is left out of the publications and that important limitations and assumptions were not described.

7. Would you have been able to complete this research without access to CSRs?

## PhUSE 2017

No

8. Would you have been able to complete this research if narratives and/or appendices (with participant listings) had been removed from the CSRs?

No

9. Would you have been able to complete this research if you had been provided with anonymised CSRs (in line with EMA policy 0070)

Anonymised CSRs are ok, but the current EMA policy redacts important information about when the adverse events appeared as well as what they were. Newer CSRs does not have individual adverse event listings and the EMA are not even in possession of them.

### **Correspondence with Professor Catrin Tudur Smith and Dr Alex Hodkinson, September 2017**

Questions were sent to Professor Tudur Smith (CTS) and Dr Hodkinson (AH) by Sarah Nevitt to further explore the use of CSRs in the project relating to the manuscript

*Hodkinson A, Gamble C, Smith CT. Reporting of harms outcomes: a comparison of journal publications with unpublished clinical study reports of orlistat trials. Trials 2016;17(1):1-11. doi: 10.1186/s13063-016-1327-z*

Professor Tudur Smith (CTS) and Dr Hodkinson (AH) provided the following responses by e-mail and is happy for the responses to be reported in this paper:

1. Could you explain the rationale for using CSRs for your research (rather than published trial reports, IPD etc.)?

CTS: As a methodological exercise to compare the information available in CSR s versus information available in published journal articles

AH: Really this was just a methodological piece of work displaying the value of CSRs when there evidence for underreporting of harms in published literature.

2. Were any specific outcomes reported within the CSRs for interest to you?

CTS: Adverse events only

AH: All harm outcomes (AEs and SAEs) but also the timing of events, grading and safety narratives if reported

3. Which sections of the CSRs did you use for your research (e.g. summary tables, narratives, appendices, participant listings, others sections etc.)

CTS: We didn't receive narratives but this may have provided additional information. The main source of information we used were the IPD listings of adverse events and tables

AH: Roche I believe have changed the format of their CSRs since this piece of work, but data were primarily obtained from the following sections:

- Module I (core report which follows the ICH E3 format): sections 2.6 (safety parameters), 2.9.4 (Safety analysis), 3.4 (safety results).
- Modules II: glossary of original and MedDRA preferred terms for AEs
- Module V: section 5, Safety analysis plan

4. If you used the narratives, can you explain exactly what you used them for? For example, did you wish to extract participant level information to create a dataset

AH: No we didn't use the safety narratives, as they were rarely reported. If they were reported consistently we may

## PhUSE 2017

have been able to create an IPD dataset.

5. Did you have any difficulties in extracting or interpreting information from the CSRs?

CTS: Large documents that are difficult to navigate at first but they were well structured and so quite straightforward to locate the information

AH: Not really, other than Roche had removed a number of pages in one of the CSRs and redacted quite a lot of information throughout. We did contact them about this, apparently there was some confidential information on these pages. I wasn't convinced.

6. Were there any key findings of your research specifically relating to the use of CSRs?

CTS: We found additional information and additional detail available in the CSRs but also additional time required to extract the information

AH: Obviously there were high numbers of AE and particularly SAEs that were not reported in the publication (see bar graphs in paper), and 5 statistically significant harm outcomes were detected using the CSRs.

7. Would you have been able to complete this research without access to CSRs?

CTS: No

AH: Due to the underreporting of harms data in publications probably not. CSRs are far more detailed than publications, and I would recommend they are used where possible for synthesis of harms.

8. Would you have been able to complete this research if narratives and/or appendices (with participant listings) had been removed from the CSRs?

AH: N/A as we didn't use the safety narratives, and the participant listing of AEs in Module IV of the report were removed.

9. Would you have been able to complete this research if you had been provided with anonymised CSRs (in line with EMA policy 0070)

CTS: Possibly

AH: I have actually looked at data that are released under the EMAs new policy 70, and they do provide fully redacted CSRs. So yes, I would say you could use these provided the drug is centrally licenced. But redactions may permit what data can be used, and they may not be of use for creating IPD datasets without the subject IDs and other patient-level info.