In principle accepted 12/20/2019 by BMC Medicine

Registered Report:

Data-sharing and re-analysis for main studies assessed by the European Medicines Agency - a cross-sectional study on European Public Assessment Reports

Authors:

Maximilian Siebert (1,2), Jeanne Gaba (1,2), Alain Renault (1), Bruno Laviolle (1), Clara Locher (1), David Moher (3), Florian Naudet (1)

Affiliations:

1. Univ Rennes, CHU Rennes, Inserm, CIC 1414 [(Centre d’Investigation Clinique de Rennes)], F-35000 Rennes, France
2. Univ Rennes, CHU Rennes, REPERES [(Recherche en Pharmaco-épidémiologie et Recours aux Soins)], EA 7449, 35000, Rennes, France
3. Center for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

Email addresses:

Maximilian.siebert@univ-rennes1.fr, fabiolagaba@gmail.com, alain.renault@chu-rennes.fr,
bruno.laviolle@chu-rennes.fr, Clara.LOCHER@chu-rennes.fr, dmoher@ohri.ca,
floriannaudet@gmail.com
Corresponding Author: Dr Florian Naudet, Clinical Investigation Center (INSERM 1414) and Adult Psychiatry Department, Rennes University Hospital, Rennes, France, +33 6 80 36 26 79, floriannaudet@gmail.com

Abstract:

Transparency and reproducibility are expected to be key features of randomized controlled trials (RCTs) used for decision-making on Marketing Authorizations for new medicines. This registered report introduces a cross-sectional study aiming to assess inferential reproducibility (i.e. Individual Patient Data is available and qualitatively similar conclusions can be drawn from a reanalysis of the original trials) for main RCTs assessed by the European Medicine Agency (EMA). 2 researchers will independently identify all main studies on new medicines and biosimilars given approval by the European Commission after 1st of January 2017 marked as main studies in the European Assessment Reports. 62 of these studies will be randomly sampled ensuring a precision of ± 12% to estimate our primary outcome, the proportion of studies where the conclusions are reproduced. Then, one researcher will retrieve the Individual Patient Data (IPD) for these studies and other necessary documents for reanalysis by contacting the study sponsors, and if necessary the EMA. For each study he will prepare a dossier containing the IPD, the protocol and information on the conduct of the study conduction. A second researcher who will have no access to study reports (including publications) or code, will use the dossier to run an independent re-analysis of each trial. All results of these re-analyses will be reported in terms of each study’s conclusions, p-values, effect sizes and changes from the initial protocol. Then a team of two researchers not involved in the re-analysis will compare results of the reanalyzes with published results of the trial. In case of conflicting results, a statistician will re-analyze the data independently before concluding to lack of reproducibility. This registered report is part of a wider project called “Reproducibility in Therapeutic Research”, funded by the Agence Nationale de la Recherche and a complementary registered report aims to assess the reproducibility of trials available on data sharing platforms.
Keywords: Reproducibility of results, Clinical Trial, Drug Approval

List of abbreviations:

AdAM: Analysis Data Model
BMJ: The British Medical Journal
CDASH: Clinical Data Acquisition Standards Harmonisation
CDISC: Clinical Data International Standard Consortium
CHMP: Committee for Medicinal Products for Human Use
CI: Confidence Interval
EFPIA: European Federation of Pharmaceutical Industries and Associations
EMA: European Medicines Agency
EPAR: European Public Assessment Report
EU: European Union
EudraCT: European Union Drug Regulation Authorities Clinical Trials
ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
ICTRP: International Clinical Trials Registry Portal
IPD: Individual Patient Data
ISRCTN: International Standard Randomised Controlled Trial Number
MAA: Marketing Authorization Application
Background:

The influence of main trials (i.e. evidence used for drug marketing approval) as assessed by the European Medicine Agency (EMA) is paramount. These studies have a major impact on drug Marketing Authorizations and can change the practices of European medical practitioners and the care offered to millions of patients in the European Union. Because of the major financial conflicts of interest inherent in the evaluation of pharmaceuticals (1, 2), stakeholders are typically more confident when the results and conclusions of these studies can be verified. For a long time, however, transparency was lacking and the individual patient data (IPD) and accompanying material (code, protocol, data analysis plan, etc.) to reproduce these analyses were unavailable. An empirical analysis suggests that only a small number of re-analyses of randomized controlled trials (RCTs) have been published to date; of these, only a minority were conducted by entirely independent authors (3). Data-sharing enabling such re-analyses is being increasingly mandated in medicine.

And indeed, the EMA aimed to pioneer transparency in this field when, in November 2010, it decided to share every piece of documentation received, in the wake of the first version of policy 0043 (4). As part of its transparency policy, the EMA publishes European Public Assessment Reports (EPAR) after the European Commission’s decision on the specific medicines. These reports include, amongst other documents, results of main trials (5). On October 2nd 2014 the EMA released its policy 0070 on
“publication of clinical data for medicinal products for human use”(6). The agency describes a two-step approach. From 1st of January 2015 clinical reports on medicines submitted for Marketing Authorization have been published. A second step includes the publication of IPD. A date for the implementation of this step still needs to be fixed. However, as a result of Brexit and the relocation of the EMA to the Netherlands, further developments and renovation have been stopped for the moment (7, 8). Efforts are therefore still needed to reach full transparency in the EMA.

On the other hand, biopharmaceutical companies (i.e. Pharmaceutical Research and Manufacturers of America [PhRMA] and the European Federation of Pharmaceutical Industries and Associations [EFPIA] endorsed a commitment ‘to enhancing public health through responsible sharing of clinical trial data’ in a manner that is consistent with 3 main principles: safeguarding the privacy of patients, respecting the integrity of national regulatory systems, and maintaining incentives for investment in biomedical research (9). Despite this commitment from 2013, an audit found that data availability was reached for only 9/61 (15%) clinical trials on medicines sponsored by the pharmaceutical industry and first published between 1 July 2015 and 31 December 2015 in the top 10 journals of general and internal medicine (10). If such low rates of data-sharing were also to be observed for main trials, it would invalidate any efforts towards reproducibility for these important studies.

However, the environment for data-sharing is changing fast. And indeed, data-sharing Platforms like ViVli, YODA project or Clinical Study Data Request are more and more widely used. In fall of 2019 these platforms gathered a large number of trials sponsored by the pharmaceutical industry. The three together reached about 8000 RCTs in November 2019 (11). Despite this available data, re-analyses are still sparse. Among the 88 published outputs we identified resulting from data-sharing on these platforms, only 3 were re-analyses: “Restoring Study 329” by Le Noury et al. which contradicted the initial publication, a trial that was already known to be misreported (12), a reanalysis of the TORCH trial suggesting an overestimation of the treatment effect in the original study (13), and the reanalysis of the “SMART-AF” trial which came to similar conclusions to the original study (14).
As part of a global research program on reproducibility in therapeutic research (ReiTheR, funded by the French National Research Agency), we designed the present cross-sectional study to assess inferential reproducibility (i.e. when IPD is available, whether qualitatively similar conclusions can be drawn from a reanalysis of the original trials) for main studies assessed by the EMA.

Our hypothesis is that for most trials (> 95%) for which we obtain the data, the results observed on the primary outcome will be fully reproducible. However, although we planned one year for data collection, we are aware that after this time some data will still not be available and thus not be re-analyzable. Nevertheless, the worst-case scenario for precision estimates is that 50% of the studies will be analyzable and reproduced.

Methods:

This is a registered report: the research protocol will be peer-reviewed before the actual research takes place. Once it is accepted, the editors undertake to publish the completed study if the protocol is validated even if there are statistically negative findings (study hypothesis not verified). This approach is expected to reduce issues such as publication bias (15). In addition, the protocol will be pre-registered on the Open Science Framework (16).

Proposed timeline

From January 2020 until January 2021 we will collect and select relevant trials and initiate all data requests. Collation and re-analysis of the data will take place from January 2020 to November 2021.

Eligibility criteria

EPARs
We will collect all EPARs on new authorised human medications, biosimilars and orphan medicines given a positive opinion by the CHMP (Committee for Medicinal Products for Human Use) after 1st January 2017 and approved by the European Commission. We will exclude EPARs concerning generics and hybrid medicine. Definitions concerning the different types of drugs can be found in additional file 1. The distinction between new biosimilars, new generics, new hybrid medicine, orphan medicines or new medicines will follow the CHMP Meeting Highlights (17).

Main Studies

Pivotal trials are referred to as “main studies” in the different EPARs. Any main study will be included, with no distinction in terms of study phase, study type, study design or intervention.

If an indication for a drug has been refused and another indication authorized, we will not consider the main study for the non-authorized indication.

Furthermore, studies with no primary outcome identified will not be included and will be listed as non-evaluable studies.

To get an idea of how many trials are eligible, numbers are provided in the pilot data below (see the “pilot data” paragraph).

Search Strategy:

Eligible main trials

Two reviewers (MS, JG) will manually extract all names of the new medicines, biosimilars and orphan medicines approved by the CHMP and enter the information on an Excel Sheet. Afterwards, a check will be performed to verify that the CHMP opinion was adopted by the European Commission (18). Next, the reviewers will identify the corresponding eligible EPARs on the EMA website (19) and will
extract all main studies reported in these EPARs. In case of disagreement, a third reviewer (CL or FN) will arbitrate.

**Sample size calculation**

A random sample of 62 of these main studies will be selected using R (rnorm function) (20). This sample size will ensure a precision of ± 12% to estimate our primary outcome (i.e. percentage of reproducible studies, see below for a definition) in the worst-case scenario for precision estimations (i.e. if the percentage is 50%).

**Main study document accessibility**

For all selected main studies, one reviewer (JG) will search for the EudraCT number and/or the Sponsor Protocol Number, and/or any other identification information in each EPAR, and will identify the official Sponsor of the study. If this information is lacking, the same reviewer will start a wildcard search using keywords (disease, drug) from the study in the European Union (EU) Clinical Trial Register (21). If this is not successful, the reviewer will go on the websites ClinicalTrials.gov (22), International Clinical Trials Registry Portal (ICTRP), World Health Organization (23) and the International Standard Randomised Controlled Trial Number (ISRCTN) allocated by BioMedCentral (24). If information on sponsor and study number is still lacking the reviewer will contact the EMA.

Once the sponsor and the study number are identified, the reviewer will contact the sponsor to collect all of the following main study documents: 1) IPD, 2) data analysis plan, 3) unpublished and/or published study protocols with any date-stamped amendments 4) all the following dates: date of the last visit of the last patient, date of database lock (if available) and date of study unblinding, 5) unpublished and/or published (scientific article) study reports.

To this end the reviewer will send a standardised email (additional file 2), presenting the research project with a link to the pre-registered protocol on the Open Science Framework. In order to improve the return rate, up to 4 emails will be sent, the original and 3 reminder emails (with a two-week interval between e-mails).
If we are asked for this information, we will indicate that the Data-Sharing of raw data is welcome in form of Study Data Tabulation Model (SDTM) which was created by the Clinical Data International Standard Consortium (CDISC) (25).

In some cases, it will be sufficient to contact the sponsor by e-mail, in other cases the sponsor will ask us to retrieve the data on a web portal. In this case we will have to use the platform.

In parallel the same reviewer will search for these documents on the EMA portal (26) and by inspecting the published reports (if available) identified using open trial (27, 28). This process is summarised in Figure 1.

Data Extraction:

The identification of main studies and the following trial characteristics will be extracted on an Excel spreadsheet by two independent researchers (JG and FN or CL).

These characteristics include patient characteristics (e.g. percentage of women, mean age of participants, pediatric indication), study design (e.g. endpoint type, description for each primary endpoint) and intervention characteristics (e.g. drug). All are described in the supplementary material.

In case of disagreement, a third independent reviewer (FN or CL) will arbitrate. An exhaustive list of the trial characteristics extracted can be found in the additional file 3. The data extraction sheet will be pilot-tested on 10 studies before being validated.

Concerning the re-analysis, a first reviewer (JG, PhD Student) will collect the information and collate data for the reanalysis. More specifically, the reviewer will prepare a dossier with the following information for each study: 1/ the protocol, 2/ all amendments to the protocol (with their dates), 3/ all the following dates: date of the last visit of the last patient, date of database lock (if available) and date of study unblinding, and 4/ the IPD. In case of information still lacking, the study authors will be contacted.
Strategy for re-analyses:

Should the IPD not be available one year after our initial request, the study will be considered as non-reproducible (primary outcome of our study).

On the basis of the dossier prepared by the first researcher, re-analyses of the primary outcome(s) of each study will be performed by a second researcher (MS, PhD student) who will have no access to study reports, journal publications, statistical analysis plan, or analytical code, in order to ensure that the analysis is as blind as possible to the primary analysis. In addition, this reviewer will be instructed not to try to find these documents or the published report.

For single-blind studies or open-label studies, analyses will be performed according to the first version of the protocol, because outcome switching is possible in these studies. For double-blind studies, all re-analyses will be based on the latest version of the protocol issued before database lock and unblinding. If this information is not available, the date of the last visit of the last patient will be used as a proxy. In case of missing information for these dates, the study authors will be contacted.

Although in therapeutic research statistical analysis is fairly simple, in some cases the re-analyses can involve difficult methodological choices. An independent senior statistician (AR) will be available to discuss any difficult aspect or choice in the analysis plan before the re-analysis, so as to choose the most consensual analyses (e.g. Intention to Treat population for a superiority trial).

Should insufficient information concerning the main analysis be provided in the protocol, the best practices for clinical research will be used, following the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH Guidelines) (29).

An analysis plan will be developed for each study included and will be recorded on the Open Science Framework.

In the supplementary material a table is provided with details of what will be taken from the ICH guidelines in case of missing information (additional file 4). Re-analyses will entail the following
different steps: 1/ identification of the primary outcome (and detection of outcome switching), 2/
definition of the study population, 3/ re-analysis of the primary outcome. Any change identified
between the first version of the protocol and the version used for the re-analysis of the primary
outcome will be tracked and described.

Procedure to assess reproducibility:

All results of these analyses will be reported in terms of each study’s 1/ conclusion (positive or
negative), 2/ p-value, 3/ effect size (and details about the outcome), 4/ changes from the initial
protocol.

These results will first be compared with the results of the analyses reported in the EPARs and, if
these are not available, with the study reports, and again if not available, with the publications. All
results from all available documents will be gathered (EPARs, study reports & publications) and will
be presented in the results section.

Because interpreting an RCT involves clinical expertise, and cannot be reduced to solely quantitative
factors, an in-depth discussion between two researchers not involved in the re-analysis (JG and FN),
based on both quantitative and qualitative (clinical judgment) factors will enable a decision on
whether the changes in results described quantitatively could materialize into a change in
conclusions. If these two reviewers judge that the conclusions are the same, the study will be
considered as reproduced. If these two researchers judge that the conclusions are not the same,
then the researcher in charge of the analysis (MS) will be given the statistical analysis plan of the
study and will be asked to list the differences in terms of analysis. If he finds a discrepancy between
the study data analysis plan and his own analysis plan, then he will correct this discrepancy in his
analysis (e.g. analysis population, use of covariates). Again, an in-depth discussion between two
researchers not involved in the re-analysis (JG and FN) will enable a decision on whether the changes in results described quantitatively could materialize into a change in conclusions, and whether the differences in terms of analytical plan are understandable and acceptable. If these two researchers judge that the conclusions are the same, the study will be considered as reproduced with verification.

If these two researchers judge that the conclusions are not the same or that the change in the analytical plan is neither justified nor desirable, then a senior statistician will perform his own re-analysis. Then a last in-depth discussion between two researchers not involved in the re-analysis (JG and FN) based on the senior statistician’s re-analysis will enable a decision on whether the changes in results described quantitatively could materialize into a change in conclusions. If these two researchers judge that the conclusions are the same, the study will be considered as reproduced with verification, otherwise the results will be considered as not reproduced. Should the study be considered as not reproduced, the authors/ sponsors of the study will be contacted to discuss the discrepancy.

This process is described in Figure 2.

Outcomes

The Primary outcome is the proportion of studies where the conclusions were reproduced (yes/no; reproduced and reproduced with verification, as defined above). In case of a divergence for two or more co-primary outcomes in the same study (i.e. one analysis is reproduced and not the other(s)) the different co-primary outcomes will be described independently but the whole study will be considered as not reproduced. All reasons for classifying studies as non-reproducible or not reproduced will be described qualitatively using the taxonomy we are developing.

In addition, we will describe in what way the data-sharing required clarifications for which additional queries had to be presented to the authors to obtain the relevant information, to clarify labels or use, or both, and to reproduce the original analysis of the primary outcomes. A catalogue of these
queries will be created and we will group similar clarifications for descriptive purposes to generate a
list of some common challenges, and to help tackle these challenges pre-emptively in future
published trials.

Concerning secondary outcomes, we will describe and compare the main outcomes, p-values and
effect sizes in the re-analyses, and the analyses reported in the EPARs, the study reports and the
publications, and we will describe discrepancies. In addition, for each paper we will assess the
presence of the following key reporting biases: selective reporting of the primary outcome and "spin"
(30).

In case of outcome switching, meaning that a secondary outcome was considered as a primary
outcome in the final analysis, both endpoints will be re-analyzed.

To analyze "spin" in the results observed for the primary outcome, we will take the definition
provided by Yavchitz et al. who described it as being “a specific way of reporting, intentional or not,
to highlight that the beneficial effect of the experimental treatment in terms of efficacy or safety is
greater than that shown by the results” (31). We will follow the approach proposed by Cosgrove and
colleagues, evaluating 3 categories: misleading reporting; misleading interpretation;
overgeneralization / inappropriate extrapolation (32).

The Modalities of data sharing are described in table 1, using the following categories: the type of
data-sharing, the time lapse for collecting the data, the reason for non-availability of data, the
deidentification of data (i.e. 18 identifiers, as required by the Health Insurance Portability and
Accountability Act) (33) and the type of the shared data (here we distinguish “computerized data”
which is not formal or ordered, “cleaned data, categorized and ordered” and “analyzable data”
meaning ready for analysis) (34).

<table>
<thead>
<tr>
<th>Type of data sharing</th>
<th>Accessibility of data:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>De-identification of data</td>
<td>- Name (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Address (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Birthdate (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Telephone Number (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Fax numbers (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Email addresses (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Social security numbers (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Medical record numbers (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Health plan beneficiary numbers (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Account numbers (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Certificate/license number (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Vehicle identifiers and serial numbers (YES/ NO/ NA)</td>
</tr>
<tr>
<td></td>
<td>- Device identifiers and serial numbers (YES/ NO/ NA)</td>
</tr>
<tr>
<td>Time lapse for collecting the data</td>
<td>- Time lapse in days</td>
</tr>
<tr>
<td>Reason for non-availability if data was not shared</td>
<td>- Privacy concerns</td>
</tr>
<tr>
<td></td>
<td>- Technical issues</td>
</tr>
<tr>
<td></td>
<td>- Non-willingness to engage in sharing data</td>
</tr>
<tr>
<td></td>
<td>- No data-sharing within the time allowed (1 year)</td>
</tr>
<tr>
<td></td>
<td>- Other (specify)</td>
</tr>
<tr>
<td>Upon request by e-mail (see additional files)</td>
<td>- Upon request on a specific website / register</td>
</tr>
<tr>
<td></td>
<td>- Available on a public register</td>
</tr>
<tr>
<td></td>
<td>- Other (specify)</td>
</tr>
</tbody>
</table>
| Type of data shared | - Analysable (e.g. AdAM)  
- Edited/cleaned (e.g. SDTM)  
- Computerised (e.g. CDASH) |
|---------------------|-------------------------------------------------------------------|
| - Web Universal Resource Locators (URL) (YES/ NO/ NA)  
- Internet Protocol (IP) addresses (YES/ NO/ NA)  
- Biometric identifiers, including finger and voice prints (YES/ NO/ NA)  
- Full-face photographs and any comparable images (YES/ NO/ NA)  
- Any other unique identifying number, characteristic, or code (YES/ NO/ NA) |

### Table 1: Modalities collected

**Data Analysis**

We will perform a descriptive analysis of the characteristics of the extracted main studies included in the EPARs selected. This will include counts, percentages and their associated 95% confidence intervals (CIs).

Effect estimates in the different studies will be expressed as standardised mean differences (SMDs) and their associated 95% confidence intervals. For binary outcomes, odds ratios and their 95% CIs will be calculated and converted into the standardised mean difference (35).

In order to compare the results of our re-analyses with the original results, the following steps will be implemented: 1/ We will compare the statistical significance in the form of the p-value. If different, the result will be considered as not reproducible. If not different, 2/ we will qualitatively compare effect sizes and their respective 95% CIs. In case of +/- 0.10 points difference in point estimates...
expressed as standardised mean differences), the difference will be discussed with a clinician in order to assess its clinical significance.

All analyses will be performed using the open source statistical software R (R Development Core Team) (20). The code will be made public on the Open Science Framework, as well as a file summarizing the process to retrieve all data-sets (16).

**Limitations**

A limitation of our study is that it is restricted to primary endpoints. While primary endpoints are paramount in RCTs, other endpoints (secondary endpoints and safety endpoints) could also be of interest to regulators.

**Risks for study completion and factors mitigating that risk**

The blinding of the reviewer may not be ensured all the time, as results of clinical trials can be found incidentally. To minimise this risk, the analyser (MS) will register his statistical analysis plan for each study on the Open Science Framework, before conducting the analysis. However, we are aware that unblinding can still occur and affect the setting up of the analysis plan, as complete blinding is impossible. If this occurs, the reviewer will document what information he has and why he has some information about the study in his statistical analysis plan.

Another risk will be the availability of data in the relevant timespan for this project. For this reason, we fixed a one-year time frame in which study sponsors are requested to send us the IPD. One year seems a reasonable time frame to balance any administrative issue with the need for fast and effective data-sharing (for instance for IPD meta-analyses).

**Pilot data**
From January 1st 2017 to November 21st 2019, 236 new human medicinal products (biosimilars, generics, etc.) received a Marketing Authorization and their respective positive European Product Assessment Reports were published.

A closer look at the last 5 approved medicines including Biosimilars by the European Commission showed that 5 EPARs contained 10 main studies of which all were eligible.

<table>
<thead>
<tr>
<th>EPAR</th>
<th>Vitrakvi</th>
<th>Xospata</th>
<th>Trogarzo</th>
<th>Inbria</th>
<th>Epidyolex</th>
</tr>
</thead>
<tbody>
<tr>
<td>International name or common name</td>
<td>Larotrectinib</td>
<td>Gilteritinib</td>
<td>Ibaliuzumab</td>
<td>Levodopa</td>
<td>Cannabidiol</td>
</tr>
<tr>
<td>Type of medicine</td>
<td>New medicine</td>
<td>Orphan medicine</td>
<td>New medicine</td>
<td>New medicine</td>
<td>New medicine</td>
</tr>
<tr>
<td>Month and Year of CHMP opinion decision</td>
<td>July 2019</td>
<td>September 2019</td>
<td>July 2019</td>
<td>July 2019</td>
<td>July 2019</td>
</tr>
<tr>
<td>Number of main studies</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Number of eligible studies</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>If one or more not eligible, why?</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 2: Pilot Data
Concerning re-analyses, our team has already explored the effectiveness of data-sharing by RCTs in studies published in BMJ and PLOS Medicine, two journals with full data-sharing policies (36). Of 37 RCTs, 17 (46%, 95% CI 30% - 62%) met the data availability requirements. Furthermore, 14 of the 17 RCTs (82%, 95% CI 59% - 94%) were fully reproduced on all their primary outcomes. In the remaining RCTs there were two where errors were found, but similar conclusions were described. For one paper it was impossible to reproduce the analyses because insufficient information was given in the Method section.

**Expected impact:**

Up to now, there has been no clear picture of the inferential reproducibility of the main studies in European Assessment Reports. These main studies are used to assess the efficacy of new medical products which will influence the lives of millions of people in the European Union. If data availability were to prove low it should urge the EMA to implement an even stronger data-sharing policy. If re-analyses of available data show low reproducibility, it would argue for independent re-analyses at the time of the approval. On the other hand, if there were to be no issues in terms of reproducibility, it would reinforce the confidence one can have in EMA’s transparency concerning processes and decisions.

In 2019, the EU Clinical Trial Regulation 536/2014 will come into force and will further extend the boundaries of data-sharing and transparency in the EMA as well as within the European Union. The EU Portal and Data Base is being developed, creating a single-entry point for submitting clinical trials in the EU. A further advantage is that not only information about clinical trials included in Marketing Authorization Applications (MAA) can be found on the portal, but the aim is to have data about every single trial conducted in the European Union, whether or not it is part of an MAA (37). Our research could help to highlight the interest of the future regulation.
To further enhance the impact of this project, the reproducibility of our results will be checked by comparison with another reproducibility project in the ReiTheR project, also submitted as a registered report.

Declarations:

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Competing interests: The authors declare that they have no competing interests

Funding: The project is funded by the Agence Nationale de la Recherche (reference number ANR-17-CE-36-0010-01). The sponsor had no role concerning preparation, review, or approval of the manuscript.

Author’s contribution: MS, JG and FN designed the work. AR, BL, CL and DM drafted the work and substantively revised it. All authors read and approved the final manuscript.

Acknowledgments: Not applicable

Availability of data and materials:

Data that supports the findings of this study will be available on our link to the Open Science Framework (https://osf.io/mcw3t/) once the article is published. Furthermore, interested researchers can contact the corresponding author via mail.
References:


41. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in The BMJ and PLOS Medicine. BMJ. 2018;360:k400.

FIGURES

Figure 1: Process of accessing main study documents

Figure 2: Procedure of assessing reproducibility

SUPPLEMENTARY MATERIAL

Additional file 1: Table with definitions of different types of medicines on the EMA website (Word)

Additional file 2: Letter to the Sponsor (Word)

Additional file 3: Table of Study Characteristics that will be extracted (Word)

Additional file 4: Table with details of what will be taken from the ICH guidelines in case of missing information (Word)