BIG THUNDER, LITTLE RAIN?

An academic viewpoint

Franz König, Michael Wolzt and Martin Posch

Medical University of Vienna
Center for Medical Statistics, Informatics, and Intelligent Systems
Section for Medical Statistics
Franz.Koenig@meduniwien.ac.at
www.meduniwien.ac.at/user/franz.koenig

FP7 HEALTH 2013 - 602552
„Hiding“ of scientific data

- unethical
- non-scientific
- uneconomical

Whether it is intended or not should not matter at all!
Should we not have access to any data due to freedom of information acts anyway?

- You can request any document from any EU institution, eg from EMA
- E.g., 2010 EMA access-to-documents policy
- Since November 2010, the EMA has released more than 1.9 million pages in response to such requests.
- **Was put on hold!** Preliminary order by the General EU Court due to two on-going legal actions of the pharma companies AbbVie and InterMune.
Two years ago (22/11/2012) at the EMA Workshop on clinical-trial data and transparency an avalanche was set off ...

Guido Rasi, Executive Director of European Medicines Agency (EMA):

“...we are not here to decide if we publish clinical-trial data, but how!”

Open access to Clinical Study Report (CSR): designates the entirety of elements submitted as study reports in CTD Module 5, following the format of the ICH E3 document.

Controlled access to Raw CT data (meaning individual patient data sets, individual patient line-listings, individual Case Report Forms (CRFs), and documentation explaining the structure and content of data sets.)
Further Clinical Trial Data Transparency Initiatives

• **BMJ Open Data Campaign**
  “As of January 2013, the BMJ will no longer publish any trial of drugs or devices where the authors do not commit to making the relevant anonymised patient level data available, upon reasonable request.”

• **FDA Transparency Initiative**
  Availability of Masked and De-identified Non-Summary Safety and Efficacy Data

• **All Trials Initiative**
  “All Trials Registered, All Results Reported”

• **Individual Pharmaceutical Industry Initiatives**
  GSK Data transparency initiative, Roche Global Policy on Sharing of clinical Trial Data, ...
  Researchers may receive access to raw data after requests have been reviewed by an independent panel of experts

• **Yale University Open Data Access (YODA) Project**
  … a model to facilitate access to patient-level clinical research data to promote wider availability of clinical trial data and independent analysis by external investigators

• **Cochrane Collaboration statement on access to clinical trial data**
  “All data from all randomised clinical trials, including raw anonymised individual participant data that do not allow identification of individual participants, and the corresponding trial protocols, to become publicly available free of charge and in easily accessible electronic formats”

• **Joint Statement of EFPIA and PHRMA**
  Principles for Responsible Clinical Trial Data Sharing

• **New EU regulation on clinical trials on medicinal products for human use**
OPEN ACCESS TO DATA
What are the opportunities, challenges and risks of sharing clinical trial data?
Life as academic researcher in medical research ...

- Enhance knowledge in medicine (patients should receive better treatments)
- Career path at universities
- Scientific metrics
  # publications (as first/last author), IF, H-factor, grants, ...
- Collect data related to interesting research questions, publish as many paper as possible (but not all type of papers/journal will count)
- Who owns the data? Do you want someone else to publish „your“ data?
- How successful have we been so far in granting access to important information?
Life as (academic) researcher ... 
... Publish or Perish

Comic from http://science2enlighten.blogspot.co.at/2012_07_01_archive.html
Medical University of Vienna
Do medical researchers already share data?

- In some areas sharing of (raw) data is common,
  - E.g., genomics
- Some journals like BMJ require already commitment to give access to raw data (for some studies ...)
- Also sharing of other documents becomes more common
  - Publication of study protocols
- We biostatisticians get also used to share our data
  - E.g., some journals require open access to software code used for analysis, simulation,

... to enhance reproducible research
Thus, is there room for improvement?

Presently, only for a fraction of clinical trials, results are published.


Clinical trial data are underutilized!


Medical University of Vienna
Transparency aspects in the proposal for new EU regulation on clinical trials on medicinal products for human use

• „For the purposes of this Regulation, in general the data included in clinical study reports should not be considered commercially confidential once a marketing authorisation has been granted or the decision-making process on an application for marketing.”

• Registration before the initiation of a trial
• Publication of summary results in a publicly and easily accessible database
• Access to clinical trial data
• Does not distinguish between academic or industry trials
Do we know which trials are currently conducted?

- Medical studies require approval by an Ethics committee before start

- Is this information publically accessible?

- Some trials are registered at public registries (WHO, ClinicalTrial.Gov, EudraCTm, ...)

- Depending on the registry more or less information on a trial is available
What happens to medical studies after Ethical Approval?

VIENNA: Publication of scientific research
[Diploma thesis J. Neugebauer, 2010; supervisor M. Wolzt]

- Clinical studies Feb 1998 – Jan 99 approved by EC of the Medical University of Vienna
- 447 studies approved
- Publication rate 35% (158/447 studies)
  - Industry sponsor 36/177 (20%)
  - Non-commercial sponsor 122/270 (45%)

FREIBURG: Fate of Clinical Research Studies after Ethical Approval
[Blümle et al., PLOS One 2014]

- Clinical studies 2000-02 approved by EC of the University of Freiburg
- Publication rate: 419 / 807 (52%) with at least 1 publication (data of ~120000 study participants hidden)
## Fate of Clinical Research Studies after Ethical Approval

[Blümle et al., PLOS One 2014)

<table>
<thead>
<tr>
<th>Study characteristics</th>
<th>Approved (column %)</th>
<th>Started at local study site</th>
<th>Of those started:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Published (row %)</td>
</tr>
<tr>
<td>Total</td>
<td>917 (100)</td>
<td>807</td>
<td>419 (52)</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomised controlled trial</td>
<td>408 (45)</td>
<td>355</td>
<td>201 (57)</td>
</tr>
<tr>
<td>Non-randomised intervention</td>
<td>72 (8)</td>
<td>65</td>
<td>33 (51)</td>
</tr>
<tr>
<td>study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic study</td>
<td>41 (4)</td>
<td>36</td>
<td>21 (58)</td>
</tr>
<tr>
<td>Cohort study</td>
<td>23 (2)</td>
<td>19</td>
<td>8 (42)</td>
</tr>
<tr>
<td>Case-control study</td>
<td>6 (1)</td>
<td>6</td>
<td>3 (50)</td>
</tr>
<tr>
<td>Cross-sectional study</td>
<td>42 (5)</td>
<td>40</td>
<td>16 (40)</td>
</tr>
<tr>
<td>Uncontrolled study</td>
<td>186 (20)</td>
<td>163</td>
<td>75 (46)</td>
</tr>
<tr>
<td>Laboratory study</td>
<td>138 (15)</td>
<td>122</td>
<td>61 (50)</td>
</tr>
<tr>
<td>Health services research</td>
<td>1 (&lt;1)</td>
<td>1</td>
<td>1 (100)</td>
</tr>
</tbody>
</table>

*Pearson $\chi^2$ (df 8) = 10.173, $p = 0.253$*

| Study size                   |                     |                            |                   |                       |
|------------------------------|---------------------|----------------------------|-------------------|
| ≥median of 120               | 449 (49)            | 391                        | 224 (57)          | 167 (43)               |
| <median of 120               | 429 (47)            | 379                        | 177 (47)          | 202 (53)               |
| Unclear                      | 39 (4)              | 37                         | 18 (49)           | 19 (51)                |

*Pearson $\chi^2$ (df 2) = 0.009, $p = 0.923$*
# Fate of Clinical Research Studies after Ethical Approval

[Blümle et al., PLOS One 2014]

<table>
<thead>
<tr>
<th>Study characteristics</th>
<th>Approved (column %)</th>
<th>Started at local study site</th>
<th>Of those started:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Published (row %)</td>
</tr>
<tr>
<td>Collaboration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-centre study</td>
<td>383 (42)</td>
<td>340</td>
<td>159 (47)</td>
</tr>
<tr>
<td>Multi-centre study</td>
<td>534 (58)</td>
<td>467</td>
<td>260 (56)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Pearson ( \chi^2 ) (df 1) = 6.257, ( p = 0.012 )</strong></td>
</tr>
<tr>
<td>Only multi-centre studies:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International</td>
<td>310 (58)</td>
<td>276</td>
<td>173 (63)</td>
</tr>
<tr>
<td>Domestic</td>
<td>221 (41)</td>
<td>189</td>
<td>87 (46)</td>
</tr>
<tr>
<td>Unclear</td>
<td>3 (&lt;1)</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Pearson ( \chi^2 ) (df 2) = 15.124, ( p = 0.00052 )</strong></td>
</tr>
<tr>
<td>Funding (as stated in protocol)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial</td>
<td>422 (46)</td>
<td>368</td>
<td>203 (55)</td>
</tr>
<tr>
<td>Non-commercial</td>
<td>140 (15)</td>
<td>131</td>
<td>75 (57)</td>
</tr>
<tr>
<td>No funding stated</td>
<td>355 (39)</td>
<td>308</td>
<td>141 (46)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Pearson ( \chi^2 ) (df 2) = 7.695, ( p = 0.021 )</strong></td>
</tr>
<tr>
<td>Only commercially funded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sponsor involved</td>
<td>362 (86)</td>
<td>318</td>
<td>182 (57)</td>
</tr>
<tr>
<td>Sponsor not involved</td>
<td>60 (14)</td>
<td>50</td>
<td>21 (42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Pearson ( \chi^2 ) (df 1) = 4.053, ( p = 0.044 )</strong></td>
</tr>
</tbody>
</table>
If they are published …

..., essential information is often missing


Potential consequences:

- a distorted information base on the risks and benefits of therapies
- impaired meta-analyses
- clinical trials that are unnecessarily repeated
Potential reasons for

Non-publication
- Competing interests (e.g. financial CoI)
- Poor project management
- Lack of time
- Low priority
- Disagreement
- Losing interest
- Moving to another institution
- Results not deemed important enough
- Journal rejection (publication bias)

Publication
- Funding (commercial or non-commercial)
- Study design (Multi-centre)
- Study size (large)
- Collaboration (international)
Thus more data need to be shared
Two types of secondary research in relation to open access to clinical trial data

- **Reproducible Research**
  - Confirm sponsor’s analysis
  - Validating the original study results and investigating their robustness
  - Transparency of regulatory decision making
  - No prospective "validation protocol" necessary

- **Investigation of additional research questions**
  - Reliable synthesis of study data (Meta-analyses)
  - Exploratory research
  - Different levels of evidence: from "quasi prospective research" (with SAP written without any knowledge on results of the trial) to full data mining
  - To interpret such results - knowledge of time lines important (data access, background knowledge when formulating research questions, ...)

How to assess the risk of "false positives" of multiple retrospective analysis of clinical trial data?

Koenig et al. Biometrical Journal 2014
Raw Data Sharing – Why?

- Reproducible research
- Patient level meta-analyses
- Planning of new studies
- Enables development of tailored study designs and statistical methodology
- Avoiding the repetition of studies
- New discoveries through exploratory research
- Provide incentive to ensure accuracy of dataset

Which data needs to be shared …

- **Aggregated Clinical Trial Results**
  - Key outcomes in clinical trial registers
  - Research Articles in Scientific Journals (ideally open access)
  - Summary reports for patients (in trial, future, ...)
  - Detailed clinical study reports (regulatory agencies, EC, ...)

- **Raw (Patient Level) Data**
  - Held by individual sponsors
  - Data Repositories
  - Regulatory Authorities
Patient level data are of particular value …

- in small populations to enhance research for orphan drugs, personalized medicines, drug development for children, ...
- Identification of patient subgroups
- Raw data of past studies may serve as historical controls
- Help to formulate prior for Bayesian analyses
- More tailored statistical models (selection of covariates, time points, …)

- However, even though small populations research may benefit most, it also poses the highest risk with regards to patient privacy.
Challenges

• **Patient Privacy**
  - „Proportionate“ De-identification of data
  - Legal obligations of data requester

• **Ensuring the Quality of Re-Analysis**
  - A pre-specified analysis plan increases the credibility (as for all clinical studies).
  - Interpretation as retrospective analysis

• **Protecting Researcher/Sponsor’s Interests**
  - Suitable timing of data release
  - Give enough credits to data-generator (e.g., co-authorship in publication?)
How to make patient level data sharing happen?

- Open access to protocols and meta-data (Data-Dictionaries, CRFs) to plan secondary analysis
- Accessible data formats (standardization preferred)
- Learn from successful examples (e.g., NIH)
"It is ironic that the organizations that most resist wider access to data are the ones that stand to benefit so much from greater transparency."

Eichler et al. NEJM, 2013.
Big thunder, little rain?

• The cat is let out of the bag!
• There is a public interest and discussion even outside the research community … (e.g., see SPIEGEL Online this week)
Big thunder, little rain?

• This debate has already resulted in a huge paradigm-shift in medical research!
• Some time ago no public information was available on which studies were actually conducted
  – … clinicaltrials.gov, EudraCT, register of ethics committee,…
• Some years ago regulatory agencies were sued for publishing summary report
  – Publication of EPARS, …
• Some years ago researchers complained that there was no way to publish results of negative trials
  – again registries, open journals
• Some years ago researchers very rarely shared raw data
  – Already after publication of EMA draft policy:
    Self commitment of industry sponsors to share raw data (past & future) in a controlled environment
Big thunder, little rain?

- This discussion has revealed that the question „who owns the data in the first place?“ might have been wrongly tackled for many years!
- Helped to re-focus on what patients expect when agreeing to participate in trials and share their data!
- In the beginning there might be some problems to judge the evidence of secondary research. So what.
- Still a long way to go (100% publication rate, open access without restriction, …)
- Overall this journey should lead to an increase of quality

There is no way back to conduct research behind closed doors!
Data is like children...

You like your own best, and do not like strangers to play with them
Selected References


Acknowledgement

Integrated DEsign and AnaLysis of small population group trials

http://www.ideal.rwth-aachen.de/

This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602552.