Data Management Practices: Data Ownership, Data Collection

Prof. Khin Mar Myint
Pro rector (Academic)
University of Medicine 1, Yangon
July 2020
Case: Falsified data

• Case scenario: falsified data.docx
Outline of the presentations

- Data in research
- Data Ownership issues
- Data collection: how researchers should collect, store, protect the research data to maintain its integrity, validity and accuracy.
Data ???
Data in Research

data when referring to research projects means **the all** of the information **collected and generated** in the course of a research project.
What is research data?
Data that has formed the basis of your research output.

Research data can be observations, survey answers, scripts, photographs, notebooks,
Data: addition to research findings

- **Sponsored projects** –

  Need to consider financial data and administrative data (subject to retention requirements)

  - **Financial data** – budget information, record of expenditures
  - **Administrative data** – project proposals, required approvals
Why Data Management Practices important?

• In order to conduct research responsibly

• Researchers’ obligations:
  • *Obligations toward other researchers*
  • *Obligations toward oneself*
  • *Obligations toward the public*

---

Ref: *On Being a Scientist A Guide to Responsible Conduct in Research: Third Edition*
Obligations toward other researchers

- new research builds on previous results
- Irresponsible actions can impede an entire field of research or send it in a wrong direction
- Imbedded in this trust is a responsibility of researchers to mentor the next generation who will build their work on the current research discoveries.
Obligations toward oneself

- maintaining a reputation as a productive and honest researcher, adhering to professional standards builds personal integrity in a research career.

“If ethics are poor at the top, that behavior is copied down through the organization”

--- Robert Noyce
Obligations toward Public

• Some scientific results directly affect the health and well-being of individuals, thereby society
• Science also is used by policy makers to make informed decisions on such pressing issues
• new knowledge speaks to our sense of wonder and paves the way for future advances.
Can you rely on the drugs that your doctor prescribes?

Silent on many questions

So, how valuable is a guideline for the management of hepatitis C, in which eight authors declared FCOI with recommended drug producers: How reliable are the recommendations from a committee that did not document how evidence was evaluated or include a public or patient representative?

https://theconversation.com/can-you-rely-on-the-drugs-that-your-doctor-prescribes-98128
March 2017

Conflict of Interest in Seminal Hepatitis C Virus and Cholesterol Management Guidelines

Akilah A. Jefferson, MD, MSc¹; Steven D. Pearson, MD, MSc¹,²

› Author Affiliations  |  Article Information

Data Ownership ??
Data Ownership

Should the person who conducts the research own the product?

- imposed by
  - Funders (Government/Private)
  - research institutions
  - data sources

- No absolute rules - highly recommended that researchers determine data ownership before they embark on a study
Ownership according to Grants and Contract

- **Grants**—researchers must carry out the research as planned and submit reports, but control of the data remains with the institution that received the funds.

- **Contracts**—researcher to deliver a product or service, which is then usually owned and controlled by the government.

***Claims of funders can do vary considerably, researchers must be aware of their obligations to them before they begin collecting data.***
Research institutions.

- Support for research is typically awarded to research institutions, not to individual researchers.
- As the recipients of research funds, research institutions have responsibilities for budgets, regulatory compliance, contractual obligations, and data management.
- Research institutions claim ownership rights over data collected; this means that researchers cannot automatically assume that they can take their data with them if they move to another institution.
The Bayh-Dole Act 1980

• allowed universities to have control of the intellectual property, such as patents, generated from federally funded research.

• With a patent in hand, universities could exclusively license the patent to businesses, recent inventions, in the form of new drugs and computer technologies, have also helped the public.

• The law has encouraged new relationships between academic researchers and companies,
Data collection (data acquisition)
Data are used;

- to confirm or reject research hypotheses,
- to identify new areas of investigation,
- to guide the development of new investigative techniques, and more;

Science and practices today cannot exist without data.
Data collection (data acquisition)

- Different collection techniques for Different types of research, it can vary !!

- How to ensure the overall integrity (both the process & information)

1. Appropriate methods
2. Attention to detail
3. Authorized
4. Recording
Appropriate methods

• study design

• Experimental design

• Appropriate statistical tests

• Data collection techniques

➢ inappropriate methods in research compromises the integrity of research data
Data Acquisition Techniques

- experiments,
- published literature sources
- surveys (SAQ, Google form, Survey monkey)
- interviews
- Observations
- documents and records

*Data acquisition: process of converting acquired data from different tools to scientific result.*
Methods can also be compromised by bias—choosing one method or set of experimental conditions so that a particular conclusion can be drawn

*E.g; Scenarios in real practice*
Attention to detail

E.g;

Manuscript submission:
The editors requested the original data, images, figures
Authorized

- Obtaining appropriate permissions

Researchers have a responsibility to know when permission is needed to collect or use specific data

E.g:

- human (and animal) subjects in research
- hazardous materials and biological agents
- information contained in some libraries, databases, and archives
- information posted on some Websites
- published photographs and other published information
- other copyrighted/patented processes or materials
Data sources

- Source of data are seeking some control over data derived from them. Countries’ data, entities with unique databases, claimed ownership of research results based on their data.

- Make sure you can answer these questions
  - Who owns the data I am collecting?
  - What rights do I have to publish the data?
  - Does collecting these data impose any obligations on me?
Recording data

• **to document** what was actually done and the results that were achieved

• Hard-copy evidence—a bound notebook with numbered pages on which **date and time of research** can be clearly enumerated

• Kept in a way that will **enable someone else to repeat** each experiment

• **Do not change records** in a bound notebook without noting the date and reasons for the change
E.g;
Electronic evidence

• validated to assure that it was actually recorded on a particular date and not changed at some later date.

• It is easy to change dates on computers and thereby alter the date a particular file seems to have been created. If you collect your data electronically, you must be able to demonstrate that they are valid and have not been changed.
### 1. Raw Data: Paste your reader OD raw data below

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.716</td>
<td>1.757</td>
<td>1.84</td>
<td>1.8</td>
<td>0.96</td>
<td>1.013</td>
<td>0.313</td>
<td>0.346</td>
<td>0.655</td>
<td>0.667</td>
<td>1.44</td>
<td>1.409</td>
</tr>
<tr>
<td>B</td>
<td>1.438</td>
<td>1.462</td>
<td>1.79</td>
<td>1.801</td>
<td>0.984</td>
<td>0.97</td>
<td>0.334</td>
<td>0.322</td>
<td>0.214</td>
<td>0.267</td>
<td>1.255</td>
<td>1.239</td>
</tr>
<tr>
<td>C</td>
<td>0.967</td>
<td>1.014</td>
<td>1.794</td>
<td>1.746</td>
<td>0.958</td>
<td>0.948</td>
<td>0.312</td>
<td>0.312</td>
<td>0.157</td>
<td>0.146</td>
<td>0.233</td>
<td>0.255</td>
</tr>
<tr>
<td>D</td>
<td>0.586</td>
<td>0.606</td>
<td>1.791</td>
<td>1.822</td>
<td>0.956</td>
<td>0.906</td>
<td>0.326</td>
<td>0.305</td>
<td>0.091</td>
<td>0.082</td>
<td>1.791</td>
<td>1.822</td>
</tr>
<tr>
<td>E</td>
<td>0.319</td>
<td>0.347</td>
<td>1.462</td>
<td>1.493</td>
<td>0.562</td>
<td>0.59</td>
<td>0.171</td>
<td>0.173</td>
<td>0.334</td>
<td>0.322</td>
<td>0.334</td>
<td>0.322</td>
</tr>
<tr>
<td>F</td>
<td>0.167</td>
<td>0.186</td>
<td>1.484</td>
<td>1.472</td>
<td>0.589</td>
<td>0.588</td>
<td>0.175</td>
<td>0.169</td>
<td>0.958</td>
<td>0.948</td>
<td>0.086</td>
<td>0.09</td>
</tr>
<tr>
<td>G</td>
<td>0.094</td>
<td>0.103</td>
<td>1.445</td>
<td>1.424</td>
<td>0.587</td>
<td>0.573</td>
<td>0.175</td>
<td>0.168</td>
<td>0.589</td>
<td>0.588</td>
<td>0.313</td>
<td>0.342</td>
</tr>
<tr>
<td>H</td>
<td>0.086</td>
<td>0.09</td>
<td>1.531</td>
<td>1.478</td>
<td>0.584</td>
<td>0.536</td>
<td>0.091</td>
<td>0.086</td>
<td>0.175</td>
<td>0.168</td>
<td>0.062</td>
<td>1.012</td>
</tr>
</tbody>
</table>

---

**Standard Curve**

Rational Function: \( y = \frac{a+bx}{1+cx+dx^2} \)

Coefficient Data:
- \(a = -0.5122547\)
- \(b = 3.005744204\)
- \(c = 0.923589601\)
- \(d = -0.419016288\)

---

**Exponential Association**

### Graph

- **OD** vs. **Concentration**
- **Rational Function**
- **Coefficient Data**

---

**Fill Sample Data** **Clear Data**
Data Protection

- data must be protected for later use, to confirm research findings, to be reanalyzed by other researchers
- Data Confidentiality: Some data may be subject to privacy restrictions (human subjects or confidential business information), a safe place that is accessible only to authorized personnel, using random codes

*PI has the primary responsibility for its protection.*
Research Misconduct: Concerned Data Management

- **Fabrication** is making up data or results.
- **Falsification** is manipulating or changing or omitting data or results.
- **Plagiarism** is the appropriation of another person's ideas or words without giving appropriate credit.
How big is the problem?

- 163 biostatisticians surveyed.
- 51% knew of at least one fraudulent project during the last 10 years.
- 31% were involved in a fraudulent project.
- 13% directly asked to fabricate or falsify data.

“Responsible Conduct of Research (RCR) and Research Ethics”, DMR (HQ), July 2020

Ranstam et al.; Control Clin Trials. 2000

http://www.thejabberwock.org/blogl/finalist3.gif

"You are completely free to carry out whatever research you want, so long as you come to these conclusions."
## Misconduct in data management

<table>
<thead>
<tr>
<th>Research Misconduct</th>
<th>Self Report %</th>
<th>Observed %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Making up research data</td>
<td>10.4</td>
<td>34.6</td>
</tr>
<tr>
<td>Changing research data without mentioning it.</td>
<td>5.4</td>
<td>30.8</td>
</tr>
<tr>
<td>Dropping “outliers” without mentioning it</td>
<td>7.2</td>
<td>36.1</td>
</tr>
<tr>
<td>Selecting only those data that support your hypothesis</td>
<td>9.0</td>
<td>47.0</td>
</tr>
</tbody>
</table>
Different forms of fraud (e.g., fabrication and falsification of data, deceptive reporting of results, suppression of data, and deceptive design or analysis)

fraud is not a negligible phenomenon in medical research
Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis

Mandeep R Mehra, Sapan S Desai, Frank Ruschitzka, Amit N Patel

Summary

Background Hydroxychloroquine or chloroquine, often in combination with a second-generation macrolide, are widely used for treatment of COVID-19, despite no conclusive evidence of their benefit. Although generally safe when used for approved indications such as autoimmune disease or malaria, the safety and benefit of these treatment regimens are poorly evaluated in COVID-19.

Methods We did a multinational registry analysis of the use of hydroxychloroquine or chloroquine with/without a macrolide for treatment of COVID-19. The registry comprised data from 671 hospitals in 41 countries. We included patients hospitalised between Dec 20, 2019, and April 14, 2020, with a positive laboratory test for SARS-CoV-2. Patients who received one of the treatments of interest within 48 h of diagnosis were included in one of four treatment groups (chloroquine alone, chloroquine with a macrolide, hydroxychloroquine alone, or hydroxychloroquine with a macrolide), and patients who received none of these treatments formed the control group. Patients for whom one of the treatments of interest was initiated more than 48 h after diagnosis or who were on mechanical ventilation, as well as patients who received remdesivir, were excluded. The main outcome of interest was in-hospital mortality and the occurrence of de-novo ventricular arrhythmias (defined as new-onset ventricular tachycardia or ventricular fibrillation).

Findings 96032 patients (mean age 53·8 years, 46·3% women) with COVID-19 were hospitalised during the study period and met the inclusion criteria. Of these, 9201 patients were in the treatment groups (1868 received chloroquine, 3783 received chloroquine with a macrolide, 3016 received hydroxychloroquine, and 6221 received hydroxychloroquine with a macrolide) and 86831 patients in the control group. 10 698 (11·1%) patients died in hospital. After controlling for multiple confounders (age, sex, race or ethnicity, body-mass index, underlying cardiovascular disease and its risk factors, diabetes, chronic lung disease, smoking, immunosuppressed condition, and baseline disease severity), those who received a treatment had a lower mortality in the control group (9·3%). Hydroxychloroquine (18·0%; hazard ratio 1·335, 95% CI 1·229–1·457), hydroxychloroquine with a macrolide (23·8%; 1·447, 1·368–1·531), chloroquine (16·4%; 1·365, 1·318–1·531), and chloroquine with a macrolide (22·3%; 1·368, 1·273–1·469) were each independently associated with increased in-hospital mortality. Compared with the control group (0·3%), hydroxychloroquine (6·2%; 2·364–9·352, 900), hydroxychloroquine with a macrolide (0·1%; 5·106, 1·406–5·983), chloroquine (4·3%; 2·330–7·050, 3·96), and chloroquine with a macrolide (6·5%; 4·011, 3·344–4·812) were each independently associated with an increased risk of de-novo ventricular arrhythmia during hospitalisation.

Interpretation We were unable to confirm a benefit of hydroxychloroquine or chloroquine, when used alone or with a macrolide, on in-hospital outcomes for COVID-19. Each of these drug regimens was associated with decreased in-hospital mortality and decreased frequency of ventricular arrhythmias when used for treatment of COVID-19.

Funding William M. Keen Distinguished Chair in Advanced Cardiovascular Medicine at Brigham and Women’s Hospital.

Copyright © 2020 Elsevier Ltd. All rights reserved.
Retraction—Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis

Mandeep R Mehra  Frank Ruschitzka  Amit N Patel

Published: June 05, 2020  DOI: https://doi.org/10.1016/S0140-6736(20)31324-6

Our independent peer reviewers informed us that Surgisphere would not transfer the full dataset, client contracts, and the full ISO audit report to their servers for analysis as such transfer would violate client agreements and confidentiality requirements. As such, our reviewers were not able to conduct an independent and private peer review and therefore notified us of their withdrawal from the peer-review process.

We always aspire to perform our research in accordance with the highest ethical and professional guidelines. We can never forget the responsibility we have as researchers to scrupulously ensure that we rely on data sources that adhere to our high standards. Based on this development, we can no longer vouch for the veracity of the primary data sources. Due to this unfortunate development, the authors request that the paper be retracted.
A single lie destroys a whole reputation of integrity.

Baltasar Gracian

- Value and belief
- Researcher’s reputation
- Public belief/perception of science
How to Ensure Data integrity?

Research Data Management Plan (RDMP template)

The Data Management Plan can be used in a number of ways to assist with data management planning.

These include:

- To identify a series of issues and underlying questions that should be considered when planning a research project and initiate discussions within a project (team)
- For the development of a data management plan
- To raise awareness of good practice when planning for data management during the life cycle of a research project.
Research Data Management Plans (RDMP)

Helps you think constructively about your research data
- Compulsory for all research projects at the University
- Part of the HDR progress plan

Use the **RDMP checklist** if you haven’t done an RDMP before, then submit a formal RDMP via the online **RDMP tool**
Research Data management Plan
RDMP

• template
Case: Falsified data

• Case scenario: falsified data.docx
References


2. **Nicholas H. S**, ORI Introduction to the Responsible Conduct of Research... Revised Edition August 2007


Thank you