Clinical Trial and Imaging

Subgroup report

Vesa Kiviniemi, FI (subgroup lead)

Aldana Rosso, DK (since October 2017)

Zsuzsanna Cserjes Szabone, HU

Martin Nyeland, DK (until October 2017)

Per Fuglerud, NO (until October 2017)

An agency of the
European Union

Table of contents

List of Abbreviations

ADR	Adverse drug reaction
ALS	Amyotrophic lateral sclerosis
BIO	Biotechnology Innovation Organization
CDISC	Clinical Data Interchange Standards Consortium
CTTI	Clinical Transformation Initiative
EDPS	European Data Protection Service
EHR	Electronic healthcare records
EU	European Union
FDA	United States Food and Drug Administration
FREEBIRD	Free Bank of Injury and Emergency Research Data
HIV	Human immunodeficiency virus
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IMI	European Innovative Medicines
IMPACT	International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury
IPD	Individual patient level data
MAH	Marketing authorisation holder
MRCT	Multi-Regional Clinical Trial Centre
NCA	National competent authorities
NIH	National Institute of Health
NNIDK	National Institute for Diabetes, Digestive and Kidney Diseases
OHDSI	Observational Health Data Science and Informatics
PAES	Post-authorisation efficacy studies
PASS	Post-authorisation safety studies
PMDA	Pharmaceutical and Medical Devices Agency, Japan
PRO	Patient reported outcome
PRO-ACT	Pooled Resource Open-Access ALS Clinical Trials Database
RCT	Randomised controlled trial
WWARN	Worldwide Antimalarial Resistance Network
YODA	Yale University Open Data Access

1. Summary

The efficacy and safety of medicinal products is generally demonstrated through clinical trials that are conducted in accordance with European Union (EU) legislation: Directive 2001/83/EC if conducted within in the EU and Directive 2001/20/EC if conducted outside the EU but submitted in an application for a marketing authorisation in the EU. On 16 April 2014, the European Commission adopted the new Clinical Trial Regulation (EU No 536/2014), repealing Directive 2001/20/EC, which comes into application in 2019.[1] Clinical trials are the fundamental basis for almost any key regulatory decision, especially for the evaluation of marketing authorisations and variations (or other post-approval decisions) and hence the link between clinical trials and regulatory decisions is obvious.

The purpose of this report is to provide an overview of the opportunities and challenges posed by the combination of data from different clinical trials into large datasets, bringing clinical trials into the domain of big data. The steps required for using big data from clinical trials in the regulatory process are several. First, data from existing clinical studies need to be characterised to understand how the data can be combined with data from other studies. Second, data from all the different clinical studies need to be combined in a database. Finally, regulatory applications have to be explored. In this report, we provide an illustrative description of different data sharing and standardisation activities. We also discuss the opportunities and challenges of combining data from several clinical trials to inform regulatory decision.

The following points characterise the status of the implementation of big data related to clinical trials in the regulatory context:

1) Clinical trial data sharing activities are currently relatively mature and they are already providing access to many thousands of clinical trials. However, it is not fully clear nor understood how these activities could be applied in regulatory or scientific procedures.
2) Data standardisation activities are critical to ensure usability and applicability of data.
3) Anonymisation in data sharing activities is a difficult balance between data utility and data privacy.
4) Imaging has promising potential in a big data context, but specific expertise in this area is lacking within the national competent authorities (NCAs), and the use of these data in the regulatory context is rare.
5) Data from single clinical trials including temporal high-frequency or high-dimensional data are not currently used efficiently for regulatory decision-making. Currently there are no guidelines or clear principles for example defining new type of outcomes based on the output from the devices (or imaging) generating high-frequency data.

Several steps could be taken to improve the points mentioned above. During the preparation of this report, we observed that there is a need to build expertise within these areas both in the NCAs and in the industry in order to incorporate results obtained by combining data from clinical trials into the

---

[1] The Regulation harmonises the assessment and supervision processes for clinical trials throughout the EU, via an EU portal and database which will be established and maintained by the EMA, in collaboration with the Member States and the European Commission. The goal of Clinical Trial Regulation EU No. 536/2014 is to create an environment that is favourable to conducting clinical trials in the EU, with the highest standards of safety for participants and increased transparency of trial information. The Regulation will require: (i) consistent rules for conducting clinical trials throughout the EU; (ii) information on the authorisation, conduct and results of each clinical trial carried out in the EU to be publicly available. This will increase the efficiency of all trials in Europe with the greatest benefit for those conducted in multiple Member States. It aims to foster innovation and research, while helping avoid unnecessary duplication of clinical trials or repetition of unsuccessful trials. The authorisation and oversight of clinical trials remains the responsibility of Member States, with EMA managing the database and supervising content publication on the public website.

regulatory process. In order to promote incentives to use big data from clinical trials, it is essential that support is given at European level to encourage applications of data sharing in the regulatory setting. Such actions may potentially involve requesting individual patient data in regulatory submissions, enhancing the use of historical data, supporting any actions of data sharing to inform scientific advice and developing methodological approaches for combined data (on a large scale). Furthermore, data collected in several clinical trials could be used for further development and assessment of currently accepted endpoints, in particular those based on scales or a single observation at a certain time point. Finally, the regulatory use of clinical trials including high-frequency data and applications of imaging needs further exploration. This could potentially involve mapping of relevant analytical approaches and providing guidance to further enhance new hypothesis generation, better definition of responses, generation of new types of outcomes and increased knowledge of patient characteristics. Based on the above, the Clinical Trial and Imaging Subgroup recommends the following actions:

1) Combining clinical trial data offers unexplored and useful opportunities for improving decision-making throughout the product life cycle, from drug discovery to safety surveillance. Regulators should encourage using approaches where individual patient data from multiple trials are used.

2) Individual patient data (IPD) from clinical trials should be requested as part of the MAH submission and assessed as part of the review of the marketing authorisation application. This would require establishing the legal basis for requesting IPD and agreeing on data format for submissions. Furthermore, NCAs may need to increase statistical capacity and skills for analysis of IPD.

3) Data harmonisation, both in terms of format and variable definitions, is a key requirement to facilitate the combination of clinical trial data. Furthermore, in the era of multiregional clinical trials standards should be as globally aligned as possible. In order to harmonise terminology, the use of international standards such as the International Classification of Diseases (ICD) and the Medical Dictionary for Regulatory Activities (MedDRA) and IDMP should be encouraged.

4) Actions to create open data platforms for proactive sharing of clinical data should be supported. In particular, efforts targeting diseases with scarce data, such as "ultra-rare diseases" or high-risk populations should be promoted.

5) Encouraging more extensive or exhaustive use of images, genomic data, data from wearables collected during a single clinical trial combined with big data analytics could benefit drug discovery and may further inform regulatory decisions. Guidance should be developed to efficiently use these data in the regulatory processes.

2. Background

Clinical trials are the main source of information when regulatory decisions are taken to assess the benefits and risk of pharmaceutical products. Clinical trials can be industry-sponsored or investigator-initiated. The primary responsibility of interpretation and analysis of data generated from the clinical trial setting regulated under the regulation (EU) No 536/2014 and the present directive 2001/20/EU on clinical trials lies with those responsible for developing the new drugs and publishing data (sponsors). In order to reach the correct conclusions from medical research, it is of paramount importance that the data used in the studies are of high quality. According to ISO 8402-1986, quality is defined as "The totality of features and characteristics of an entity that bears on in its ability to satisfy stated and implied needs". In the case of data obtained from clinical trials, this definition implies that the data should be accurate, complete, consistent and representative of the population to which the results are going to be extended to. Representativeness is of outmost importance in clinical trials. Clinical trials

are conducted under idealised and rigorously controlled conditions to minimise the possibility of bias regarding the effect of the investigated product. However, the disadvantage is that the results of the trial may not be directly generalised to real life patients. A review of randomised controlled trials (RCT) in fields of cardiology, oncology and mental health indicates that a high proportion of the general disease population is often excluded from trials (Kennedy-Martin et al., 2015). Thus, the trial samples are not fully representative of patients treated in clinical practice (Kennedy-Martin, Curtis et al. 2015). On the other hand, no individual clinical trial can be expected to be totally representative of future users because of the strict conditions under which the trial is conducted: inclusion and exclusion criteria, influences of geographical location, the time when it is conducted, the medical practices of the particular investigator(s) and clinics, and so on (ICH 1998).

Despite the above-mentioned limitations, data from clinical trials offer several advantages. In clinical trials, the objectives, design and analytical methods are mostly pre-planned. The study population, primary and secondary variables, type of comparison are chosen to fulfil the objectives of the trial. Furthermore, several aspects of the study, such as the protocol, design, definition of study population, outcomes and follow-up times are publicly accessible on websites such as for example EudraCT and ClinicalTrials.gov. The main advantage of clinical trials is that the treatment is assigned in a randomised form, and thus, no selection bias is expected. In general, clinical trial data are well-structured, and the internal quality of the data (integrity and veracity) is very high. Unstructured data such as images and text may appear in clinical trials, but the actual data are usually transformed into a structured numerical format and (in that sense) are straightforward to process and analyse. The potential for combining clinical data from several studies is therefore feasible (Babre 2013).

During the last years, there have been several initiatives to share data from clinical trials in order to maximise insight, increase power, inform new trials and prevent duplication of effort and the exposure of patients to unnecessary treatment and investigations (Lo and DeMets 2016, Rockhold, Nisen et al. 2016). Publicly available data not only increase transparency, it also allows for independent confirmation of the results, secondary analyses and investigation of new research questions (Lo and DeMets 2016, Rockhold, Nisen et al. 2016, Warren 2016, Bertagnolli, Sartor et al. 2017). From the methodological perspective, sharing individual patient data (IPD) from clinical trials offers valuable opportunities for using the data in IPD meta-analyses. IPD meta-analyses give a more reliable estimation of treatment effects than those obtained using only aggregated data (Kawahara et al 2018). This is mainly because detailed covariate information is available and consequently also the possibility to examine the effect of risk factors and effect modifiers. Patients who participated in trials have also expressed consensus to share their data as long as their integrity is protected[2].

One of the main challenges of sharing data across different trials is standardisation since a common definition of variables and data collection are essential (Rockhold, Nisen et al. 2016, Rockhold 2017). Moreover, transparency in the anonymisation approaches applied is required in order to enable pooling of patients across trials. Therefore, data sharing from clinical trials and the corresponding standardisation activities are identified as key issues. In this report, we first describe the current situation regarding data sharing activities. Later, we focus on the opportunities and challenges of implementing clinical trial big data for regulatory decision-making.

---

[2] Patient panelists at the NEJM meeting in June 2017 expressed consensus that their trials data should be shared (with their privacy protected) and were surprised that there was debate over this issue. Patients seemed unaware of tension between trialists and data analysts, and of the existence of disincentives to data sharing. One speaker, Nancy Nagler, who has participated in multiple cancer trials, said she now feels even more strongly that data should be shared, the sooner the better, and in as much detail as possible. Nagler also advocated for a formal process to "close the loop" and share trial findings in an appropriate format with participants. Information obtained from Clinical Informatics News, August 22, 2017. http://www.clinicalinformaticsnews.com/2017/08/22/data-sharing-stakeholder-perspectives-on-transparency-in-clinical-trials.aspx, accessed 20-11-2017

Objectives:

The task of the Clinical Trial and Imaging Subgroup are:

- To identify main sources of accessible clinical trial data (data sharing activities).

- To discuss issues on data standardisation and harmonisation activities.

- To identify regulatory value and challenges related to use of these big data sources to inform decision-making throughout the pharmaceutical product life cycle.

In terms of imaging, the following should be noted. Overall, imaging may be linked to clinical trials, but it is also a domain of its own and at this stage, this area is not covered in depth. However, due to the importance of this area in the big data context, a few preliminary thoughts are expressed in Section 5.

3. Methods

Definitions and scope

In this report, the ICH definition of a clinical trial is adopted: *Any investigation in human subjects intended to discover or verify the clinical, pharmacological and/or other pharmacodynamic effects of an investigational product(s), and/or to identify any adverse reactions to an investigational product(s), and/or to study absorption, distribution, metabolism, and excretion of an investigational product(s) with the object of ascertaining its safety and/or efficacy*.

The clinical trial data considered in this report are as follows:

- Data obtained from studies on human subjects (including healthy volunteers and patients) are considered.

- Interventional studies performed under a study protocol where the subjects are assigned to receive one or more interventions (or no intervention). The assignment is determined by the study protocol. This includes phases 1 to 3 and pragmatic studies.

- Interventional Post-Authorisation Safety (PASS) and Post-Authorisation Efficacy (PAES) studies.

Investigational products considered here include:

- Any pharmaceutical forms of an active ingredient or placebo being tested or used as a reference in a clinical trial. This also includes a product with a marketing authorisation when used or assembled (formulated or packaged) in a way different from the approved form, or when used for an unapproved indication, or when used to gain further information about an approved use.

Endpoints considered here include:

- Efficacy, effectiveness, safety, pharmacokinetics, pharmacodynamics, quality of life and patient-reported outcomes (PRO).

Type of data format included in this report:

- Individual patient data (IPD).

The following topics are excluded from this report and fall outside the scope:

- Data from veterinary medicinal products, animal data and in silico studies.
- Data from studies on medical devices and diagnostic tools.
- Pharmacoeconomic studies and their outcomes.
- Methodological approaches such as meta-analyses and other literature review-based approaches.

Search strategy and criteria to identify and select data sources

The Wellcome Trust foundation published a thorough assessment report on access to clinical trial data, which was our main source of information (Varnai, Rentel et al. 2014). Data sharing activities were identified from this report and further characterised according to the following elements (as applicable):

- Data structure e.g. terminology, structured vs unstructured data etc.
- Provenance of data.
- Data Quality (integrity and veracity).
- Data heterogeneity (variability).
- Speed of change; rate of accumulation.
- Completeness and opportunities to capture data.
- Methods to analyse the data.
- Accessibility of data for NCAs.

Search strategy to identify opportunities for implementation of large clinical datasets in the regulatory process

The applications of combining clinical trial data mentioned in this report were found using a general search in google. A directed search in PubMed using the terms "clinical data sharing" was also performed. The examples included here are to highlight different applications of combining clinical trial data in the regulatory process and are not exhaustive.

Search strategy limitations

The literature search performed when writing this report is not exhaustive. Data standardisation and harmonisation activities in order to improve sharing of data from clinical trials are continuously ongoing. The initiatives presented here are examples to illustrate how clinical data sharing could be implemented.

4. Data characterisation

4.1. Volume – structure and size of the data source

Clinical trial data are usually a set of values of qualitative or quantitative variables. The data are usually structured following the "one patient per row" principle. In this report, we focus on the data defined in the study protocol and collected from human subjects in a clinical trial.

The data of interest are coded, transcribed, abstracted and corrected from raw data sources and made accessible in the final cleaned and locked analysable database (based on the Institute of Medicine's document "Discussion Framework for Clinical Trial Data Sharing").

In overall terms, the number of patients typically recruited into a single clinical trial varies from very few subjects (e.g. 30 patients) to tens of thousands subjects (e.g. 20,000 patients). The size of a study mainly depends on its nature and objectives, the outcomes and the presumed size of the effect being studied. The number of patients recruited into a single clinical trial is typically estimated prior to patient recruitment, using sample-size calculations. However, it can be modified when the study is ongoing if pre-specified interim analyses are performed. A single clinical trial without high-frequency data does not itself represent a big data source as analysis can be conducted using traditional statistical methodology and software. However, data volume, variability and level of complexity dramatically increase when high-frequency data (i.e. signal data from wearables, images, etc.) is included into a clinical trial.

Combining clinical data sources clearly increases the size of the data source dramatically. The ClinicalStudyDataRequest is a consortium of clinical study data providers established in 2013, and it is an example of large database of individual clinical studies. Beginning in January 2014, requests for individual patient data could be made from this consortium's public website, clinicalstudydatarequest.com. The requests were subject to approval by an independent review panel. Today, data from 3,049 trials conducted by 13 pharmaceutical companies are available through the website. These data concern both approved medicines and data from terminated research programmes. Both raw and analysis-ready datasets are provided, along with supporting documentation including the protocol, data specifications, annotated case report forms and clinical study reports.

Another data sharing activity is the PRO-ACT database coordinated and implemented by the non-profit organisation Prize4Life. This database was created in 2012 and includes nearly 11,000 ALS (amyotrophic lateral sclerosis) patients from 23 phase 2 and 3 studies. For most trials, both treatment and control arm data are included, whereby trials generally "failed" (i.e. results were clinically and statistically not significant), with only one 'modestly effective' treatment currently available on the market. The database is open to anyone with an acceptable research proposal. A data access committee has agreed eligibility guidelines, and Prize4Life staff review individual requests according to these guidelines.

More examples on data sharing activities are provided in Section 4.2.2 of this report.

4.2. Veracity and variability

Clinical trial data are usually generated to support drug development, representing some of the most structured, well controlled, complete and reliable data available. To some extent, and because of regulation and guidelines, clinical trials are relatively homogeneous by structure and quality compared to other relevant sources of big data. However, to be able to bring together several clinical trials and promote the use of them in an efficient manner, various attempts to further develop standardisation have been introduced. These are described in detail in 5.2.1. Such activities are also a prerequisite for data sharing and the creation of big data repositories which consist of data from several clinical trials. Such activities are described in Section 5.2.2.

4.2.1. Data standardisation and harmonisation activities

During the last two decades, an increasing number of initiatives have been launched to increase standardisation and harmonisation of clinical trials and clinical trials data globally. Guidelines have been published to supervise and clarify requirements mainly for the industry and regulators (ICH 2017). The overall mission for these initiatives and activities are to improve the efficacy and safety of clinical trials and to optimise the regulatory process. Some of these initiatives are listed below.

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use)

ICH brings together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of drug registration. Since its inception in 1990, ICH has gradually evolved, to respond to the increasingly global face of drug development. ICH's mission is to achieve greater harmonisation worldwide to ensure that safe, effective, and high quality medicines are developed and registered in the most resource-efficient manner (http://www.ich.org/home.html).

CTTI (Clinical Trials Transformation Initiative)

The CTTI is a public-private partnership of over 80 members that strive to identify and drive adoption of practices that will increase the quality and efficiency of clinical trials, co-sponsored by the FDA. CTTI

was established in 2007 through a partnership between the FDA and Duke and is administered through the Duke Translational Medicine Institute. CTTI's approach includes conducting projects to better understand the range of current practices, assess alternative approaches, understand barriers to change, and propose recommendations for improvement (https://www.ctti-clinicaltrials.org/).

MRCT (Multi-Regional Clinical Trial Center)

The MRCT Center engages expert stakeholders from industry, academia, advocacy groups, nonprofits, and regulatory agencies to take on critical issues in the conduct and oversight of clinical trials. It is a neutral convening organisation associated with Brigham and Women's Hospital and Harvard University. Working in the pre-competitive space, their multidisciplinary teams collaborate to identify challenges and deliver ethical, actionable, and practical solutions for the global clinical trial enterprise, with a focus on emerging economies. Their efforts have resulted in the implementation of best practices, greater transparency, and improved safety for research participants (http://mrctcenter.org/).

CDISC (Clinical Data Interchange Standards Consortium)

CDISC is a global, open, multidisciplinary, non-profit organisation that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. Given that many trials are now multi-regional, the availability of globally accepted standards is very important. The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. CDISC standards are vendor-neutral, platform independent and freely available via the CDISC website (https://www.cdisc.org/).

CDISC develops data standards to foster smarter research and enable connections to healthcare by allowing data to speak the same language and providing common formats for data collection, data sharing and data analyses. The ultimate objective is to maximise the valuable information offered by patients participating in research studies around the globe, enabling researchers to discover new treatments, find breakthroughs, and unlock cures.

All CDISC standards (fundamentals standards and extended standard initiatives) and how they are connected appear from Figure 1 below, as presented at the CDISC webpage.
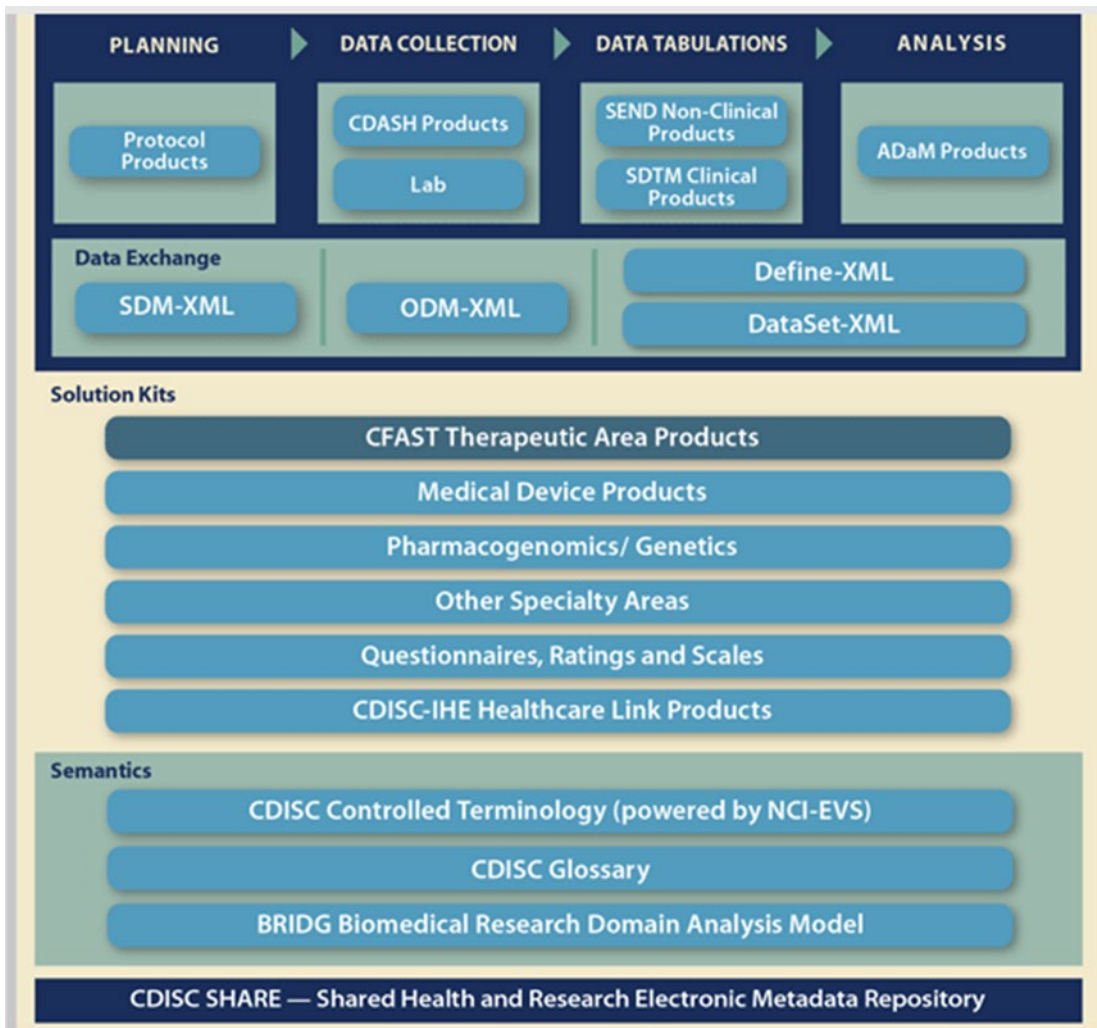
Figure 1. CDISC Standards (Standards | CDISC).

CDISC standards are required for regulatory submissions to the FDA in the USA and the PMDA in Japan, they are endorsed by the CFDA in China, and are requested to be used by the European Innovative Medicines Initiative (IMI). Currently in Europe, no single standard (such as CDISC) is endorsed, and currently the EMA does not require the use of CDISC.

As CDISC standards are continually refined this will lead to better transparency as data are submitted in a format which systems are able to easily recognise. In general, health data mainly consist of diagnosis, procedures, medicines and observations – and mainly focus on observations such as direct primary patient, meta-observations, context observations and analysis observations. CDISC data structures may eventually bridge the gap between the raw data and the structured clinical trial data. This gap could be bridged as various sources of data collection to analysis and reporting through regulatory submission and electronic data archive, based on standard data models and standards for exchange of nonclinical data and the clinical data acquisition standards harmonisation. It is expected that these developments will gradually benefit the standardisation between healthcare records and clinical trial data. One challenge is how these standards are interpreted by the industry in the context of their own level of standards – a challenge/difficulty in having a full proof concept of standardisation. This interpretation should be accurate for it to facilitate the availability of compliant standards. Overall, from the standardisation and harmonisation perspective, it would be desirable if the standards were as global as possible, and endorsing CDISC in Europe could therefore be a valid option to consider.

However, several other aspects need to be considered for such endorsement, and this is not within the mandate of the Clinical Trial and Imaging Subgroup.

4.2.2. Data sharing activities

The past few years have seen considerable interest in the sharing of patient-level data from clinical trials. These data sharing activities have been driven by the wish to permit activities ranging from verification of the original analysis to testing of new hypotheses. This interest has resulted in many debates and meetings, attention from the Institute of Medicine, proposed changes in journals policies and considerable effort from pharmaceutical sponsors and other groups to provide access to patient-level data.

The Wellcome Trust assessment report on access to clinical trial data identified 18 data sharing activities. These activities were grouped into the following five categories:

- Collaborative groups of trialists/trial sponsors (Critical Path (C-Path) Institute):

  - Research collaborations, rather than initiatives to enable broad data access.

  - Database staff harmonise the data on receipt.

  - Data providers retain control over their datasets and can veto requests for access.

- Disease-specific data repositories (PRO-ACT; database for ALS patients):

  - Created with the aim of accelerating development of treatments through enhanced data. Access for a wider research community.

  - Database staff harmonise the data on receipt.

  - Database staff grant access following guidelines agreed with the data provider.

- Public-funder mandated repositories (publicly funded research e.g. NIH):

  - Created as a platform for depositing data and often linked to other types of data (genetic data, observational studies).

  - Some databases require the data provider to standardise data to their requirements prior to submission, others leave this to the user of the repository.

  - Database staff grant access (administrative).

- Commercial trial repositories and data portals (ClinicalStudyDataRequest):

  - Created as a platform to allow access to data from commercial clinical trials.

  - Standardised according to company data standards.

  - Access is granted by an independent review board following an agreed application process. Companies can retain a right to deny access.

 Open data sharing by individual research groups/units (FREEBIRD):

  - Data available for download to allow broad access to individual patient data without the need to contact the original researchers.

  - Datasets are currently held on many different platforms, in distributed locations. Therefore, researchers may need support to be able to find and combine these to maximise data use.

  - Part of the dataset may be withheld from the open access (e.g. the randomisation code).

The activities representing all five categories described by the Wellcome Trust assessment report were selected and characterised further. The activities that have been identified are shown in Table 2. Each activity has been characterised in terms of accessibility, content, usability, sources and analytics.

Table 2: Clinical Trial Data Sharing Activities

Name	Accessibility	Level of processed data	Clinical information	Usability	Sources	Analytics
BioLINCC – National heart, lung, and blood institute (NHLBI) (public-funder mandated)	Researchers have to provide an IRB approval for any level of access to data. Administrative review by allocation committee	Individual patient data	Data from 82 clinical trials and 33 observational studies	Data are downloaded in the format they receive it. No customisation of data	Academic, government	Data files
ClinicalStudyDataRequest (Commercial trial repositories)	Approval by an independent review panel. Some sponsors may decline access to their data	Individual patient data	3049 trials performed by 13 pharmaceutical companies	Sponsors provide access to data in a private workspace. Controls to prevent researches from downloading data to their computer	Industry sponsors	SAS and R analyses can be downloaded
Critical Path Institute Data Collaboration Center (Collaboration of trialists/disease specific)	Mainly for internal use within consortiums except Alzheimer's database which is available for external researchers	Individual patient data	More than 50,000 patients from 88 trials, covering Alzheimer's, Parkinson's, kidney diseases, multiple sclerosis and tuberculosis	Data are loaded into an online data repository	Academic and industry sponsors	Analysis data extracts are provided to the consortium
Early breast cancer trialists collaborative group (EBCTCG) (Collaboration of trialists)	Access to datasets in the database to conduct own analyses, but approval must be sought from data owner before transfer	Individual patient data	Data from around 700 clinical trials in breast cancer	Data are downloaded in a structured format	Academic and industry sponsors	
IMPACT (International Mission on Prognosis and Analysis of Clinical Trials) (Collaboration of trialists)	Mainly for internal use within the group	Individual patient data	Data from 12 randomised clinical trials and 5 observational Studies in the field of traumatic brain injury	Data merge into one dataset	Academic and government	
Industry sponsors home page (Commercial	Approval by an independent review panel.	Individual patient data	Data from industry sponsored studies. E.g.	Sponsors provide access to data in a	Industry sponsors	

Name	Accessibility	Level of processed data	Clinical information	Usability	Sources	Analytics
trial repositories)	Some sponsors may decline access to their data		Merck and Bristol-Myers Squibb (SOAR initiative). After regulatory approval of a product	private workspace. Controls to prevent researches downloading data to their computer		
NIDDK – National Institute for Diabetes, digestive and kidney Diseases (public-funder mandated)	Researches have to provide an IRB approval for any level of access to data. Administrative review by allocation committee	Individual patient data	Data from 43 clinical trials	Data are downloaded in the format they receive it (usually SAS)	Academic and government	SAS
PRO-ACT (Disease specific)	Approval by a data access committee. Based on eligibility guidelines	Individual patient data	Nearly 11,000 ALS patients from 23 phase 2/3 trials	Download	Academic and industry sponsors	Excel or text files
Project Data Sphere (Collaboration of trialists/disease specific)	Open Sign up Free of charge	Individual patient data	More than 62,000 patients from 89 phase 3 oncology trials, covering multiple tumour types	Download, share, integrate, and analyse patient-level data	Academic, government, and industry sponsors	SAS
Sylvia Lawry Centre (Disease specific)	Full dataset only accessed locally. Open access to synthetic dataset	Individual patient data	More than 26,000 patients from 29 trials in Multiple Sclerosis	Download Synthetic dataset	Academic and industry sponsors	Internal developed Online Analytic Processing tools
TransCelerate Biopharma INC	Requires membership in the network	Data from industry-sponsored studies	TransCelerate provides access to data	Individual patient data. Mostly standard of care and placebo groups.	Industry sponsors	
Vivli (Collaboration of trialists; launches in July 2018)	Approval by a data access committee	Individual patient data	Not disease, country, sponsor, funder specific.	Data and metadata can be identified, hosted, shared and analysed	Academic, government, and industry sponsors	Vivli analytical tools
WWARN (Worldwide Antimalarial Resistance Network) (Collaboration of trialists)	Data ownership remains with the primary data provider. Mainly for internal use within the network. Data access for external	Individual patient data	Data from approximately 100,000 individual patients, generated in 350 clinical trials (phases 2-4).	Transforms submitted data to a common format and re-analyses the study. Data mining system to identify the best	Academic, government and industry sponsors	

Name	Accessibility	Level of processed data	Clinical information	Usability	Sources	Analytics
	research groups with specific research question			data to use. The data may not be used for commercial purposes		
YODA (Yale University Open Data Access) (Commercial trial repositories)	Approval by a review board	Individual patient data	Data from over 250 clinical trials	Sponsors provide access to data in a private workspace. Controls to prevent researches downloading data to their computer Data from rhBMP-2 trials can be downloaded	Industry sponsors (Johnson & Johnson and Medtronic)	SAS, R and other software on request

The EMA developed external guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use referred to as Policy 0070 (EMA 2017). The scope of phase 1 of Policy 0070 covers clinical study reports for all medicines submitted to the EMA to support a request for a marketing authorisation under the centralised procedure from January 2015 (EMA 2017). In addition, from July 2015, data submitted to the agency as part of a new indication or line extension application relating to existing centrally authorised products also fall under the scope of Policy 0070. Importantly the policy applies to all clinical data regardless of the outcome of the regulatory procedure. Phase 2 of the policy has not yet started, but it will review available options for sharing of individual patient level data (IPD). Phase 1 of Policy 0070 has required the creation of a dedicated portal (clinical data.ema.europa.eu) where data are shared in an open access manner with minimal restrictions (EMA 2017); hence data must be fully anonymised before being hosted on the site. In order to support marketing authorisation holders (MAHs) and following consultation with the European Data Protection Service (EDPS), the EMA provided guidance on the anonymisation of clinical reports which highlights to MAHs/applicants the available techniques and those it considers most relevant, but stops short of mandating any specific methodology.

Despite the availability of data sharing platforms, the requests of data and publications of studies based on these data are few (Sydes, Johnson et al. 2015). The process to access the data differs between the different initiatives. In most cases, the research plan should be approved by a committee before access to the data is given. For transparency reasons, these committees are usually, independent from the companies providing the data. However, difficult access procedures based on unclear criteria may present barriers to some qualified requesters. Another issue that limits the use of these data is related to the anonymisation of the trial participants. It is our understanding that the Data Protection Directive in the EU does not provide explicit guidelines for how data should be protected through anonymisation (Lo 2015). De-identifying data does not completely eliminate all risk of re-identification. Reducing that risk to zero by collapsing cells with few individuals may make the data worthless. Therefore, a balance between research objectives and patient protection should be facilitated by, for example, providing safe environments to perform the calculations and reviewing the results. The costs of and workload needed to keep the data sharing platforms up to date with emerging

computer tools to improve data protection are increasing continuously. Since the complexity of these issues is enormous, all stakeholders should be involved to find a sustainable solution.

4.3. Velocity

Velocity, defined as speed of change and/or rate of accumulation of data, is not relevant in the analysis of traditional single clinical trials but is relevant in the context of the data sharing platforms as more data accumulate within a specific disease or for controls. The relevance of velocity may vary from one platform to another, depending on the rate of accumulation. However, velocity is potentially most relevant in the context of wearables or other censor data which are increasingly being used in clinical trials. This aspect is not addressed in this report.

4.4. Value

Data standardisation and harmonisation activities require constant coordination, however, the concepts and platforms etc. are generally available and ready to use. From the point of view of data accessibility, most of the clinical trial data remain confidential (at least partly because of commercial confidentiality and intellectual property rights), and they are not open to public or research communities.

Clinical trials have intrinsic limitations that cannot be fully overcome by data sharing and harmonisation of standards. Some of these limitations concern the use of surrogate endpoints (especially when not properly validated), restricted follow-up time, missing data issues, possible lack of relevant comparators in clinical trials etc., which will, at least partly, remain even after data from several trials are combined. In addition, a robust validation process of results from pooling and aggregation across trials has to be developed.

Despite these limitations, clinical data sharing has incredible potential to improve medical research (see Section 4.4.1). Clear advantages of data sharing can be seen for example in the field of Alzheimer's disease in which data sharing could enable better understanding of trial failures and re-direct efforts. However, regulatory acceptability is still somewhat limited and may require further actions (see Section 4.4.2). Furthermore, clinical data sharing allows IPD meta-analysis which may provide clear advantages compared to meta-analysis based on summary statistics. Examples of such approaches are provided in Kawahara et al. 2018 including a list of approved research proposals for meta-analysis in the Clinical Study Data Request (Kawahara, Fukuda et al. 2018).

4.4.1. Applicability and acceptability of data sharing and harmonisation activities

Currently, it is mostly only trials showing positive results which are published in scientific journals (Kien, Nußbaumer et al. 2014) while those that fail, the great majority, are rarely being published. The Biotechnology Innovation Organization (BIO) reported in 2016 that approximately 70 % and 40 % of Phase II and III trials fail respectively; and only 10 % of drug development programs successfully make it to market (Organization 2016). There are of course multiple reasons why clinical trials fail but it is reasonable to assume at least some of them could be addressed by proactive and non-selective sharing of clinical trial data. A list of applications of clinical data sharing is presented in Table 3.

Table 3: List of applications of historical clinical trial data

Areas of current/potential applications	Benefits
Control arm substitution with historical	Fewer recruited patients. Increased statistical

Areas of current/potential applications	Benefits
controls	power.
Standard of care arm substitution	Fewer recruited patients. Increased statistical power
Comparison between the trial and the real-world patient population	Assessment of the external validity of the trial
Sample size calculations	Estimation of the target difference for a randomised controlled trial
Improvement in handling of missing data	Identification of factors related to drop out in clinical studies
Reproducibility of the findings	Confirmation of the results by independent sources
Biomarker development	Identification and confirmation of biomarkers
Predictive models	Identification of patient population, improved description of the expected response
New hypotheses	Facilitation of novel discoveries
Safety information	More accurate safety information, including molecules not currently marketed
Educational data sets	Training data sets, testing of novel statistical methods and harmonisation of statistical methods

In some specific cases, pragmatic or "less orthodox" clinical trials could be an alternative or a complement to traditional clinical trials with randomised treatment arms. For example, historical data from previous trials could be used to avoid or reduce the need to recruit patients to the control or the standard care arm. Miyamoto et al. 2013 conducted a pilot study on the combination of bortezomib and high-dose melphalan as conditioning regime following stem cell transplant in patients diagnosed with multiple myeloma (Miyamoto, Yoshimoto et al. 2013). The investigators used historical controls in which participants had received high-dose melphalan only followed by stem cell transplant (SCT).

During the marketing authorisation process, the sponsor presented a matched pair analysis to provide further evidence in support of the benefit of bortezomib in combination with dexamethasone in relapsed or refractory multiple myeloma compared to bortezomib monotherapy. In this analysis, data from three previous clinical studies were combined to create statistically matched pairs between the bortezomib arm and that receiving bortezomib in combination with dexamethasone arm (Orlowski, Facon et al. 2013, Dimopoulos, Orlowski et al. 2015). Based on this analysis, Bortezomib can be used in multiple myeloma in combination with dexamethasone (or as monotherapy) without a formal trial. Findings from these analyses were similar to those obtained in an observational study performed afterwards (Fourrier-Reglat, Noize et al. 2014).In addition to the ethical advantages of such approach, the re-use of data reduces expenses due to fewer patients and less monitoring. The disadvantage of this approach is that the study is conducted without treatment randomisation and thus could suffer from treatment selection bias.

A randomised clinical trial is optimised to ensure the internal validity of the trial, i.e. in order to show that there is a treatment effect. The population included in the trial differs from real-world patients in many aspects; the clinical trial populations are generally healthier, younger and have fewer comorbidities than ordinary patients. Therefore, it is plausible to expect somewhat better results in the comparator (standard of care or placebo) arm of a clinical trial than those observed in patients treated in clinical practice. The performance of comparator arm data from different trials could be compared with real-word data sources from patient registries or electronic healthcare records to provide direct information about the external validity of the trial.

Sample size calculations performed during the planning phase of the trial will determine the number of patients included in the trial. The most important piece of information for the calculation is the

specification of the target difference, which is the minimum difference in treatment benefit that has to be observed between the treatment arm and the comparator arm to claim that there is a clinically relevant treatment effect. In many cases, the target difference appears to be determined in an informal basis (Fayers, Cuschieri et al. 2000). A comparison, as described above, between previous control arms and real-world patients selected using the same inclusion/exclusion criteria allows a better estimation of the effect of the current standard of care in healthier patients. Such information is valuable when assessing the minimum effect that has to be achieved in the trial, and this will consequently help to power the trial correctly.

Handling of intercurrent events and missing data is a critical issue that could jeopardize the interpretation of the results of a trial. Even though several statistical techniques have been developed to minimise bias in the results due to missing data, it is highly desirable to reduce the number of patients leaving the study earlier than planned. Investigation of risk factors related to drop out is therefore of paramount importance. Data from previous clinical trials are useful to identify factors that may increase drop out in a patient population. Krishnan et al. presented a study based on historical clinical trial data that identified factors related to loss-to-follow up in HIV trials (Krishnan, Wu et al. 2011).

The identification and confirmation of biomarkers is another area where clinical data sharing could provide added value since these studies require large amounts of data. As an example, the Biomarkers Consortium of the Foundation for the National Institutes of Health organised a collaboration project between several pharma companies to investigate whether adiponectin is a predictor of glycaemic response to peroxisome proliferator–activated receptor agonist drugs used in the treatment of diabetes. This hypothesis was successfully confirmed in 2009 by the Consortium using data from 2,000 patients recruited in several clinical trials (Medicine 2013).

Further possible applications for the combination of clinical trial data include predictive models for patient recruitment and efficacy estimations. During the planning phase of a trial, predictive models help to identify which patients may benefit most from a certain therapy before conducting any study. Importantly data sharing also allows other researchers to explore the data and to test new hypotheses. The combination of several trials makes it possible to tailor the characteristics of the patients included in the data analysis to answer other questions than those proposed by the original investigators. In 2016, the New England Journal of Medicine organised the SPRINT Data Analysis Challenge, in which people from all over the world were encouraged to reanalyse data from the Systolic Blood Pressure Intervention Trial (SPRINT) to identify novel biological or clinical findings (Burns and Miller 2017). The winner of the competition was a research group that had developed a weighted risk-benefit calculator for examining advantages and disadvantages of intensively treating an individual patient with hypertension.

Another application of clinical data sharing is the improvement of the reporting of adverse events. Combining several trials may enable adverse events to be reported earlier, as well as potentially identify risk factors for the occurrence of adverse drug reactions (ADRs) in specific patient populations or help define underlying mechanisms. Pharmacovigilance-related databases mostly include marketed products. The collection of clinical trial data in a uniform system would facilitate the analysis of adverse reactions observed in the clinical trials, and information about the safety of experimental drugs could thus be improved.

Finally, historical clinical data could be used for educational purposes, testing of novel statistical methods and harmonisation of applications of statistical methods across different regulatory agencies. Historical clinical data represent one of the best sources of material for training since they are real-world cases.

4.4.2. Potential for improving regulatory applicability and acceptability of data sharing activities

Despite several potential applications of data sharing activities in medical research, some gaps exist and further actions could be taken to improve regulatory applicability and acceptability of such activities and sharing of clinical trial data.

Firstly, patient level data are not currently being requested for storage and further analysis by the European regulatory network. Currently, the EMA does not require the submission of the original clinical trial data for assessment during review of the marketing application, and hence the assessment is based on the results provided by the sponsor. Individual patient data (IPD) from clinical trials should be requested as part of the MAH submission and assessed as part of the review of the marketing authorisation application. This would require establishing the legal basis for requesting IPD and agreeing on data format for submissions. Furthermore, building a data repository by systematically requesting original individual patient level data could enable meta-analyses, product class comparisons and indirect comparison of closely related medicinal products, identification of safety signals, etc. Consequently, requesting clinical trial data systematically could provide some further regulatory value of clinical trial data beyond the replication of the submitted analyses. However, implementing such actions requires increased statistical capacity in the NCAs and other regulatory agencies. In addition to increased requirements for statistical capacity in the NCAs, this would also require infrastructural changes, including funding thereof.

Secondly, there are several challenges to be addressed from a regulatory perspective when utilising, partially or completely historical clinical trial data in the trial design. As mentioned in the previous sections, a key issue is the lack of standards for study protocols and variable definitions.

Comprehensive and overarching European guidelines describing the regulatory acceptability of clinical data standardisation, data sharing and statistical methodology are needed. Errors due to misinterpretation of the harmonised or pooled data may occur if the data are analysed by people not involved in the original study. CDISC (see Section 4.2.1) may probably be the most developed approach to standardise clinical data collection and to develop platform-independent data standards (CDISC standards are required for regulatory submission to the FDA in the USA and the PDMA in Japan). Overall, from the standardisation and harmonisation perspective, it would be desirable if the standards were as global as possible, and endorsing CDISC in Europe could therefore be an option to consider. Furthermore, there is a need to provide guidance for both regulators and industry on the current legislation regarding informed consent and data protection in the context of clinical data.

Thirdly, it should be noted that there is insufficient data available in some disease areas, especially in rare diseases or high-risk populations. Hence, data sharing in the area of (ultra)-rare diseases could particularly inform scientific advice procedures and ultimately regulatory decision-making. Supporting any actions to create data platforms for proactive sharing of clinical data related to "ultra-rare diseases" is of value.

Fourthly, since the statistical methods employed to analyse large amounts of data generally differ from those traditionally applied in the regulatory process, the EMA and the NCAs need to further educate the assessors in this matter. To support assessments based on causality, new statistical methodology and tools/applications may need to be developed (multidimensional/multi-criteria-decision analyses). Harmonisation at the European level regarding acceptable methodology for big data analyses for regulatory decision-making is needed.

In addition to greater efforts into data sharing and harmonisation, a more exhaustive use of single clinical trial data may provide additional value. For example, more extensive use of images, genomic data, signal/censor data collected during a single clinical trial combined with big data analytics (especially machine learning) could provide several new insights and opportunities for the analysis of clinical trials. Such data and related analytics could lead to new hypothesis generation, better definition of responses, new type of outcomes and better knowledge of characteristics of patients, and eventually to more effective treatments for better selected patients. It seems likely that from the regulatory perspective such data linkage could lead to an added value. This area currently lacks overarching guidance, and regulatory acceptability may be rather unpredictable.

5. Imaging

Medical imaging has enormous potential for early disease prediction, but is impeded by the difficulty and expense of acquiring data sets before symptom onset (Miller, Alfaro-Almagro et al. 2016). Furthermore, imaging (when linked to clinical data) has several applications related to drug-development, disease target identification, biomarker development etc. Fields of imaging applications include domains from radiology, pathology, ophthalmology and neuroscience, just to name a few.

Medical imaging is widely used in clinical trials. There is an enormous unexploited potential in these data, which are currently not routinely analysed or explored. Such analysis may require big data analytics. Reading images is a very complex and time-consuming process. Development of automated procedures to analyse images is currently ongoing in many fields (Schuster, Hardiman et al. 2016, Trebeschi, van Griethuysen et al. 2017). This process may facilitate the combination of images for the same and several patients to add further information to the analysis of traditional endpoints such as progression free survival.

In the field of imaging, there are also several initiatives, such as the UK Biobank, which incorporate high-quality imaging data collected outside the clinical trials in the context of a longitudinal cohort study. Such projects aim to tackle the problems related to accessibility of imaging data and the ethical challenges associated with sharing images in addition to the challenges of integrating their analysis with electronic health records, genomics and other complex phenotyping (see more at http://www.ukbiobank.ac.uk/).

Due to the promising applications in large-scale imaging analysis in the regulatory context imaging should be considered much more extensively in a separate context.

6. Discussion and conclusions

In this report, the Clinical Trial and Imaging Subgroup presented information about current data sharing activities for clinical trials and the opportunities and challenges of using clinical trial big data for regularity decision-making.

Our main findings are:

1) Clinical trial data sharing activities are currently relatively mature, but they are already providing access to many thousands of clinical trials. However, it is not fully clear nor understood how these activities could be applied in regulatory or scientific procedures. Actions to create open data platforms for proactive sharing of clinical data should be supported. In particular, efforts targeting diseases with scarce data, such as "ultra-rare diseases" or high-risk populations should be promoted.

2) Individual patient data (IPD) from clinical trials should be requested as part of the MAH submission and assessed as part of the review of the marketing authorisation application. This

would require establishing the legal basis for requesting IPD and agreeing on data format for submissions. Such actions would have impact on resource use in NCAs.

3) Data standardisation activities are critical to ensure usability and applicability of data. Clinical trial standards should be as globally aligned as possible.

4) Anonymisation in data sharing activities is a difficult balance between data utility and data privacy.

5) Encouraging more extensive or exhaustive use of images, genomic data, data from wearables collected during a single clinical trial combined with big data analytics could benefit drug discovery and may further inform regulatory decisions. Regulatory Guidance should be developed to efficiently use these data in the regulatory processes.

Data standardisation and harmonisation activities are a prerequisite for efficient use of shared data. Standardisation may also gradually increase compatibility of healthcare records and clinical trial data, which enables new applications of combining data sources (bridging the gap between clinical trial data and other data sources). In the USA, sponsors are required to present the data from the clinical studies when applying for a marketing authorisation. CDISC is the accepted format for submissions to the FDA. In Europe, phase 2 of Policy 0070, which requires the submission of part of the clinical data during the marketing authorisation application for centralised procedures, has not yet been implemented. Thus, the required format for the submission to the EMA has not been presented in this report.

Potential applications of clinical data sharing include enhanced use of historical data, more efficient identification of target populations and outcomes, improvements in missing data handling, new opportunities for biomarker and predictive model development, possibilities to generate new hypotheses and new options to identify safety signals. Educational applications can also be derived from shared data. Some data privacy issues related to sharing individual data exist, and anonymisation approaches and access mechanisms should be carefully considered. Despite their huge potential in medical research, clinical trial data sharing activities have currently limited applicability in the regulatory setting. We believe that regulatory value could be improved by requesting individual patient level data and setting up a repository for clinical trials used in regulatory submissions. Since 2015, Policy 0070 of the EMA has required the submission of the clinical study reports. In the next phase of implementation of Policy 0070, the publication of individual patient data will be required. Discussions are currently ongoing on how to combine strong data protection regulations, protection of commercial interests and to increase transparency by allowing external parties access to data used for assessment by the EMA. Furthermore, overarching guidelines on applications of big data from clinical trials in regulatory settings and data standards are needed to promote incentives to use big data in regulatory decisions.

One of the very promising applications of big data from clinical trial is when clinical trial data are combined with other sources of big data, such as electronic healthcare records (EHR). For example, if the patient's healthcare history is available in an EHR, those records could be combined with those collected during the trial when the patient was using wearables and other sensors and devices. Extensive analyses of imaging data are also expected to provide valuable information to study the effect of a medicinal product on the disease course. These aspects of big data need further consideration to fully assess the opportunities and challenges of their implementation in the regulatory context.

7. Sources

Literature

Babre, D. K. (2013). "Clinical data interchange standards consortium: A bridge to overcome data standardisation." Perspectives in Clinical Research 4(2): 115-116.

Bertagnolli, M. M., O. Sartor, B. A. Chabner, M. L. Rothenberg, S. Khozin, C. Hugh-Jones, D. M. Reese and M. J. Murphy (2017). "Advantages of a Truly Open-Access Data-Sharing Model." New England Journal of Medicine 376(12): 1178-1181.

Burns, N. S. and P. W. Miller (2017). "Learning What We Didn't Know - The SPRINT Data Analysis Challenge." N Engl J Med 376(23): 2205-2207.

Dimopoulos, M. A., R. Z. Orlowski, T. Facon, P. Sonneveld, K. C. Anderson, M. Beksac, L. Benboubker, H. Roddie, A. Potamianou, C. Couturier, H. Feng, O. Ataman, H. van de Velde and P. G. Richardson (2015). "Retrospective matched-pairs analysis of bortezomib plus dexamethasone versus bortezomib monotherapy in relapsed multiple myeloma." Haematologica 100(1): 100-106.

EMA (2016). Adaptive pathways workshop briefing book (EMA/693142/2016v).

EMA (2016). Guidance for companies considering the adaptive pathways approach (EMA/527726/2016).

EMA (2017). "External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use."

EMA (2017). Questions and Answers (Q&As) on the External Guidance of Policy 0070 on Clinical Data Publication (CDP) (EMA/14227/2017 Rev. 1).

Fayers, P. M., A. Cuschieri, J. Fielding, J. Craven, B. Uscinska and L. S. Freedman (2000). "Sample size calculation for clinical trials: the impact of clinician beliefs." British Journal of Cancer 82(1): 213-219.

Fourrier-Reglat, A., P. Noize, T. Facon, J. P. Fermand, O. Fitoussi, G. Marit, P. Thomare, P. Robinson, E. Bignon, J. Jove, R. Lassalle, M. Rouyer, A. Grelaud and N. Moore (2014). "Real-life patterns of use and effectiveness of bortezomib: the VESUVE cohort study." Leuk Lymphoma 55(4): 848-854.

ICH (1998). ICH Topic E 9: Statistical Principles for Clinical Trials.

ICH (2017). ICH E8(R1): Revision of General Considerations for Clinical Trials.

Kawahara, T., M. Fukuda, K. Oba, J. Sakamoto and M. Buyse (2018). "Meta-analysis of randomized clinical trials in the era of individual patient data sharing." Int J Clin Oncol 23(3): 403-409.

Kennedy-Martin, T., S. Curtis, D. Faries, S. Robinson and J. Johnston (2015). "A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results." Trials 16: 495.

Kien, C., B. Nußbaumer, K. J. Thaler, U. Griebler, M. G. Van Noord, P. Wagner, G. Gartlehner and U. P. C. on behalf of the (2014). "Barriers to and facilitators of interventions to counter publication bias: thematic analysis of scholarly articles and stakeholder interviews." BMC Health Services Research 14: 551.

Krishnan, S., K. Wu, M. Smurzynski, R. J. Bosch, C. A. Benson, A. C. Collier, M. K. Klebert, J. Feinberg and S. L. Koletar (2011). "Incidence rate of and factors associated with loss to follow-up in a longitudinal cohort of antiretroviral-treated HIV-infected persons: an AIDS Clinical Trials Group (ACTG) Longitudinal Linked Randomized Trials (ALLRT) analysis." HIV Clin Trials 12(4): 190-200.

Lo, B. (2015). "Sharing clinical trial data: maximizing benefits, minimizing risk." Jama 313(8): 793-794.

Lo, B. and D. L. DeMets (2016). "Incentives for Clinical Trialists to Share Data." New England Journal of Medicine 375(12): 1112-1115.

Medicine, I. o. (2013). Sharing Clinical Research Data: Workshop Summary. Washington, DC, The National Academies Press.

Miller, K. L., F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. Andersson, L. Griffanti, G. Douaud, T. W. Okell, P. Weale, I. Dragonu, S. Garratt, S. Hudson, R. Collins, M. Jenkinson, P. M. Matthews and S. M. Smith (2016). "Multimodal population brain imaging in the UK Biobank prospective epidemiological study." Nat Neurosci 19(11): 1523-1536.

Miyamoto, T., G. Yoshimoto, T. Kamimura, T. Muta, S. Takashima, Y. Ito, M. Shiratsuchi, I. Choi, K. Kato, K. Takenaka, H. Iwasaki, Y. Takamatsu, T. Teshima and K. Akashi (2013). "Combination of high-dose melphalan and bortezomib as conditioning regimen for autologous peripheral blood stem cell transplantation in multiple myeloma." Int J Hematol 98(3): 337-345.

Organization, B. I. (2016). Clinical Development Success Rates 2006-2015- The Biotechnology Innovation Organization. T. B. I. Organization.

Orlowski, R. Z., T. Facon, P. Sonneveld, K. C. Anderson, M. Beksac, L. Benboubker, H. Roddie, A. Potamianou, C. Couturier, H. Feng, O. Ataman, H. van de Velde and P. G. Richardson (2013). "Retrospective Matched-Pair Analysis Of The Efficacy and Safety Of Bortezomib Plus Dexamethasone Versus Bortezomib Monotherapy In Patients (Pts) With Relapsed Multiple Myeloma (MM)." Blood 122(21): 3177-3177.

Rockhold, F., P. Nisen and A. Freeman (2016). "Data Sharing at a Crossroads." New England Journal of Medicine 375(12): 1115-1117.

Rockhold, F. W. (2017). "Statistical controversies in clinical research: data access and sharing-can we be more transparent about clinical research? Let's do what's right for patients." Ann Oncol 28(8): 1734-1737.

Schuster, C., O. Hardiman and P. Bede (2016). "Development of an Automated MRI-Based Diagnostic Protocol for Amyotrophic Lateral Sclerosis Using Disease-Specific Pathognomonic Features: A Quantitative Disease-State Classification Study." PLoS One 11(12): e0167331.

Sydes, M. R., A. L. Johnson, S. K. Meredith, M. Rauchenberger, A. South and M. K. Parmar (2015). "Sharing data from clinical trials: the rationale for a controlled access approach." Trials 16: 104.

Trebeschi, S., J. J. M. van Griethuysen, D. M. J. Lambregts, M. J. Lahaye, C. Parmar, F. C. H. Bakers, N. H. G. M. Peters, R. G. H. Beets-Tan and H. J. W. L. Aerts (2017). "Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR." Scientific Reports 7: 5301.

Varnai, P., M. Rentel, P. Simmonds, T. Sharp, B. Mostert and T. de Jongh (2014). Assessing the research potential of access to clinical trial data. A report to the Wellcome Trust. Study led by Technopolis Group (UK).

Warren, E. (2016). "Strengthening Research through Data Sharing." New England Journal of Medicine 375(5): 401-403.

Links

Accessed on 27-06-2018

https://eudract.ema.europa.eu/

https://www.clinicaltrials.gov/

http://www.ich.org/home.html

https://www.ctti-clinicaltrials.org/

http://mrctcenter.org/

https://www.cdisc.org/

http://www.hl7.org/

https://www.cdisc.org/standards/domain-information-module/bridg

"A global clinical research network: moving beyond harmonisation", Koski G, Widler B, Olmsted A, Clinical Trials Network, Regulatory Affairs Pharma, July 2011
http://www.appliedclinicaltrialsonline.com/data-standards-harmonization

"Data standards Harmonization – A look at the past, present and future of interoperability across the clinical research spectrum", Mcllwain J, Applied Clinical Trials, May 2010
http://www.appliedclinicaltrialsonline.com/data-standards-harmonization

http://www.cochranelibrary.com/

http://www.ukbiobank.ac.uk/