Open Science and Reproducibility 30s workshop series

Open Data

Open Science

Open Mind

UNIL | Université de Lausanne
Faculté de biologie et de médecine

CHUV

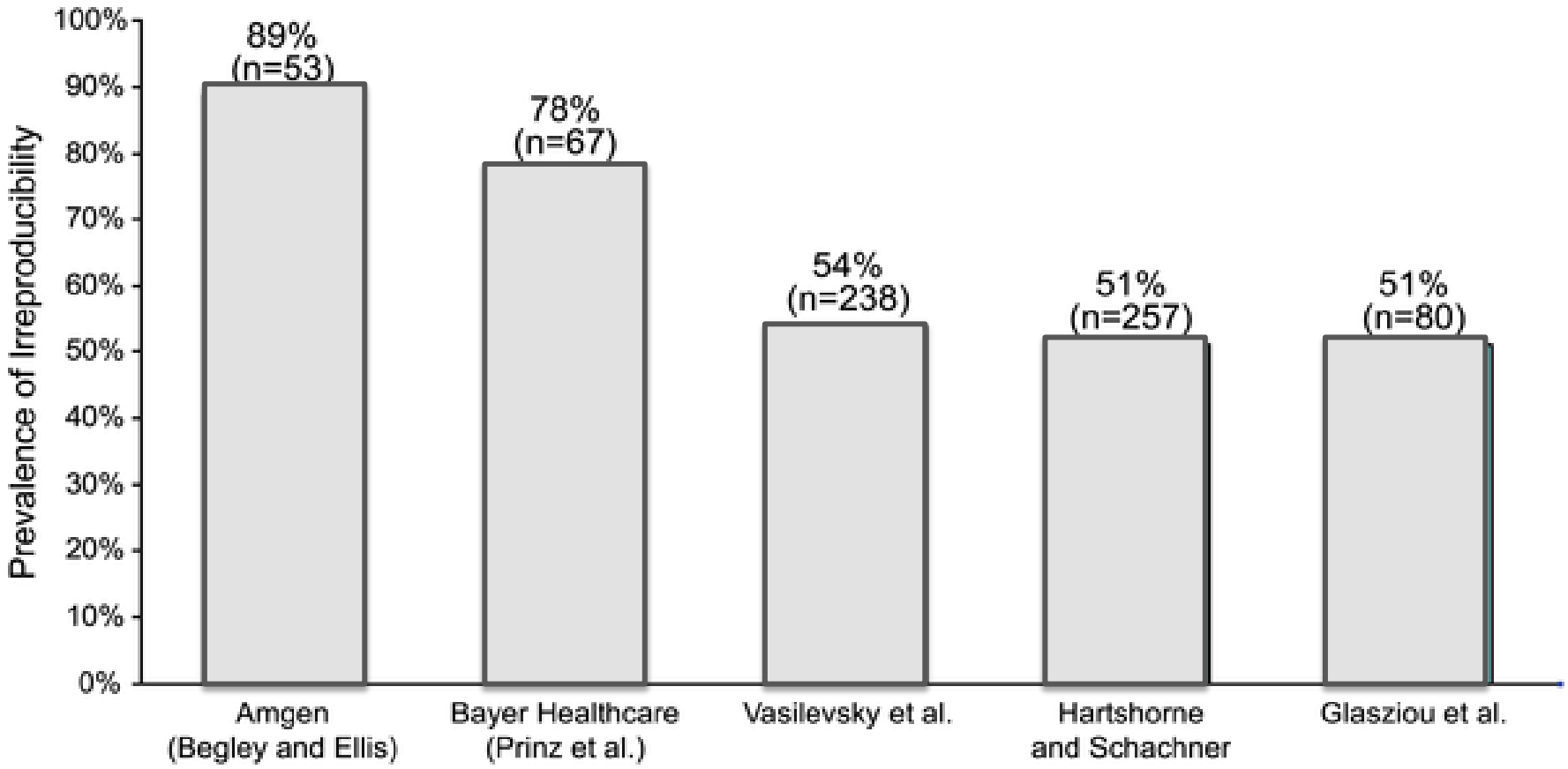
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

SIB Vital-IT High Performance Computing Center

Scientific Information School

BIOTELLIGENCES

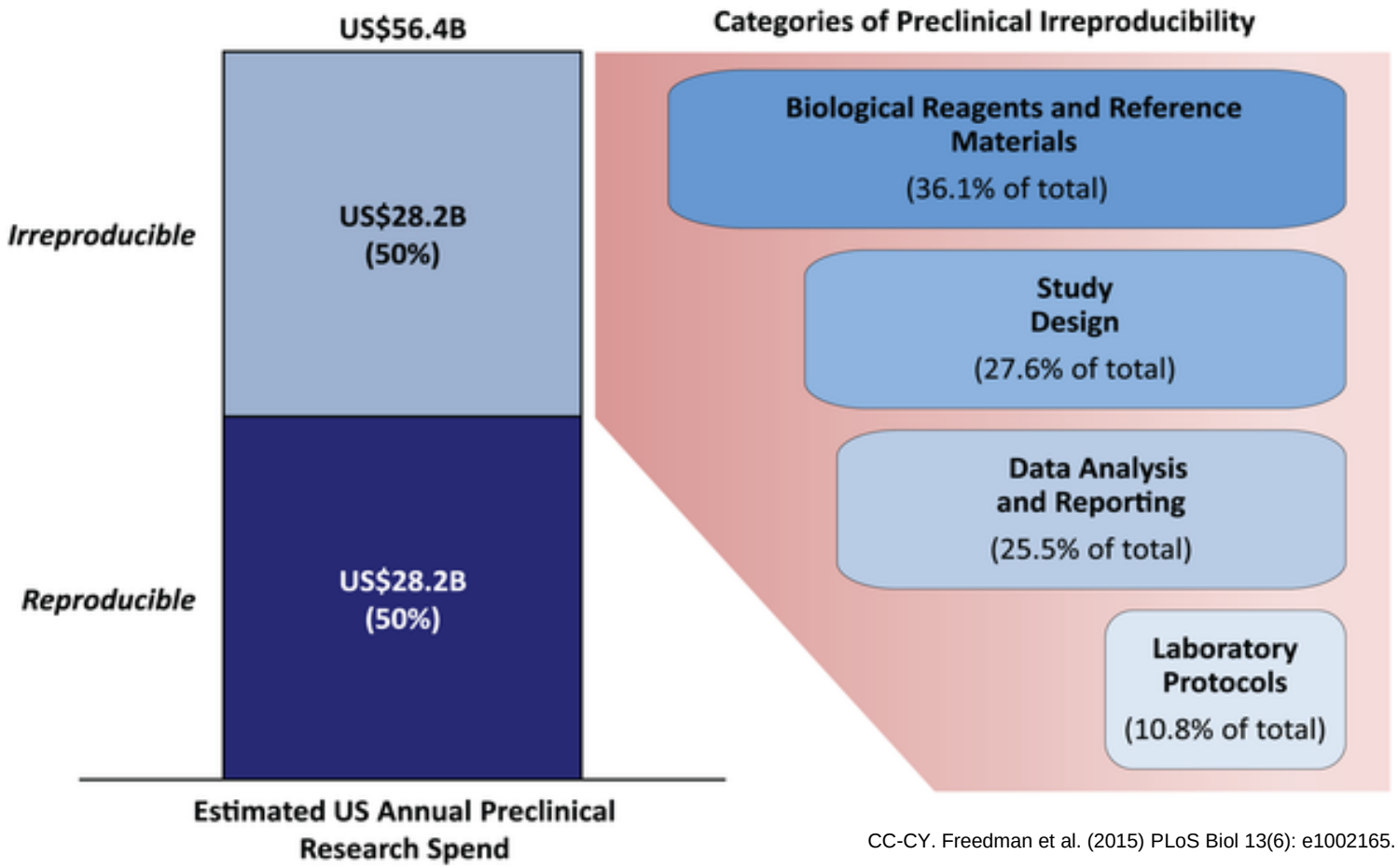
Context: Biomedical & translational research validity under controversy



 Freedman LP, Cockburn IM, Simcoe TS (2015) PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165

PLOS

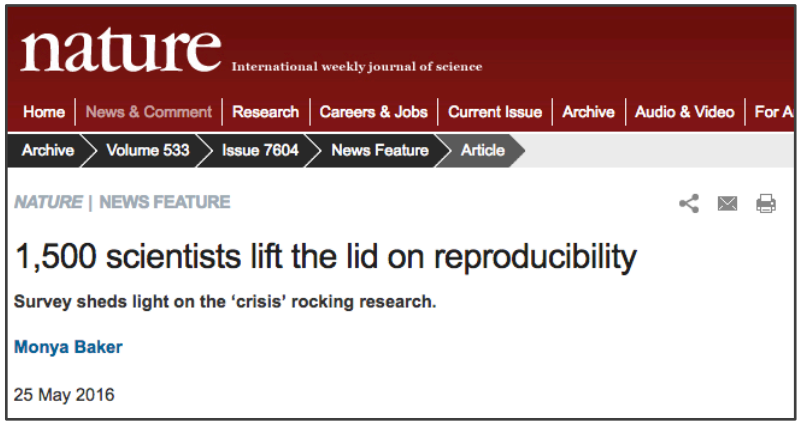
Preclinical research spend and errors that contribute to irreproducibility

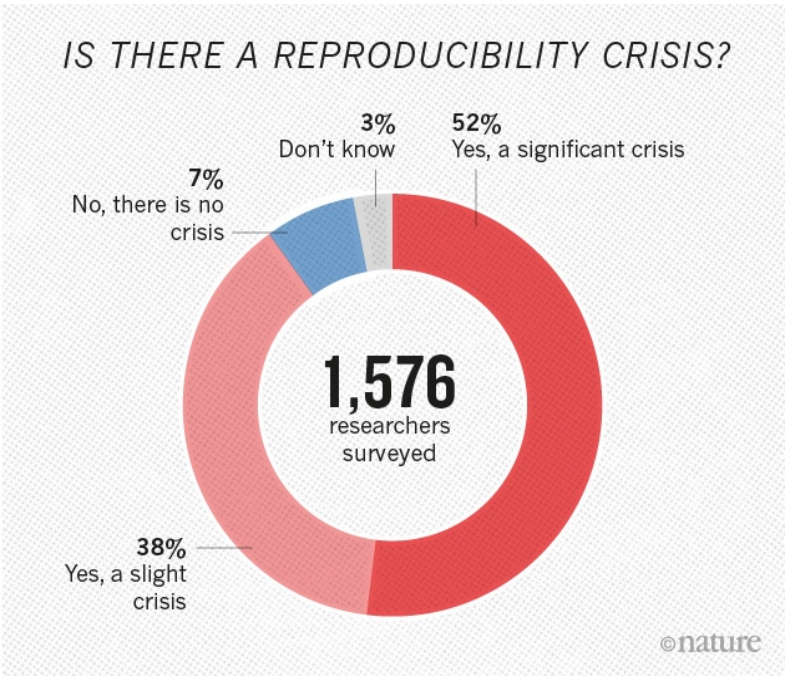


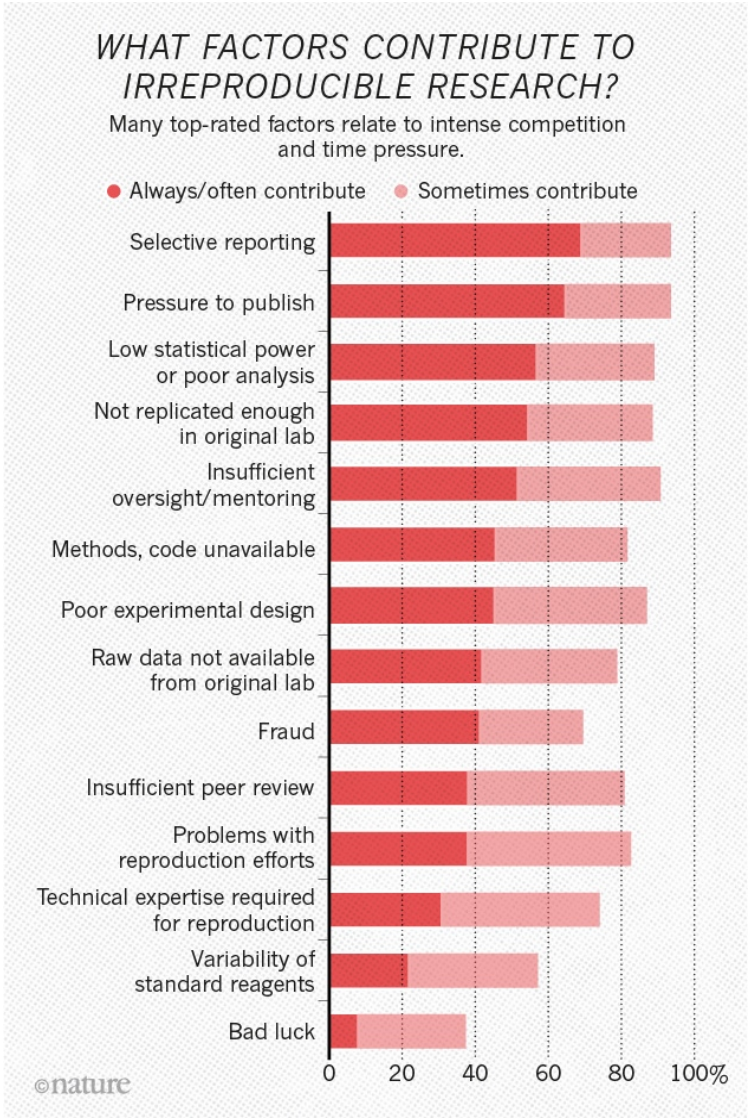
CC-CY. Freedman et al. (2015) PLoS Biol 13(6): e1002165.

Begley, CG, and Ellis L L. "Drug development: Raise standards for preclinical cancer research" Nature. 2012 Mar 28;483(7391):531-3.
Begley, C G, and Ioannidis, J. PA. "Reproducibility in science improving the standard for basic and preclinical research." Circulation research. 2015; 116.1: 116-126.
Chalmers I, Glasziou P. Avoidable Waste in the Production and Reporting of Research Evidence. Lancet. 2009; 374(9683): 86–89.
Howells, D. W., Sena E.S., and Macleod, M.R. Bringing rigour to translational medicine. Nat Rev Neurol. 2014 Jan;10(1):37-43.
Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA."Reproducible Research Practices and Transparency across the Biomedical Literature." PLoS Biol. 2016. 14(1): e1002333.

Reproducibility crisis: researchers' point of view

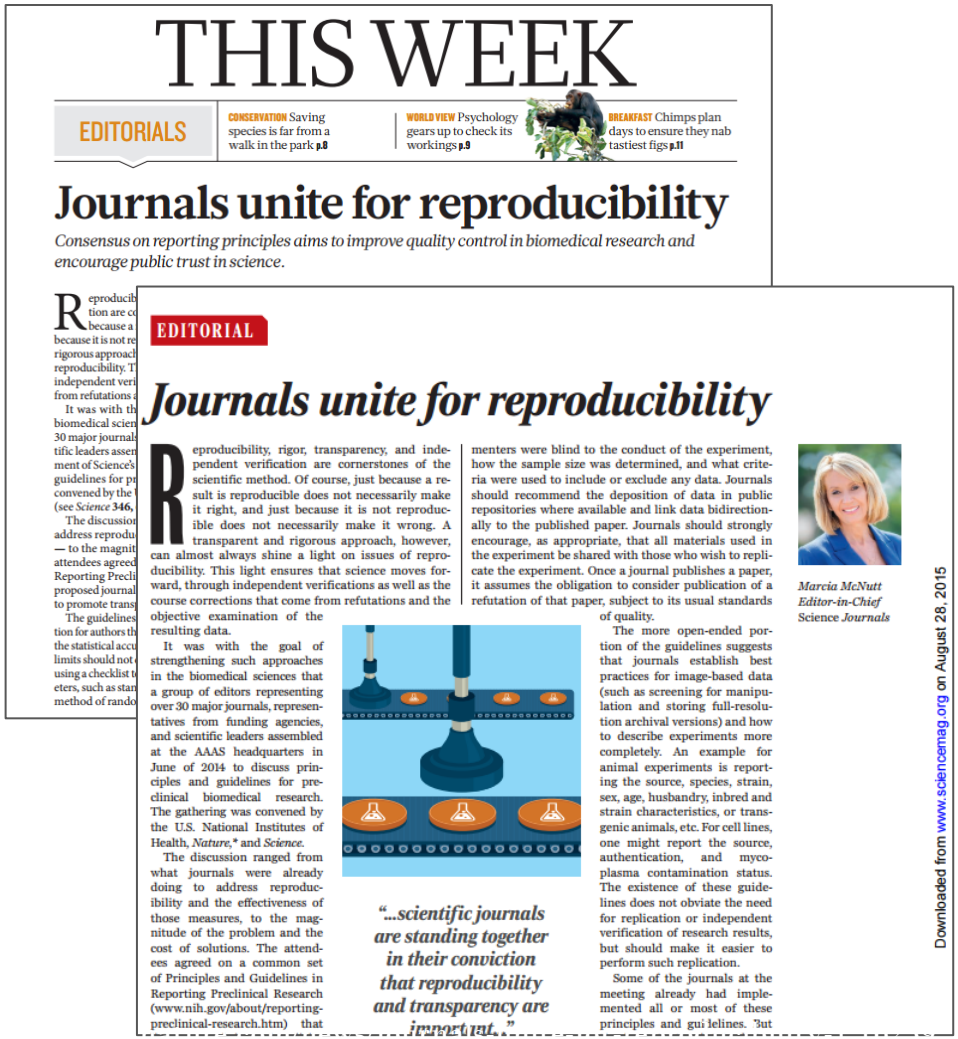






http://www.nature.com/news/reality-check-on-reproducibility-1.19961

Journals: Open Data directives & reporting standard guidelines



THIS WEEK

EDITORIALS

CONSERVATION Saving species is far from a walk in the park p.8

WORLD VIEW Psychology gears up to check its workings p.9

BREAKFAST Chimps plan days to ensure they nab tastiest figs p.11

Journals unite for reproducibility

Consensus on reporting principles aims to improve quality control in biomedical research and encourage public trust in science.

EDITORIAL

Journals unite for reproducibility

Reproducibility, rigor, transparency, and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not necessarily make it right, and just because it is not reproducible does not necessarily make it wrong. A transparent and rigorous approach, however, can almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come from refutations and the objective examination of the resulting data.

It was with the goal of strengthening such approaches in the biomedical sciences that a group of editors representing over 30 major journals, representatives from funding agencies, and scientific leaders assembled at the AAAS headquarters in June of 2014 to discuss principles and guidelines for preclinical biomedical research. The gathering was convened by the U.S. National Institutes of Health, *Nature*,* and *Science*.

The discussion ranged from what journals were already doing to address reproducibility and the effectiveness of those measures, to the magnitude of the problem and the cost of solutions. The attendees agreed on a common set of Principles and Guidelines in Reporting Preclinical Research (www.nih.gov/about/reporting-preclinical-research.htm) that menters were blind to the conduct of the experiment, how the sample size was determined, and what criteria were used to include or exclude any data. Journals should recommend the deposition of data in public repositories where available and link data bidirectionally to the published paper. Journals should strongly encourage, as appropriate, that all materials used in the experiment be shared with those who wish to replicate the experiment. Once a journal publishes a paper, it assumes the obligation to consider publication of a refutation of that paper, subject to its usual standards of quality.

The more open-ended portion of the guidelines suggests that journals establish best practices for image-based data (such as screening for manipulation and storing full-resolution archival versions) and how to describe experiments more completely. An example for animal experiments is reporting the source, species, strain, sex, age, husbandry, inbred and strain characteristics, or transgenic animals, etc. For cell lines, one might report the source, authentication, and mycoplasma contamination status. The existence of these guidelines does not obviate the need for replication or independent verification of research results, but should make it easier to perform such replication.

Some of the journals at the meeting already had implemented all or most of these principles and guidelines. "but

Marcia McNutt
Editor-in-Chief
Science *Journals*

"...scientific journals are standing together in their conviction that reproducibility and transparency are important."

Transparency and Openness Promotion (TOP) guidelines
560 journals and 49 associations
B. A. Nosek et al. Science 2015;348:1422-1425

Principles and Guidelines in Reporting Preclinical Research (NIH, Nature, Science).

Data sharing policies in instructions for authors in the majority of journals. NPG, Cell Press, PLoS, Science, EMBO, PNAS, Lancet, BMJ, BMC, ...

27 May 2016: Europe has announced that all scientific papers should be free by 2020

- EU ministers declared Open access to all scientific papers by 2020.
  - This decision is extended to scientific data behind the articles.

Funding agencies Open-Access and Open Data policies

Funding agency	Policies
SNSF Switzerland Regulations on information, valorisation and rights to research results (PDF, 178 KB)	•Obligation for Gold-OA or Green Road (Self-archiving) within 6 months; •Support costs of Gold-OA APCs (3000 CHF) •Does not support costs of Hybrid-OA •DMP mandatory (2017) •Open data policy in preparation?
Horizon 2020 Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020	•Obligation for Gold-OA or Green Road within 6 months •Self-archiving reporting is requested in all cases •Support costs of Gold-OA or Hybrid-OA APCs •No compliance = funding is reduced •Deposit of the research data recommended and Open data policy in preparation

Open Science Definition

"The conduction of science in a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, with terms that allow reuse, redistribution and reproduction of the research"

https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

Open Science Goals

- Transparency in experimental methodology, observation, and collection of data
- Public availability and reusability of scientific data
- Public accessibility and transparency of scientific communication
- Using web-based tools to facilitate scientific collaboration

Dan Gezelter, http://www.openscience.org/blog/?p=269

Benefits to Open Access (OA) & Open Data

To the author

- Protection against data entropy;
- Improved data management & methodologies;
- Higher diffusion and visibility
- Higher citation rate of your publications (+20-40 %);
- Save your copyright;
- Fulfillment of funding mandate (FNS, H2020,..).

To the community

- Verification of published data;
- Preserving accessibility to data;
- Allowing reuse of data;
- ↗ quality et ↗ reproductibility;
- Foster collaboration;
- Accelerate innovation;
- Educational opportunities;
- Public trust in science;

Reproducibility in science I :

Systematic Review of animal and human studies

15 February 2017

Dr. Sylvie Vullioud (SIS)
Dr. Cécile Lebrand (UNIL/CHUV)

Scientific Information School

UNIL | Université de Lausanne
Faculté de biologie et de médecine

CHUV

Lectures I: Systematic Reviews

Systematic Review of animal studies demo

(Dr. S. Vullioud, SIS)

- How systematic review helps for science validity
- Formulating a suitable and specific research question
- Developing literature search strategies
- Risk of bias assessment

Practical Workshop I:

Systematic Review of animal studies: methodology (4 hrs)

Dr. Sylvie Vullioud
Dr. Cécile Lebrand

Systematic Review (SYRCLE)

- Pubmed/Zoreto

(Search components/field tags/free and Mesh terms /Boolean operators/Pubmed search builder/SYRCLE animal filter)

Publication risk of bias Review (RoB)

- Internal validity
- External validity

Reproducibility in science:

Experimental design

Professor of Statistics in Medicine
Dr. Romain-Daniel Gosselin (Biotelligence)

BIOTELLIGENCES

Lectures II: Experimental design

1. Professor of Statistics' experiences
2. Biostatistics and Design (Dr. R-D. Gosselin, Biotelligence)

➲ Importance of biostatistics and design in reproducibility

➲Introduction to statistics

(Sampling methodology/ Replication/ Independence/ Controlling bias/ Power and sample size/ Outlook of statistical tests/Interpretations of p-values/Data dredging)

➲Null results and publication bias

Practical Workshop II: Experimental Science (2 hrs)

Understand:

Existing guidelines in experimental science
Pseudo replication in the lab
Confounding variables
Importance of pilot studies
Inflation of Type I and Type II errors

How to:

Estimate sample sizes
Reduce sample sizes
Increase power
Blind in experimental research
Block / stratify in the lab

Publish "negative" results
Read publications
Perform post-publication peer-reviewing

Practical Workshop II: Clinical Research (2 hrs)

Understand:

Existing guidelines in clinical science
Observational studies
Clinical trials
Safety *vs.* Efficacy
Non-inferiority and equivalence

How to:

Increase power in clinical science
Reduce the impact of confounders
Reduce bias in patient enrolment
Block / Stratify in clinical science

Read clinical publications

Perform post-publication peer-reviewing

Reproducibility in science III:

Data Management

22 May 2017

Dr. Mark Ibberson (VitalIT/SIB)
Dr. Aude Dieudé (EPFL)
Jan Krause (EPFL)
Dr. Cécile Lebrand(UNIL/CHUV)
Gérard Bagnoud (UNIL/UNIRIS)

Vital-IT
High Performance Computing Center

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Lectures III: Data Management

22 May 2017

1. Experiences from researchers
2. Big Data management (VitalIT/SIB)
3. Data Management Plan (Dr. Aude Dieudé, EPFL)

Data Management

- Increased the quality of your data
- Prevent the loss, preserve the accessibility and reuse of your data
- Ensure the integrity and reproducibility of you research work
- Reinforce visibility and impact, as well as the relevance of your research
- Fulfillment of funding mandate (DMP directives FNS, H2020,..).

Practical Workshop:

Data Mangement Plan

Dr. Aude Dieudé (EPFL)
Jan Krause (EPFL)
Carmen Jambé (UNIL/UNIRIS)
Dr. Cécile Lebrand (UNIL/CHUV)

Data management plan (DMP).

- requirements of the financing agencies (FNS/H2020).
- anticipate in detail the management of your research data (analyses, organization, storage, security and sharing)
- specify the type of data.
- budget, intellectual property, and monitoring over time.

Reproducibility in science:

Sharing Data:

Open Data

22 May 2017

Dr. Cécile Lebrand (UNIL/CHUV)
Jérôme Zbinden (UNIL/CHUV)
Raphaël Grolimund (EPFL)

UNIL | Université de Lausanne
Faculté de biologie et de médecine

CHUV

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Lectures: Open Data

22 May 2017

1. Experiences from researchers
2. Policies from funding agencies (FNS or H2020)
3. Policies from publishers (eLife or PLoS)
4. Data repositories (figshare or Zenodo)

- Open data experiences from researchers
- Benefits to data sharing
- Policies for open data from funding agencies/publishers
- Guideline and standards for improving studies reproducibility
- Deposit their datasets accompanying their publication
- Policies on confidentiality and intellectual property.

Practical Workshop:

Data Sharing

Dr. Cécile Lebrand (UNIL/CHUV)
Jérôme Zbinden (UNIL/CHUV)
Raphaël Grolimund (EPFL)

- Search for datasets
- Benefits to data sharing
- Publish and share data on Zenodo or figshare
- Metadata standards
- File formats for long-term preservation/re-use
- Citation for a dataset
- Confidentiality and intellectual property

For FBM (UNIL/CHUV)

Dr. Cécile Lebrand-Research & Publications Officer
FBM Publication Management Unit-BiUM
Tél. ++41 (0)21 314 50 81
Cecile.lebrand@chuv.ch
http://www.bium.ch/en/publication-open-access/data-management/

For SV/EPFL

Pierre Devaud- Scientific librarian
Pierre.devaud@epfl.ch

Aude Dieudé- Specialist Research Data
aude.dieude@epfl.ch
http://library.epfl.ch/research-data-services/en
datamanagementplan@epfl.ch

For UNIGE

Eliane Blumer- Swiss DLCM Project Manager
eliane.blumer@unige.ch
http://www.dlcm.ch/datacycle

For VitalIT

Vital-IT GroupSIB Swiss Institute of Bioinformatics
https://www.vital-it.ch/about/team